Sentiment Analysis in Electronic Negotiations

Michael Körner

Table of Contents

Index of Figures

Index of Tables

Index of Abbreviations

B2B	Business to Business
BATNA	Best Alternative To a Negotiated Agreement
CMC	Computer-Mediated Communication
CV	Cross-Validation
DSS	Decision Support System
DT	Decision Tree
EASI	Emotions As Social Information
HTML	Hypertext Markup Language
IDF	Inverse Document Frequency
IG	Information Gain
IS	Information Systems
LR	Logistic Regression
kNN	k-Nearest Neighbour
LAP	Language-Action Perspective
NB	Naïve Bayes
NSS	Negotiation Support System
RBF	Radial Basis Function
SIP(T)	Social Information Processing (Theory)
SVM	Support Vector Machine
TF-IDF	Term Frequency – Inverse Document Frequency

1. Introduction

1.1. Motivation and Research Objective

As electronic communication has further and further diffused into minuscule areas of entrepreneurial routine, its importance as a means to conduct a variety of economic tasks effectively and efficiently is not only undisputed, but has shifted to be a core area of improvement of both practitioners and researchers alike. In this respect, appropriate media usage becomes a substantial factor regarding effective and efficient communication processes, both intra- and interorganisational, so that communication media can serve as an enabler of a fully globalized business world. Curiously, many entrepreneurial tasks are potentially not well-suited to be conducted using electronic media due to their inherent complexity. As such, negotiations serve as a paradigmatic example of such a complex, mixed-motive interaction, the conduct of which is more and more shifting towards electronic media – in practice mostly via electronic mail. The stance that marks one of the foundations of this thesis is that it should be a core task of the communication medium itself, to offer additional supporting means in order to increase communication efficiency and effectiveness, especially to counteract detrimental effects that are caused by the usage of the medium itself. Concerning the domain of study in this thesis, namely electronic negotiations, Negotiation Support Systems (NSSs) seek to provide decisional aids to negotiators to reduce cognitive complexity of the task, as well as communication support functions in order to clarify ambiguous conveyances of meaning in the negotiation intercourse. Here, research has acknowledged that there is a need of a shift in paradigms of support towards a more proactive approach where systems follow the negotiation process and activate support functions as necessary (Druckman et al. 2012). To achieve such a task, it is mandatory for the system to possess the ability to evaluate a given negotiation situation in order to judge whether it should become active – therefore, the system needs a basic understanding of regularities and patterns occurring in negotiations that are indicative of specific outcomes. Decision Support functionalities, for example, use concession patterns to estimate the final outcomes in multi-attribute negotiations (Vetschera 2016). Negotiators' communication unfortunately rarely possesses explicit, known patterns that are common with respect to negotiation outcome.

Concerning this, methods of Predictive Analytics provide a viable possibility to implicitly work on communication patterns and, more specifically, differences in communication between successful and failing negotiations. Especially since the advent of Web 2.0 and social media, Predictive Analytics on unstructured documents have become a core task in Information Systems (IS) research and practice. For this reason, research fields such as

Sentiment Analysis and Text Classification have boomed over the last 15 years, offering means to convert unstructured data from subjective documents such as product reviews or social media statements into economically meaningful knowledge and recommendations to take action. As electronic negotiation transcripts are essentially series of opinionated documents exchanged between the acting entities in the negotiation, methodological concepts from Sentiment Analysis should translate well to electronic negotiations as an application domain. Harnessing these principles and approaches, this thesis seeks to provide a mechanism that can be utilized by NSSs in order to distinguish negotiation outcomes based on the communication exchanged and thus providing a situational evaluation for an NSS on which a judgement to proactively enter the negotiation process can be based. Hence, this thesis contributes to answering the following research question:

How can methods of Predictive Analytics and Sentiment Analysis provide an adequate means of detection of electronic negotiation outcomes, based on the communication exchanged?

In order to answer this research question, a series of studies presented in the following chapter were conducted that represent – in IS research terminology – a Design Science cycle in which classification models for electronic negotiation transcripts are constructed and evaluated regarding their classification performance, with knowledge from Computer-Mediated Communication and relevant communication and negotiation theories serving as the theoretical foundation.

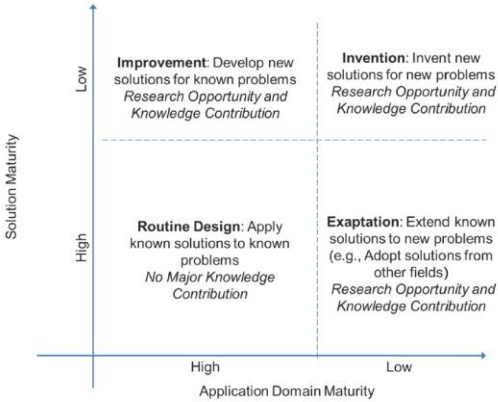

Figure 1: Design Science Knowledge Contribution Framework (Gregor & Hevner 2013)

Gregor & Hevner (2013) give a basic framework for knowledge contribution in IS design research shown in Figure 1. The research conducted in this thesis can be placed into the improvement quadrant, which is described as an area where the problem domain is known (there exists previous work on negotiation classification), and a novel way of solving the problem is discussed and presented (in our case a Sentiment Analysis approach). This thesis seeks to improve existing means of negotiation classification by drawing on fundamental knowledge from the domain field of Computer-Mediated Communication and electronic negotiations (as intended in Hevner & Gregor, 2013 p.346).

As such, this thesis focuses on communication processes within negotiations. Whilst in overarching phase models of complete negotiation processes, communicating in the *preparation phase* to the actual negotiation and communication in the *post-settlement* phases is accounted for as well, it is considered out of scope for the remainder of this work.

1.2. Structure of the Thesis

In order to answer the research question posed earlier and to apply Sentiment Analysis and Predictive Analytics methods to electronic negotiations, four studies were conducted, which are depicted in figure 2:

First, an overarching theoretical introduction is given in the following chapter, including existing research on communication support in electronic negotiations, discussing the influence and function of communication with respect to the negotiation outcome and arriving at a synthesis discussing the applicability of Sentiment Analysis to the electronic negotiation domain.

Study I marks the first approach to the topic, in which purposes and potentials of the application of Sentiment Analysis to electronic negotiations are presented and discussed. Furthermore, since electronic negotiations were identified as a rather specific application domain for Sentiment Analysis, the construction of a domain-dependent sentiment lexicon is recommended. To this end, different approaches to generate such a lexicon are analysed and a variation of these approaches is conducted on an existing negotiation data set. The result of Study I is said lexicon, which serves as foundation to Study II, where it is used in the training of feature-based Machine Learning classifiers. Here, we formulate the negotiation classification task explicitly in the context of Predictive Analytics, using Shmueli and Koppius' (2011) model approach. At the end, performances of the classification model on data sets of complete, half and three quarters of negotiation transcripts are presented and discussed.

Study III builds upon an approach which in Study I was already briefly mentioned as a means to improve classification quality. The classification models developed over the course of study II act on the document granularity only in order to evaluate complete negotiations regarding their success or failure. Sentiment polarities are employed in a passive fashion, simply as additional classification features without much differentiation. Oftentimes in Sentiment Analysis, a more fine-grained approach is desired which is interesting in that it allows separation of subjective and factual content and more detailed statements about which elements contribute more to the overall evaluation than others. Hence, we employ means of subjectivity and polarity classification at the sentence-level of our negotiation data. To achieve this, a corpus of 28667 sentences from negotiations has been manually prepared by human coders, similar to Content Analysis methods, and – using these manual codes – subjectivity and polarity classifiers for negotiation data that act on the sentence level are trained and evaluated.

The concluding Study IV is the synthesis of the previous studies. Here, the main focus lies on the integration of the information given by the sentence-level classification into the document-level classification process. How exactly this integration should ideally be conducted is subject to ongoing research in Sentiment Analysis and, most likely, dependent of the application domain. Therefore, we discuss different integration alternatives and select variations to train a cascaded classification model that enriches negotiation outcome classification with sentence-level subjectivity and polarity information.

Lastly, we give an overall discussion in order to put the conclusions fragmented over the studies back into the larger research context, which is to employ such classification models in Negotiation Support Systems and use their predictions as an entry point for a proactive system to actually activate itself and to further deduct diagnostic information regarding the reasons for failing negotiations. Based on this, possible further recommendation steps for the NSSs are discussed, including diagnosis, advice and visualization. Furthermore, limitations of the overall approach are discussed and ideas are presented as to what future research on this topic could encompass.

Preface: Theoretical foundations

- Negotiator's communication in electronic negotiations
- Applicability of Sentiment Analysis
- Text Classification and Sentiment Analysis techniques

Study I:
Sentiment-Based Assessment of Electronic Mixed-Motive Communication - A Comparison of Approaches

- Discussion of approaches in Sentiment Analysis
- Development of a sentiment lexicon for electronic negotiation data
- Discussion of application potentials of the sentiment dictionary

Study II:
Feature-Constructed Prediction of E-Negotiation Outcomes based on their Communication Content

- Construction of a feature-based Machine Learning classifier
- Discussion of model variations and selection procedure of a model
- Training and application of the model to transcripts of complete and incomplete negotiations
- Performance analysis

Study III:
Micro-Level Sentiment Assessment of Negotiators' Utterances

- Fine-grained application of Sentiment Analysis to sentences from negotiation transcripts
- Training of sentence-level classifiers predicting subjectivity and polarity of negotiator utterances
- Performance evaluation

Study IV:
Enriching Negotiation Transcripts with Sentence-Level Information to Improve Negotiation Outcome Classification

- Synthesis of Study II and III
- Discussion of integration variants to incorporate sentence-level information into the negotiation classification process
- Training & analysis of different classifiers on negotiation level

Reprise: Conclusion & Outlook

- Critical assessment
- Proactive communication support activities
- Limitations & future research

Figure 2: Structure of the thesis

2. Theoretic and Methodological Background

2.1. Existing Communication-Centric Approaches to Electronic Negotiation Support

The main focus of the research presented in the course of this thesis is on *electronic Business to Business (B2B) negotiations*. In order to achieve a common understanding, this section will give basic definitions of the negotiation terminology and scope that is followed in the thesis.

Bichler et al. (2003) view a negotiation "*as an iterative communication and decision making process between two or more agents […] who: 1. Cannot achieve their objectives through unilateral actions; 2. Exchange information comprising offers, counter-offers and arguments; 3. Deal with interdependent tasks; and 4. Search for a consensus which is a compromise decision"* (Bichler et al. 2003, p.316). Negotiation is thus a specific form of a mixed-motive interaction process (Komorita & Parks 1995), where actors have to keep a balance between reaching their individual goal, (e.g. in the form of a result that benefits them individually) and reaching a joint goal, i.e. making the final consensus also acceptable to the other parties involved so that an agreement can be implemented at all. To achieve these goals, parties enter a complex communication process involving decision-theoretic as well as socio-emotional aspects that have to be considered.

The advent of Computer-Mediated Communication in B2B interactions during the 1980s sparked the notion that such complex interaction processes will be conducted using electronic media instead of the classic face-to-face approach, affecting and reshaping the nature of the communication process (Kiesler & Sproull 1992). As a means to alleviate the impediments that a negotiator will suffer from and in order to instead exploit the additional means that a Computer-Mediated Communication scenario allows for, the concept of Negotiation Support Systems (NSSs) was devised. Originally, NSSs were perceived as a system that one the one hand offers a communication means between negotiators, and on the other hand includes a Decision Support System (DSS), where negotiators can specify their preference structures for a given negotiation agenda, and which allows to directly evaluate received and sent offers during the negotiation process (Lim & Benbasat 1992). Whilst in the original perception of NSSs, the DSS aspect was already described as allowing multifaceted types of individual and joint support capabilities, the communication channel as second component was rather understood as a given to exchange task-oriented information than as a component that can enrich, structure and thus support negotiators' communication processes – which rather is a perception of purely "electrifying" the communication without any adaptions to the new medium.

This understanding of NSSs prevailed for several generations of systems, until the need for supporting not only the decision-theoretic aspect but also the communication process itself was identified (Schoop & Quix 2001, Weigand et al. 2003, Yuan et al. 1998). Whilst structured communication flows in the form of protocols had already been briefly discussed in the context of one of the first NSS prototypes (Jarke et al. 1987), further integration of communication-structuring aspects had been rare up to this point. Novel NSSs such as Negoisst (Schoop et al. 2003, Schoop 2010) centred on a more thorough support in this aspect, by structuring content and message exchange and linking this content to a document management component which derives and tracks contract development over time (Schoop and Quix 2001). The theoretical foundations of this type of communication support centred on the Language-Action Perspective (LAP, Winograd 1988), and respectively Habermas's theory of communicative action (Habermas 1984) and Searle's speech act theory (Searle 1969).

The LAP is a research stream that focuses on giving principles and paradigms on how communication systems should be designed to provide ideal support to their users (Schoop 2001, Schoop & Quix 2001). Seminal works mostly focus on the ideal that communication via systems should be well-structured and as clarified as possible and that a communication system should be specifically designed to provide functionalities that allow for structuring and avoidance of unclear communication (Winograd 1988). As the communication-theoretic roots for LAP mainly lie in the works of Searle and Habermas, LAP can thus be perceived as a form of synthesis of principles of these theories with a focus on system design. Negoisst emphasizes this notion in that it similarly transforms Habermas' and Searle's principles into functionality:

Speech act theory emphasizes a separation of any communicative act into propositional content which is what is actually uttered and illocutionary force, which is the intention that the speech act should achieve. Hence, Negoisst provides a similar separation through the concept of message types which explicitly state the illocutionary point of any message and actual message content. The message types available are predefined by the system and can be mapped to the different types of speech acts that Searle defines (Schoop et al. 2003).

Furthermore, Habermas' theory of communicative action defines four validity claims for successful communication: Comprehensibility, truth, truthfulness and appropriateness. If during a communication activity, one of these claims is challenged, conflict arises which can be resolved via discourse. This notion is reflected in the system through the allowance of question/clarification cycles which do not contain any commitment but serve to implicitly clarify arising challenges in validity claims between the negotiators (Schoop et al. 2003, Schoop 2010, Weigand et al. 2003). The appropriateness claim is further supported by a communication protocol, which limits the available message types at a given point in the negotiation.

What all these supporting functions have in common is that, following the principles of LAP, they mainly offer structuring and passive, regulating support. Components that proactively aid the negotiator during the process are not explicitly considered in these previous works. Partly this is due to the fact that communication support is still maturing (Schoop 2015), as the complex socio-emotional interaction process during electronic negotiations is not well understood up to now. An important precondition for the provision of proactive means is for an NSS to be able to detect *entry points* into the human negotiation process, i.e. to detect at some point during the negotiation that it is time to activate itself and offer supporting means to the negotiators. Kersten & Cray (1996) defined this entry point notion as a necessity in the case of decision support and automation of the negotiation. Similarly, for proactive communication support this condition should hold true as well – there exist numerous examples of "unwanted" supporting functions in systems being perceived as annoying and aggravating (a well-known example being the infamous Microsoft Office Assistant, better known as "Clippy"). Since on the communication layer of a negotiation, a huge number of subliminal socio-psychological processes are at work, the direct detection of such entry points where adequate support can be provided should be especially difficult. However, this thesis seeks to provide a means for such an entry point detection, using sophisticated technological means on negotiators communication, which may allow for an NSS to deduct potentially arising conflict situations using the negotiator's communication data. To do so, we will first elaborate on communicational influences in electronic negotiations, more specifically what exactly is contained in negotiator's communication and how it may be related to the negotiation outcome, which is the goal of the following chapter.

2.2. Socio-Emotional Effects in Electronic Negotiation Communication

Research on socio-emotional effects and electronic negotiators' communication originates in a variety of streams. Whilst originally, most research on negotiations put a strong focus on decision making, communication was often interpreted from a prescriptive point of view, as means of negotiators to exchange priority information on the issues to be discussed, ideally in the form of complete information disclosure ("Full, Open, and Truthful Exchange", Raiffa 1982). During the 1980s, this view was extended by scholastic work laying a stronger focus on how communication influences and shapes negotiations in a more descriptive perspective. Contrary to previous research, which was mostly conducted by groups with a mathematical or economic background, an increasing number of scholars from the social sciences and linguistics analysed negotiation processes (e.g. Donohue 1981, Putnam & Jones 1982). In these perspectives, communication is not viewed as only a part of negotiation, but to be at the very core of the whole interaction process, integrating transmission of information and development of (shared) meaning, reflecting power relations, influence and persuasion processes as well as aspects of identity management (Putnam & Roloff 1992).

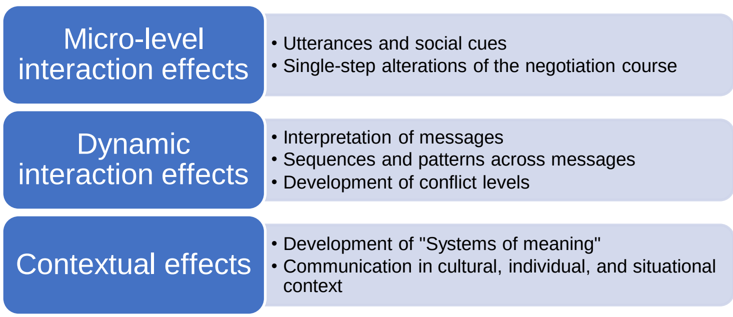

Figure 3: Effects of Communication in Negotiations (Putnam & Roloff 1992)

Figure 3 shows a brief outline of the impacts and effects (and thus, potential research foci) of communication defined by Putnam and Roloff (1992), roughly dividing interaction effects in communication into three dimensions. On the micro-level, single utterances and exchanges of social cues shape the negotiation process step by step. On a higher granularity, these utterances form sequences and patterns which, combined with the meaning given to them by the communicators through interpretation, guide the overall development of conflict level. Again in combination with contextual factors, of cultural, individual or situational nature, so-called “Systems of meaning” are developed between the negotiators. In this view, language itself (spoken or written) is viewed as a form of micro-level manifestation of the bargaining process, the course of the negotiation is determined by what is conveyed on the micro level, which in turn is influenced by a variety of contextual factors.

From a researcher’s point of view, this leads to different possible perspectives how exactly communication impacts negotiation processes and outcomes. Here again, Putnam & Roloff (1992) distinguish two different general approaches: The effects approach, which reflects a classical behaviouristic perspective in viewing communication as an influencing variable in different research flavours (mediating, moderating, direct influence, black box) and the key components approach, where communication is viewed as a set of sequences, developments and patterns in messages which in turn construct the bargaining process (Putnam & Roloff 1992). Concrete research rarely assumes a pure form of either of these approaches, but oftentimes induces a tendency to favour one over the other. Similarly, this thesis favours a keys components perspective, where the concrete manifestations, interaction sequences and patterns of communication frame the negotiators’ perspective and evaluation of the negotiation situation and thus produces subjective evaluations which in turn again manifest themselves in the negotiation communication, and ultimately take impact on the negotiation outcome. Hence, it is especially important to understand the function of communication in negotiations i.e. what exactly is communicated by the negotiators and how negotiation outcomes are affected by it.

To this end, Duckek (2010) distinguishes basic communication functions into a *procedural*, a *factual*, and a *relational* layer. These layers are not explicitly selective, i.e. many communication elements and constructs discussed in negotiation literature simultaneously act on multiple layers, but they provide a concise framework for discussion and explanation of communication functions. Similarly, it reflects the notion in Computer-Mediated Communication models of a separation of a communication task into a cognitive and an affective dimension (Te’eni 2001) as well as the distinction often made in negotiations into a substantive and a relational dimension (Lewicki et al. 2010).

Communication is oftentimes seen as a means to manage the cognitive complexity when conducting a joint task (Te’eni 2001). Likewise, the *procedural* communication function’s main concern is to structure the negotiation process by separating the discussion into different segments, so that simpler negotiation subtasks emerge that can be solved more easily. Thus, it contains a form of collaborative development of a shared understanding and sense-making (Swaab et al. 2004, Putnam 2010) and serves as a vehicle to enable identification of joint or diverging interests (Duckek 2010). Typical communication content here is process-specific meta-communication (e.g. clarifying the sequence in which negotiation issues should be discussed), development of a shared definition of terms and aspects specific to the negotiation and other process-structuring exchanges (e.g. timeframes for the negotiation, scheduling future interactions etc.).

The *factual* layer is concerned with the actual joint decision-making process itself. Here, offers and counteroffers are exchanged and discussed, the offer space is explored, integrative potential is identified and, lastly, solutions are developed, reframed, or abandoned. Thus, it is the layer in which preference information is requested and exchanged, reasoning for offers is given and where negotiation tactics are applied ranging from integrative logrolling over persuasion to hard negotiation tactics (Keough 1992). Whilst this layer should as per its content be dominated by rationality, in reality this is often not the case. An integral part of offer communication is how exactly offers are formulated and framed in order to achieve a desired outcome from the negotiation counterpart (Tutzauer 1992). Hence, there is often an interconnection to the relational layer, especially when persuasion (such as an appeal to the mutual relationship and fairness norms) and hard tactics (such as threats to leave the negotiation table) are applied.

The *relational* layer itself emphasizes the management of the relationship between the negotiators that implicitly or explicitly takes place in every negotiation. Adequate relationship management is the foundation of building an atmosphere of trust and information disclosure, which allows for a better exploration of integrative potentials and achieving common ground (Clark & Brennan 1991, Morris et al. 2002). Therefore, this layer contains exchanges of assessments of the negotiation situation (Gibbons et al. 1992), explicit discussions of the relationship situation and, most importantly, affective communication, which directly leads to another important research field that is inherently intertwined with communication – the study of emotions in negotiation. Emotional evaluations of different negotiation aspects (e.g. the behaviour of the counterpart, the quality of offers, or the general negotiation progress) are present throughout negotiation transcripts and communication function layers discussed above, providing a guideline of how negotiators perceive the progression of the negotiation and are perhaps the strongest indicator, at least to the human mind, how likely a settlement given a certain negotiation context is. This mindset also induced the classical perception of a rather simplistic linkage between emotions during the process and negotiation outcomes. The argumentation is as follows: The expression of positive emotions in the negotiation process facilitates relationship- and trust-building, induces a friendly atmosphere, opens the negotiation parties to share information and to put in effort towards collaborative, integrative solutions and, again, reciprocation of the positive emotions displayed (see for example the emotional contagion stance in Thompson & Nadler 2002 or the affect infusion model in Forgas 1998). This process of reciprocation leads to higher likelihoods of settlement, a better quality of negotiated agreements (Forgas 1998) and, lastly greater satisfaction of the negotiators with process and outcome (Morris & Keltner 2000).

However, in recent research a more differentiated view has spawned, namely in the form of the Emotions as Social Information (EASI) model (van Kleef 2009). According to EASI, the perception of emotional expressions is not only influenced by affective mechanisms, but also by cognitive inferential processes. Whilst affective mechanisms depend on socio-relational factors, cognitive inference is moderated by the rational information processing capability of the actor. Both of these mechanisms in turn influence the reaction to the expressed emotion. This means, if an observer is able to rationalize negative/positive emotional expressions by the counterpart and evaluates these expressions as justified and/or socially appropriate given the context in which they are uttered, he may adapt his reactions accordingly. Note that this view is not directly opposed to the classical interpretations of reciprocity, but rather complementary, putting the classical interpretations in a more differentiated context.

Now, regarding the impact of these findings on negotiators' communication it is underlined that the relationship between emotions and outcome does not necessarily follow a simplistic pattern, where positive emotions are generally good and negative emotions are bad in principle. Rather, it becomes obvious that aside from the context it becomes especially important *how* emotions are conveyed and regarding *which aspects* of the negotiation process, since affective communication seeps through all the layers of communication function discussed above. Furthermore, it is also noted that emotions are not only conveyed explicitly, but also in an implicit manner through lexical constructs that are not only used in subjective statements but that can also provide an emotional framing of otherwise factual statements (Gibbons et al. 1992, Griessmair & Köszegi 2009). Hence, statements of emotional polarity may influence the negotiation process already on the level of single utterances and thus the conveyance of emotions becomes an inseparable aspect of research on communication influence on negotiation outcomes.

In electronic negotiations, another important aspect is how the usage of an electronic medium as opposed to traditional, face-to-face settings influence and alter the communication process between the negotiators. Originally, when the concept of

Negotiation Support Systems was devised, using an electronic communication medium was perceived as advantageous for the process, leading to a more task-oriented, de-emotionalized approach to negotiating, which should result in more objective, and ultimately, better negotiation outcomes (Lim & Benbasat 1992). This understanding has changed greatly over the recent decades of research, with the main influential body of work to understand electronic negotiators' communication stemming from theories of Computer-Mediated Communication (CMC). In fact, using electronic media to conduct negotiation has a differentiated and profound impact, altering the communication and decision making process.

Generally, the CMC theories that were used to explain electronic negotiation behaviour can be subsumed under two perspectives, the *cues filtered out view* and the *cues filtered in view* (Walther & Parks 2002). These terms date back to the concept of *social cues*, which are exchanged between actors in every communication process. Social Presence Theory (Short et al. 1976) first introduced the notion that communication effectiveness is influenced by the intensity of the perceived *social presence* of actors in a communication system. The main determinant for this social presence is ability of a medium to exchange social cues, ranging from lexical choice of words to paralinguistic cues such as pitch, tone of voice, mimics, or gestures. Social cues thus serve as indicators for the actors to determine their relationship as well as the affective state of the interaction. The more channels for the exchange of these cues are available, the stronger the social presence of actors becomes, and – according to Social Presence Theory – the more powerful a communication medium is to conduct effective communication resulting in improved task performance. Hence, face-to-face meetings are deemed to be the richest medium available, whilst electronic communication (especially written communication only) has a rather low social presence, due to many channels for the exchange of social cues being unavailable.

A similar connection between choice of communication medium and task performance was coined in Media Richness Theory (MRT, Daft et al. 1987), which assumes a more prescriptive view in recommending to select appropriate communication media based on the nature of the task to be conducted. Task nature mainly is distinguished by the information processing capability necessary for an efficient conduction, which in turn is influenced by *complexity* of the task itself as well as the *equivocality*, i.e. how likely it is that ambiguous situations occur over the course of the communication process. Daft et al. then distinguish media according to their richness based on similar aspects than Social Presence Theory, and argue that the more complex and ambiguous a task becomes, the richer the used medium should be.

The third theory that is strongly in the tradition of the Social Presence approach is contained in the work of Sproull and Kiesler (1986). According to their line of argumentation, the reduction of social cues in electronic media introduces a tendency of CMC actors towards antisocial behaviour in the form of increased aggression (“flaming”) as well as a greater likelihood to embark on risky decisions and adversarial communication (Sproull & Kiesler 1986, Kiesler & Sproull 1992). With regard to this aspect, the Social Identity and Deindividuation Effects Model (SIDE, Lea & Spears 1992) argues in a similar direction, suggesting that impression formation is obstructed in lean media, which leads to deindividuation effects and, as a result, less inhibited communication behaviour and an impression formation of the counterpart that is mainly based on in-group and out-group perceptions. Similarly, Byron (2008) notes that there also exist distortions in CMC actors’ perception of messages in that electronic communication tends to be interpreted in a more negative fashion than intended by the sender.

While the cues-filtered out approaches in CMC hold their own value in analysing communication efficiency, their prescriptive aspect of selecting appropriate media never made it into B2B communication reality. For years there has been a strong tendency to employ lean media, such as electronic mail or even shorter text exchange media such as WhatsApp in the context of complex tasks. Hence, the focus of research has rather undergone a shift from “Which media should be used for which task?” to “How can the actually employed media be used efficiently, so that even complex communication tasks can be fulfilled appropriately?”. In fact, early research already indicated that actors engaging in complex tasks using a lean medium already try to employ compensation strategies in order to increase media efficiency – and their failure to do so may even be an antecedent of the alleged negative effects of CMC (Markus 1994). In line with this, many classical theories departed from the notion that a certain medium has a “fixed” suitability for a given task, and that other factors such as experience with the medium may have a stronger influence on CMC effectiveness (Walther 2011).

In strong contrast to the cues-filtered out approaches Walther thus proposed the Social Information Processing (SIP) approach to CMC, laying the foundation for what was later called the “cues filtered in”-perspective (Walther 1992, Walther & Parks 2002). While SIP acknowledges that “lean” media, such as written communication, lack the traditional channels of social cue exchange, it proposes that these channels are – given repeated interaction over a frame of time – not needed for effective communication as well as relationship-development between actors in CMC. Instead, the remaining channels gain in importance for the interpretation of the counterparts’ message and in turn for impression formation about the counterpart. Additionally, novel forms of social cue exchange are devised by the actors to compensate for the existing ones, such as including socio-emotional markers in the content of the messages, alterations to writing style or placing importance on other meta-attributes such as the timing of messages (Walther 2011). In a concurrently developed theoretical approach, the Hyperpersonal model (Walther 1996), it was furthermore pointed out that unique abilities of the communication medium used are exploited by the users to increase the quality of the message transmitted, e.g. in asynchronous e-mail scenarios, actors take their time to revise, review and improve the quality of their message before transmission.

In recent CMC research, this approach to understanding media effects on communication had a strong prevalence. Whilst other theories – especially media richness – are still often used in research work, there is currently a state of coexistence of these approaches, with a tendency of research trends to shift into the direction of cues-filtered in approaches. As an exemplification, revisions of the SIDE model saw a departing from the strong deindividuation assumptions of the original model – it was found that with the introduction of a temporal component, deindividuation effects decline (Postmes et al. 2005, see also Walther 2011). Nonetheless, the understanding of how communication media shape interactions is still a subject under ongoing discussion.

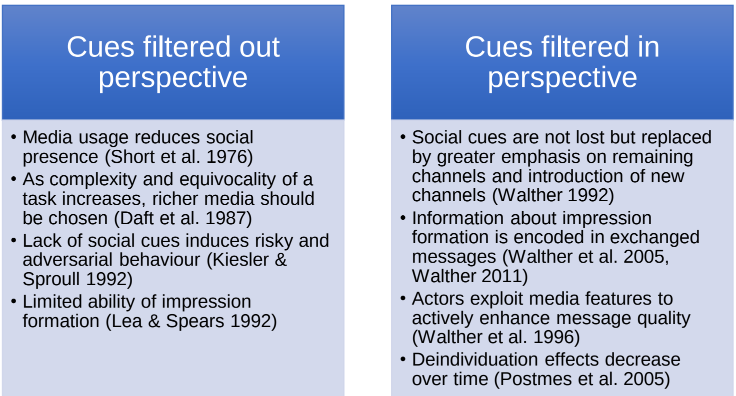

Figure 4: Theoretical summary of media alterations to communication

Figure 4 provides a brief overview on the previous subsection, summarizing the alleged alterations to communication, impression- and relationship-formation in prevalent CMC theories.

Now, putting these alterations under the lens of electronic negotiation communication a picture emerges that mostly classifies the propositions of the cues-filtered out approaches as detrimental for negotiation tasks, whilst the cues-filtered in approaches rather tend to predict no influence, or at least no harmful influence of the electronic medium on negotiations (see Figure 5). Social Presence Theory as well as Media Richness Theory suggest that written electronic communication should be unsuitable for complex negotiation tasks, either because the reduced social presence decreases communication efficiency (SPT), or because the task is too complex and equivocal to avoid misunderstandings – again resulting in a decrease of communication efficiency (MRT). Additionally, detrimental shifts in behaviour can occur, resulting in overly risky decisions, adversarial communication styles and application of hard, aggressive negotiation tactics (SIDE). Similar effects have been described in e-negotiation literature. Morris et al. (2002) describe that in e-mail only interactions, rapport building between negotiators is obstructed by the medium, resulting in a greater likelihood of impasse – whilst negotiators that already had previous contact aside from the actual negotiation medium (face-to-face, or even a brief telephone conversation) showed no signs of failing to build rapport. Thompson & Nadler (2002) discuss different communication biases that occur frequently in electronic negotiations. According to these descriptions, e-negotiators employ more risky strategies such as trust-testing and hard interpersonal tactics that have a high chance of resulting in a defective relationship (also see Purdy et al. 2000). Furthermore, the emotional style tends to be more aversive and negotiators are prone to assume ulterior, sinister motives behind the communication of their counterpart, negatively affecting trust-building and thus leading to less information sharing which makes it more difficult to identify mutually beneficial outcomes.

On the other hand, using an electronic medium introduces opportunities for the negotiators to take a more active stance in managing the communication themselves. The possibility to structure the exchange already remedies parts of the detrimental effects of the task complexity if the medium is used in an adequate manner (Te'eni 2001). As already mentioned in the previous chapter, early NSS research assumed that communication would change towards a de-emotionalized, task-oriented style which should improve overall outcomes (Lim & Benbasat 1992). Hence, it has been argued that while the conduction of negotiations electronically may not be as efficient as a face-to-face meeting, it induces an atmosphere of neutrality, where status difference effects are lessened (Poole et al. 1992), exchanges are well-documented and reviewable at any point in time (Friedman & Currall 2003), and the focus on the negotiation task itself is greater, inviting a better framework to explore the negotiation space in order to identify joint outcome potentials. Furthermore, since negotiators in asynchronous settings possess the opportunity to revise their own messages before sending, they also can be more conscious of their argumentation lines as well as the general impression they make on their counterpart (Friedman & Currall 2003) – which, again, if the medium is managed properly by the negotiators should positively affect trust- and relationship-building, especially in repeated interaction scenarios (Walther 1992, Jarvenpaa & Leidner 1998).

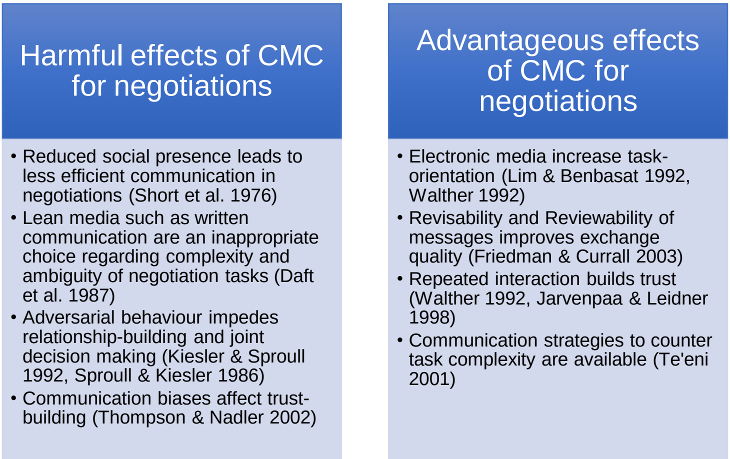

Figure 5: Electronic negotiation communication effects according to CMC theories

All in all, the overall role of CMC effects on negotiations remains disputed and ambiguous, strong arguments for harmful as well as for beneficial effects having been made. Nonetheless, entrepreneurial reality suggests an increasing importance of electronic negotiations, asynchronous computer-mediated exchanges are in fact common for complex negotiation tasks (Schoop et al. 2008). Therefore, management of the communication itself, by the negotiators as well as by the medium becomes a core research task in electronic negotiation, avoiding the pitfalls connected to CMC while at the same time harnessing potential benefits.

2.3. Synthesis: Negotiations as Classification Domain to Enable Proactive Communication Support

As an interim conclusion of the previous section, communication is undisputedly a crucial factor in negotiations, acting as the main vehicle in socially constructing the negotiation process. Especially in the socio-emotional exchange, relationship-building, establishment of trust, transmission of social cues as well as expressing evaluative (emotional) statements about the state and development of the negotiation, communicative details such as timing and lexical choice are crucially altering the negotiation process.

Putnam and Roloff (1992) early on suggested the conveyed language as a potential research focus, analysing how small stepwise alterations shape the negotiation process. Their understanding of communication effects is quite similar to the perspective taken by SIP. According to SIP as well as to research on emotions in electronic negotiations, as well as the research directions proposed in early negotiation communication research, this thesis takes the perspective that electronic negotiations aside from the actual offer-communication consist of a stream of evaluative expressions that constantly reflect the perceptions of the negotiators about the decision-theoretic progress of the negotiation, the state of relationship- and trust-building between them and their emotional evaluation of the negotiation situation and counterpart. Guided by the assumption that there exist context-invariant regularities in these exchanges, we assume that these regularities differ between negotiations that end in a successful conclusion and those that reach an impasse where at least one party is ready to stop negotiating altogether. This perspective is inherently close to what Sentiment Analysis assumes (see the following chapter) which enables methods and techniques from Sentiment Analysis to be applied on the domain of electronic negotiations.

An important aspect of this work regarding the CMC-Theoretic perspectives is that we purposefully decline taking a definite adherence to either of the two cues-based directions, but rather acknowledge the existence of influential factors predicted by both sides. Electronic negotiation research contains a multitude of results favouring either a cues-filtered out (e.g. Morris et al. 2002, Thompson & Nadler 2002) or a cues-filtered in perspective (e.g. Burke et al. 2002). This dual nature is also consistent with Walther (2002) noting that the two CMC perspectives should not be viewed as antithetic but rather as complementary, until an explicit unification is found. Hence the perspective taken in this thesis accepts that detrimental effects such as communication biases and adversarial behaviour may occur in electronic negotiations, but at the same time relationship-building may function as well given repeated interaction over the negotiation course and that this information is at least partially encoded in the communication exchanged by the negotiators.

Now, given the downsides of CMC for electronic negotiations, it has been stated multiple times (Weigand et al. 2003, Griessmair & Köszegi 2009) that it should be the task of the system (in this case NSS) to counteract the disadvantages of employing it, by providing adequate means of support regarding the communication aspects of the negotiation process. Based on this understanding NSSs such as Negoisst (Schoop et al. 2003) employed passive means of communication support, offering to structure the negotiation process using communication protocols, to clarify intentions expressed with message types and to provide clear linkages between negotiation messages and contract development, thus seeking to resolve potentials for ambiguous situations. In recent research, calls for more proactive means of communication support were expressed as well (Curhan & Pentland 2007, Kersten & Lai 2007, Druckman et al. 2012), providing NSSs with the capabilities to actively intervene into the negotiation process when they detect irregularities that may endanger a successful negotiation outcome. In terms of negotiators' communication this means in order for such a proactive support to be practical, the system should be able to differentiate between negotiations being likely to succeed or fail as early in the negotiation process as possible, in order to provide a notion of an "entry point" where the system can decide to take action in an ongoing negotiation (cf. Kersten & Lai's (2007) understanding of proactive systems in this respect).

Previous research on negotiation data strongly suggests that there exist differences in communicational content between succeeding and failing negotiations which can provide such entry points, ranging from meta-attributes such as message length and timing (Kersten & Zhang 2003), over emotional intensity (Griessmair & Köszegi 2009), to concrete communication aspects such as conveyance of positive and negative emotions (Hine et al. 2009), integrative/distributive connotation of utterances (Twitchell et al. 2013) and informational content of messages (Sokolova & Lapalme 2012) and that these differences are also prevalent early in the interaction when the conflict still develops[1]. According to these findings, communication data can indeed be used as training input for systems that learn to differentiate potential negotiation outcomes. In line with the expressed notion of evaluative statements to be of central role in judging negotiation results, this thesis seeks to employ methods of Text Classification and Sentiment Analysis in order to train e-negotiation classifiers that can eventually be applied in an NSS.

Aside from the theoretical considerations due to which Sentiment Analysis might prove a useful tool in electronic negotiation research, there are also methodological considerations

1 Further elaborations on previous research can be drawn from the studies in the body of this thesis, hence this section only briefly presents core findings for the sake of the argumentation.

regarding communication analysis and subsequent communication support: Existing means of communication analysis in negotiations are almost exclusively consisting of ex-post manual qualitative-quantitative methods, such as content analysis (Srnka & Köszegi 2007) or discourse analysis (Putnam 2005) which involves labour-intensive manual classification of transcripts and therefore is not feasible during ongoing negotiations. While there exist attempts to computerize the content analysis process in negotiations (Nastase et al. 2007), the results were not overly promising, which is mainly a direct result of existing content analysis schemes in negotiation research being too complex and diversified to be understood by a Machine Learning model with reasonable accuracy. Furthermore, these schemes mainly would yield information about what has been exchanged in the communication process – information about the contribution towards the outcome of the negotiation can only be conveyed indirectly, through known contributions of certain categories towards the outcome of the negotiation (e.g. discussed in Schoop et al. 2014).

The availability of data is another consideration to be taken into account. Drawing from several years of experimental research at the University of Hohenheim, the data sets used in this thesis are among the largest comparable sets available for complex, multi-attributive negotiations. The factual negotiation result as successful or failing has furthermore been explicitly formalized through usage of the NSS Negoisst, hence each negotiation available for final analysis has been concluded with an explicit form of agreement or abandonment of the negotiation. Hence, the opportunity to conduct predictive research on negotiation data a larger scale is almost unique.

This explicit finalization of each negotiation is also a reason, why this thesis adheres to the notion of treating negotiation result as a binomial target variable, as previous research applying data mining to negotiations did as well (Kersten & Zhang 2003, Twitchell et al. 2013). Negotiations that are perceived successful end in an explicit form of consensus, i.e. a final acceptance of the terms specified and a decision to implement said terms in the future steps. Unsuccessful or failing negotiations on the other hand do not encompass a compromise decision and result in impasse, i.e. a final rejection of the specified terms and the decision to abandon the negotiation as a whole. Whilst this most certainly is an oversimplification of reality, it at least allows to determine differences between those two extreme outcome possibilities and is in line with the overall research objective to provide a system with capabilities to detect potential negotiation failure.

2.4. Text Classification and Sentiment Analysis – Methodological Foundations

The vast amount of content generated in the World Wide Web has sparked growing interest in methods of automatic data analysis, which combined with advances in the fields of Business Intelligence and Data Mining has led to the automatic data analysis hype in practice as well as research that is commonly known under the term *Big Data*.

Especially the shift towards user-generated content with the advance of the web 2.0 paradigm and social media has opened a whole new channel of feedback and potential to apply analytic methods for companies and businesses all over the world. Businesses or internet vendors can use the data provided in order to adapt to opinions provided directly from the customer. Hence, it becomes an important corporate task to be able to aggregate and extract customer's opinions from the vast amount of data available. In this context, Sentiment Analysis was popularized as a way to achieve this task, also giving birth to a new area of research, which is now among the most important areas regarding the analysis of user generated content (Liu 2012).

The core task of Sentiment Analysis is described as to analyse "…people's opinions, sentiments, evaluations, appraisals, attitudes and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes" (Liu 2012, p.7). Hence, at the core of Sentiment Analysis (also known as *Opinion Mining*) lies any form of opinionated document that expresses an attitude of the author towards an entity of any kind. With this very generic claim, a vast variety of potential application domains has already been discussed, such as the tracking of people's opinions regarding political parties and issues, innovation diffusion, customer's opinions of products, brands and companies, financial market development, citation analysis and literary reputation in research etc. (Pang & Lee 2008, Feldman 2013). Furthermore, CMC discussions in particular have been identified as a potentially valuable area for Sentiment Analysis, where it would be possible to detect antagonistic language (Razavi et al. 2010) or monitor the general development of the discussion (Hassan et al. 2010). For negotiation data, Sentiment Analysis has rarely been explicitly discussed (aside from a brief mention in Sokolova & Szpakowicz 2007), but given the theorizations in the previous chapter, electronic negotiations in particular should prove a vital area for Sentiment Analysis application.

In general, two main goals of Sentiment Analysis can roughly be distinguished: *Summarization* and *Classification* of opinionated documents. The summarization goal focuses on providing concise aggregations of opinions expressed over different documents, in order to generate a brief overview from large corpora of data, which can then be used to extract plans of action (for example, improving a feature of a product that is generally considered negative). The classification goal is, more generally, to analyse the general polarity of an opinionated document and to automatically evaluate whether the document expresses a negative or a positive opinion. This evaluation can reach from a very general classification in positive or negative classes to a detailed listing of what exact aspects are considered good or bad – thus basically fading into a summarization. Unsupervised techniques have been employed for this task, using scoring functions based on sentiment lexica that provide a mapping between words in the document and an evaluation of their polarity. However, the majority of the approaches use a form of supervised Machine Learning methods, often combined with the notion of sentiment lexica. These methods rely on training a Machine Learning classifier on a corpus of sentiment data, in order to enable it to automatically evaluate unseen instances.

As the remainder of this thesis will have a strong focus on Sentiment Analysis using supervised learning methods drawn from Text Classification, an introduction of them methodological steps required will be given here. Detailed introductions to Sentiment Analysis are given over the course of the studies in chapters 3-6 of this thesis, so to reduce redundancy, they are omitted here in favour of Text Classification and Machine Learning processes that will only receive very brief introductions in the following studies.

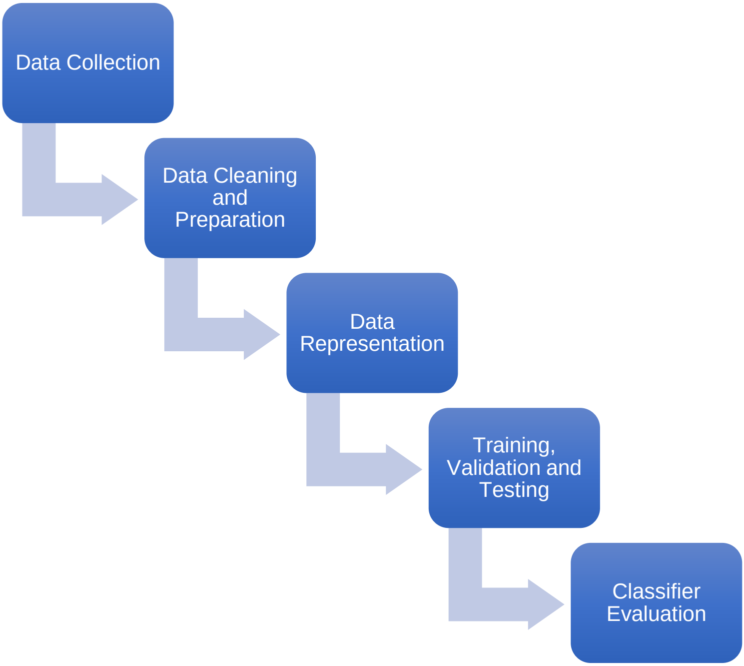

Figure 6: A model Text Classification Process

The application of Text Classification methods typically consists of a multi-step process which is shown in Figure 6. In the Data Collection step, relevant labelled data is extracted from its sources, for example by crawling web pages for reviews, transcription of audio documents or any other usual means of data sourcing. Secondly, a further preparation step which is often necessary involves cleaning the data with respect to noise such as unintentionally included HTML-Tags in the transcripts, missing values, and sometimes removal of outliers in the data set. The result of this process is usually a set of documents of the chosen classification domain, with associated labels (the terms category or class are often used synonymously) that define the class the document belongs to. These class labels either occur naturally in the context of the data set (e.g. star ratings for online product reviews) or have to be determined – most of the time by using human coders that assign labels to the documents similar to a content analytic process (e.g. Srnka & Köszegi 2007).

Now, as the transcripts are readily available, the next step is probably the most crucial one throughout the classification process: To decide on an adequate means of representation of the data set. Since most classifiers that will be used later on only accept numerical input and cannot simply process textual data, the documents have to be converted into a machine-readable form. Here, numerous variations of the *bag of words*-model are by far the most common choice. In its basic form, the *bag of words*-model views the entirety of terms present over the documents as a feature space. The documents can now be represented in this feature space in the form of a term-document-matrix M, where an entry a_{ij} denotes whether document i contains term j or not (Manning et al. 2008). For example, a simple document collection consisting of two documents D_1: “I am a cat” and D_2: “You are a cat” would result in the term-document-matrix shown in table X.

	a	am	are	cat	I	you
$\boldsymbol{D_1}$	1	1	0	1	1	0
$\boldsymbol{D_2}$	1	0	1	1	0	1

Table 1: A sample Term-Document-Matrix

It is important to note two things here. First, the columns in this example are sorted alphabetically on purpose, in order to underline that in the *bag of words*-model, any information about the ordering of the terms in the document is lost. This is a deliberate concession that this model makes in its simplest form. The second important aspect is that this example denotes presence of a term using binary information only, which omits any information about how frequent the terms occur in the document and assesses each term as equally important regarding the classification decision.

There exist several representations which are able to introduce parts of this information into the term-document matrix and which currently coexist in Text Classification. An intuitive variant is to denote counts of term occurrences instead of using a binary representation or, similarly, to use relative term frequencies in the document as a measure. However, these representations come with a downside, specifically they tend to distort the matrix in that they put an overly great emphasis on terms that occur very frequently in the corpus but that don’t carry any informative value regarding the classification goal. Such terms (e.g. “the”, “are”, “is”, “a” etc.) are commonly referred to as *stopwords* (Manning et al. 2008). Hence, measures such as *Inverse Document Frequency (IDF)* and the composite of Term Frequency and Inverse Document Frequency *TF-IDF* have been developed. The TF-IDF-Measure is given as

$$TF - IDF = tf * log \frac{N}{df}$$

where tf is the term frequency in the document multiplied by IDF which in turn is defined as the logarithm of the overall number of documents N divided by document frequency df, which denotes the relative frequency of documents a given term occurs in across the corpus. This measure has proven to be very robust regarding a purposeful representation of term counts and is an established quasi-standard in Text Classification. Nonetheless, the choice of an adequate measure to this day remains strongly dependent of the classification problem that is evaluated and oftentimes is chosen retrogradely based on classifier performance measures (Pang et al. 2002).

Aside from the actual numeric representation, it is also possible to vary the construction of the feature vectors in order to retain contextual information about sequences of terms in the document collections. This is reached via the employment of n-gram representations of the terms, where the occurrence of sequences of terms is used as matrix columns, as opposed to the example approach where only single terms (unigrams) were denoted as columns. This means, each document is tokenized into these sequences and then represented the same way as in the example. The granularity of the sequences is up to the choice of the researcher depending on how important contextual information is perceived. Usually sequences of 2 (bigrams) or 3 (trigrams) are used, since especially in Text Classification problems, unigram representations tend to generate representations that do not generalize well to unseen data points (Mayfield & Penstein-Rosé 2010). To illustrate this, see table 2 for a bigram matrix of the previous example.

	a_cat	am_a	are_a	I_am	you_are
D_1	1	1	0	1	0
D_2	1	0	1	0	1

Table 2: A sample bigram representation

Note that it is also possible and common to combine n-grams of different length into a single term-document matrix. Usually, lower-length n-grams are almost always included as well.

Whilst longer representations are obviously desirable regarding the information retained, the usage of n-grams greatly amplifies one of the most challenging problems in Text Classification, in that it causes a huge increase in the dimensionality of the feature space. For longer document collections, this space can easily extend well over hundreds of thousands of dimensions. Not only does processing these huge matrices pose a significant challenge to computing power, for longer documents these matrices also tend to only contain very few values that are not zero. This *curse of dimensionality* easily distorts classification models, since features that carry predictive value simply get lost in the complexity of the initial matrix and cannot be detected by the classifiers anymore. Lastly, this form of inflated representation also tends to be very specific regarding the data set it was generated on, which can result in overfitting issues and stop the classification models from being transferrable to other, similar data sets (Sebastiani 2002), or in the worst case to any set that is not the training set itself.

This problem directly leads to *dimensionality reduction* as the next important data preparation task. Generally, dimensionality of a feature set is reduced by either *collapsing* similar dimensions into one, or *filtering out* of irrelevant features (e.g. Forman 2008, Sebastiani 2002 coined the terms *term extraction* and *term selection* for these approaches).

Collapsing dimensions can range from simple applications like *stemming* – where all terms are reduced to their word stems, so that words that convey the same meaning but have different suffixes (e.g. “process-*ing*”, “process-*ed*”, “process-*or*”, “process”) get subsumed under one word stem or stopword removal where frequent terms that do not contribute to the classification decision (e.g. “a”, “the”, “and”) are removed from the documents – to advanced applications like Latent Semantic Indexing (Deerwester et al. 1990), attempts to cluster features of similar meanings, automatic feature construction using genetic algorithms (Mayfield & Penstein-Rosé 2007), or knowledge-engineering approaches where dictionaries are crafted to use feature categories instead of the original terms in the document. The latter approach is not uncommon in Sentiment Analysis and also subject to Study I in the remainder of the thesis.

Likewise, filtering techniques also include a broad range of possible approaches that are commonly subsumed under the name *feature selection*. Forman (2003 and 2008) discusses different variations of filtering methods, which, at their core, consist of a scoring function that estimates the predictive value a feature has regarding the class decision. Afterwards, the features can be ranked according to these scores and the features that are below a predefined threshold are excluded from further analysis. The threshold can either be defined as an absolute value of the scoring function, as an absolute number of top performing features or as a relative value (e.g. the top 5% of features). Usually applied methods of scoring include Information Gain, Odds Ratio, and Chi^2-Feature Selection which is used over the course of this thesis and is also among the most prevalent methods (Sebastiani 2002, Yang & Pedersen 1997).

Alternatively, it is also possible to apply wrapper methods, which interpret the different possible feature sets as a high dimensional search space and apply local search techniques common in artificial intelligence in order to identify the optimal feature subset (Forman 2008). Lastly, embedded methods view feature selection as an inherent property of the classifier itself, and thus only perform it implicitly. However, these approaches face some limitations when confronted with large feature spaces as they are common in Text Classification. Especially wrapper methods are notoriously scaling poorly, aside from other issues such as getting stuck in local maxima and not identifying ideal solutions (Russell and Norvig 2009). Therefore, in Text Classification, filtering methods are the most common approach in feature selection.

The next step in the classification process consists of actually training the Machine Learning model using the input data from the previous steps. For a purposeful estimation of classification performance, the existing data set is usually split into different parts. A *training set* which is used as input values for the classifier and which will be used as a foundation of the models and a *test set*, consisting of data points unseen by the trained models, which is then used to evaluate the actual performance of the classifier. This approach is necessary to avoid biasing the classifier with previously seen data during the training phase which results in an overestimation of performance and can result in strong overfitting to the training set[2]. An optional third *validation set* can be used as a test set for parameter tuning purposes. Common splitting percentages are around 66-80% of the data being used to train and the remaining 34-20% to test the models – if a validation set is used, a 70-20-10 ratio of training/test/validation sets can be used.

Oftentimes, this split means reducing a data set that does not contain overly many entries even further, which can make the entire classification task unfeasible. Fortunately, there exist compensation approaches that can be applied if only few data points are present. The main proponent of these approaches is Cross-Validation (CV). Instead of using a single training/test split, the original data set is divided into multiple splits, called folds. By far the most common amount of folds used is 10, resulting in the term 10-fold Cross-Validation. Each of these 10 folds is used as a test set for a separate classifier which is trained using the other 9 folds. The overall evaluation of the model is the assessed by using the average performance values over all of the folds. Whilst this still can lead to slight overestimations

[2] It is important to note that feature selection techniques must be performed *after* the splitting into training and test set, because otherwise information about term distribution from the test set would be leaked into the classifiers (Forman 2003).

of performances, the effect is significantly diminished and the approach allows for training a classifier when only limited data is available (Han & Kamber 2006, Witten & Frank 2005).

The classification models that are being trained can roughly be distinguished into two broad categories: *Statistical classifiers*, which seek to learn probabilities for class membership and *vector-space classifiers* which interpret data points as feature vectors in n-dimensional space and aim to place separating (hyper-) planes between feature vectors of different classes. Whilst there exist numerous classifiers in each of the categories, we will only give a brief introduction of the most common ones that are also used in the scope of the thesis (Studies II, III, and IV). The interested reader is referred to Machine Learning literature (Mitchell 1997, Han & Kamber 2006, Feldman & Sanger 2007, Manning et al. 2008 for further detailed information on the classifiers. The most widely used statistical classifier (and likely the most widely used classifier overall) is Naïve Bayes.

Naïve Bayes, as the name suggests is a comparatively simple classifier that is based on Bayes' theorem. In its basic form it maximizes the conditional probability that an instance represented by the set of terms T belongs to a class C_i out of n classes $C_1 \dots C_n$:

$$P(C_i|T) = \frac{P(T|C_i)P(C_i)}{P(T)}$$

Applying Bayes theorem as in the above equation, this probability can be estimated by maximizing $P(T|C_i)P(C_i)$, since $P(T)$ is constant over the classes. $P(C_i)$ is the a priori probability of class C_i being present, which is estimated from the class distribution in the given training data set. $P(T|C_i)$ is also estimated from the available training data by computing

$$P(T|C_i) = \prod_{j=1}^{m} P(t_k|C_i)$$

where $P(t_k|C_i)$ is simply estimated as the share of instances in the training set of class C_i containing the term t_k (Han & Kamber 2006). This approach to classifying data introduces a set of theoretically problematic assumptions, most importantly the assumption of *conditional independence* of the terms – the occurrence of a specific term in an instance does not make the occurrence of another specific term more likely. It can easily be seen that in textual data the conditional independence assumption can almost never be held up, as there exist a manifold of standard sentence construction patterns that make term occurrences dependent of each other (e.g. "there is", "I am", etc. etc.). Mainly for this assumption Naïve Bayes derives its name prefix of being "naïve" and has been under scrutiny for decades. However, its actual performance on varieties of classification problems has been surprisingly well, leading to Naïve Bayes still being the most widely used classifier. One assumption to explain Naïve Bayes performances is that even though its estimates should be inaccurate, these inaccuracies rarely change the maximization order used for the final class decision (Hand & Yu 2001). In Text Classification literature, Naïve Bayes is often used to create a classification benchmark for other classification algorithms to be compared against.

Another notable classifier that determines classification rules by scoring methods is the Decision Tree classifier. Popularized by the research of Quinlan (Quinlan 1979, Quinlan 1986), Decision Tree classifiers seek to detect the attributes that are best suited to split the training data set with regard to the class decision. In order to identify these attributes a scoring function is applied on each attribute via iteration over the attribute values (or predefined value ranges for continuous attributes) and the attribute achieving the highest score is selected from the set. A wide variety of scoring functions have been employed, such as Information Gain (as in the original ID3 version presented by Quinlan 1986), Gain ratio (used in the C4.5 extension in Quinlan 1993), or Gini index scoring. After selecting the optimal attribute, the training set is split according to the attribute values and the same scoring methods are applied to the resulting subsets, leading to further attribute splits and so on, until a stopping criterion is satisfied. Stopping criteria can include that a threshold in the scoring function (i.e. the best attribute does not provide a sufficient gain in information), if the remaining set contains no further attributes, if only a minimal number of data points remains in the subsets, or if all data points in the subset belong to the same class. Through this process, a hierarchy of splitting attributes with according value ranges is defined, which can be represented in a tree structure. Compared to other classifiers, Decision Trees usually yield results that can readily be interpreted by humans – which makes it a very popular method for exploring identified classification criteria and explaining model quality. In Text Classification and Sentiment Analysis, they have also been applied regularly in their most well-established implementation, C4.5 (Quinlan 1993). However, faced with large attribute sets, Decision Trees can experience scalability issues (Han & Kamber 2006), and are often subject to overfitting, i.e. yielding results that cannot be generalized from the data set the trees were trained on. This oftentimes makes additional tweaking of the methods such as pruning of complex subtrees or strict feature selection in advance to training the trees necessary (Witten & Frank 2005). Nonetheless, Decision Trees are known as a well-performing method in Text Classification.

The second large category of Machine Learning methods, vector-space classifiers, have seen a huge increase in popularity in Text Classification over the last 15 years, due to their ability to specifically cope with classification problems of large dimensionality. Especially *Support-Vector Machines (SVM)* (Cortes & Vapnik 1995) and *Logistic Regression* (Cox 1958) are known for strong performances in this respect.

For SVMs the data points in the training set are interpreted as high-dimensional feature vectors, i.e. an instance with n attributes represents a point in an n-dimensional decision space. The basic idea is now to find a hyperplane in this space that separates the data points belonging to different classes. If a representation of this hyperplane is found, new data points can be classified based on which side of the hyperplane they are on. Furthermore, if many separating hyperplanes are available, SVM selects the hyperplane with *maximum margin* towards any of the data points in order to ensure a maximum of generalizability of the classification and to decrease the risk of misclassifying new data points. Those data points in the training set that define said margin are the *support vectors*, of which the name of the approach is derived. Figure 7 illustrates the classification approach of SVM in an example for the two-dimensional space.

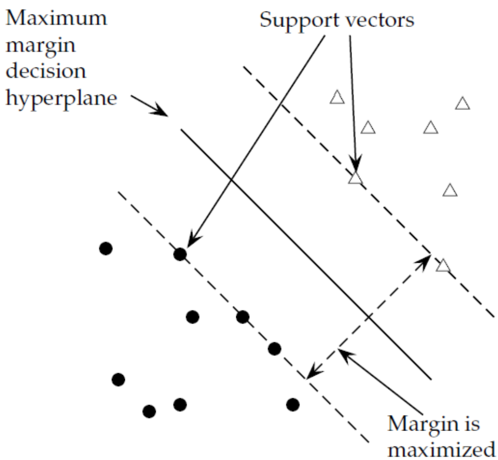

Figure 7: Support Vector classification (Source: Manning et al. 2008)

It is often the case that the data points are not linearly separable, i.e. there exists no linear hyperplane that is able to completely distinguish the different classes. To alleviate this issue, concepts such as slack variables and the kernel trick are common in SVM classification. Slack variables basically define a cost function, which enables SVM to accept a certain amount of misclassification (Manning et al. 2008). Kernel functions allow to project the data points into a higher dimensional space, where previously non-separable data may become separable and thus a decision hyperplane may be created.

SVM is known for being very well performing when confronted with a high-dimensional feature space and to have a comparatively small risk of overfitting to the training set. This makes it one of the most popular classification methods, especially for Text Classification problems. The downside of employing SVMs is that human interpretation of the model is difficult. For linear kernel functions, the weights assigned to attributes can provide some information on how important an attribute is regarding the classification decision, but for any other kernel function, even this interpretation becomes difficult, because the weights are distorted from the transformation. Hence, SVMs are difficult with regard to exploring the classification domain.

Originally stemming from statistics (Cox 1958), Logistic Regression is similar to SVM in that it seeks to identify a separating hyperplane between data points of different classes (Zhang & Oles 2001). However, instead of using a maximum-margin criterion as shown for SVM, Logistic Regression makes use of a logit function that generates a logistic curve between the data points of the different classes, the main focus being on error minimization. The weightings for the function representing the final decision plane are usually estimated using a maximum likelihood approach for class membership. It thus contains weightings similar to SVM but also gives a probabilistic interpretation of the final class decision.

In Text Classification, Logistic Regression was originally thought not to be able to perform well (Schütze et al. 1995), until techniques such as regularization of weightings were introduced into Regression approaches, after their positive impact on classification performance was confirmed in SVM approaches (Zhang & Oles 2001). Since then, Logistic Regression has been established as a well-performing technique for Text Classification, often leading to results comparable to, but slightly worse than SVM. Classifiers generated through Logistic Regression tend to yield models that contain weightings for the majority of attributes, which makes the method more sensitive to outlier values than SVM, where the final hyperplane is only described by the support vectors, with all other weightings being zero.

K-Nearest Neighbour (kNN) classifiers are a bit peculiar compared to the other classifiers presented in this section in that they do not require any model training phase at all, since the training set itself makes up the classification model, kNN is thus an example of an *instance-based* classification method. Basically, these classifiers only consist of a metric to measure distances between data points in the feature space. If a data point with unknown label is encountered, kNN measures the distances to all the data points contained in its model, and compares the class memberships of the *k nearest neighbours* found. The class assignment for the unseen data point is subsequently decided by a simple majority vote regarding the class labels of these k neighbours. Hence, it is common to use only odd values for k, so that there can be no tie between the class labels (at least in the binary classification case). Usual values for k range from 1-7, in Text Classification problems with a large amount of data points also higher values have been used (e.g. k=15 in Sokolova & Lapalme 2012).

After training the classifiers and applying the trained models to the test set, classifier performance can be evaluated based on how accurately the models labelled the data points contained in the test set. Most of the frequently used performance measures are based on the *confusion matrix*, wherein classifier label decisions and actual labels are compared.

classifier decision actual label	A	B
A	true positive	false negative
B	false positive	true negative

Table 3: Confusion matrix of a binary classifier

The example in table 3 shows a confusion matrix for a binary (i.e. 2-class) classification task with classes A and B. It simply lists all classifications from the test set comparing the actual labels with the classifier decisions, resulting in four possible cells. The convention for naming these cells stems from information retrieval where usually the two classes of “relevant” and “irrelevant” documents for a query were discussed and from where these measures originate – the derived names of true positive (tp), false positive (fp), true negative (tn) and false negative (fn) are commonly used for other problems as well, regardless of whether classes A and B represent any kind of valence or polarity. Ideally, all test examples lie on the main diagonal of the table, i.e. all examples are classified correctly. Directly derived from this table are the standard performance measures of accuracy, precision, recall and the F-measure.

Accuracy, as the name indicates simply measures the proportion of examples that have been classified correctly, i.e. it can be computed from the confusion matrix as

$$Acc = \frac{tp + tn}{tp + tn + fp + fn}$$

This measure has for years been the most common to compare classifiers with, even though it suffers from several drawbacks leading to other measures being preferred nowadays. One main problem of accuracy is that it does not represent the initial class distribution. Suppose an example where 90% of the data points belong to class A. Using a completely simplistic classifier that allocates each example to the majority class, a performance of 90% would be reached, which completely obfuscates that this classifier is not capable of detecting any example of class B whatsoever.

Hence, precision, recall and the f-measure are much more common in assessing classifier performance although accuracy is still reported and used as a very basic comparison means.

Precision and recall are defined per class, i.e. for each class a separate precision and recall score exists. These scores can be averaged to arrive at a global precision and recall measure. In the example given above, precision for class A is defined as

$$\mathrm{Pr} = \frac{tp}{tp + fp}$$

As such, precision measures the percentage of documents assigned to class A that actually belong to class A. It thus serves as an indicator of how distinctive a specific class is. If many examples that do not actually belong to the class are allocated to it by the classifier, precision drops.

Recall, conversely measures the percentage of documents belonging to a class that were assigned correctly.

$$Rec = \frac{tp}{tp + fn}$$

Recall thus is an indicator of how well documents of a specific class can be detected. If many examples of said class are misclassified into other classes, recall drops.

Most of the time, there is a trade-off situation between precision and recall, where it is possible to modify the parameters of a classifier in order to increase one of the measures at the expense of the other. Therefore, these measures should not be discussed in isolation to each other. An option to combine precision and recall into one single measurement is to compute the weighted harmonic mean of the two scores, in Machine Learning known as the F-Measure:

$$F = \frac{(\beta^2 + 1)Pr}{\beta^2 Pr + Rec}$$

where β defines whether one wants to emphasize precision or recall in the measure. $\beta = 1$ yields an equal weighting, while $\beta < 1$ weighs precision more strongly and, conversely, $\beta > 1$ favours recall.

The last basic measure that is repeatedly used in Machine Learning is the kappa statistic. Stemming from the social sciences (Cohen 1960) the kappa statistic (sometimes better known as *Cohen's κ*) was introduced as a way to measure agreement between human coders in a manual classification task as common in content analysis. In its adaption in the Machine Learning realm, it measures "agreement" between a classifier and reality. It is defined as

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$

Where p_o is the proportion of units where coders agreed (or, the classifier assigned correctly), and p_e the proportion for which chance agreement is to be expected (Cohen 1960). The peculiar thing about the kappa statistic is that it accounts for chance agreement, i.e. it is not possible to reach a high kappa score only through correct classification by chance. Consider the majority classification example from the accuracy measure – here, the kappa statistic would simply evaluate to 0. As such, the kappa statistic provides a convenient measure for classification quality, which is less biased towards a specific direction and has become an acknowledged tool in judging classification accuracy (Carletta 1996).

Aside from these basic measures a variety of other measures exist, but are not as extensively employed over the scope of this thesis. When these measures are used to evaluate classifiers employing a cross-validated design, measures of the single folds are usually averaged in order to achieve an overall estimation of classifier performance.

These theoretical introductions serve as a foundation for the application of the methods to negotiation data, as will be conducted in the following studies. Note that some aspects were skipped in this introduction such as domain dictionary creation methods or cascading of classifiers since they will be addressed in deeper details in the body of the thesis.

3. Study I: Sentiment-Based Assessment of Electronic Mixed-Motive Communication[3]

Sentiment-Based Assessment of Electronic Mixed-Motive Communication

A Comparison of Approaches

Michael Körner and Mareike Schoop
Chair of Information Systems I, University of Hohenheim, Stuttgart, Germany.
Email: Michael.koerner@wi1.uni-hohenheim.de, schoop@uni-hohenheim.de

Abstract

In this paper, we seek to analyse specific types of bilateral electronic communication processes, namely such processes where there is a distinction between individual goals of the communicating parties and their joint goals. We argue that there exists a distinction between successful and unsuccessful processes. This distinction is manifest in the communication patterns used by the participants. Sentiment analysis can enable researchers to identify these distinctions automatically, based on a classification model previously trained for the exact type of communication process. This paper discusses an adaption of sentiment-based techniques for the domain of electronic business negotiations.

Keywords: sentiment analysis, electronic communication, negotiation, mixed-motive interaction

---

3 The contents of this study are already published as: "Sentiment-Based Assessment of Electronic Mixed-Motive Communication - A Comparison of Approaches", in L. Brooks, D. Wainwright, & D. Wastell (eds.), *Proceedings of the UK Academy for Information Systems Conference,* Oxford, UK, 08.04-09.04.
The content of the published study is identical to the content presented here, the only changes applied consist of text formatting and numbering of figures and tables in order to be consistent throughout this thesis.

3.1. Motivation – Mixed-Motive Communication Processes

A mixed-motive communication process is characterized by the interplay of each participants' individual goal and all participants' joint goals (Komorita and Parks 1995). In such a scenario, parties communicate their intentions via their evaluation of the other parties' statements, as well as via disclosing pieces of information about their own intentions. Since joint goals can only be reached if all communicating parties in the end agree to a specific result of the discussion, there is an inherent difference between mixed-motive processes that are successful and those that are unsuccessful.

In an electronic scenario, where the parties do not have visual or aural access to each other (e.g. using e-mail), the role of exact language usage increases to a level, which is crucial for the success of said processes, because of the absence of other communication channels (Walther and Parks 2002, Berger 2002). Therefore, we argue that there exists a clear difference in the language (i.e. choice of words) of successful and failing interactions. This point of view is, to a degree comparable to basic assumptions of Discourse Analysis (Bavelas et al. 2002) especially to the approach that language acts as a manifestation of mental processes of the utterer, as a means to explicate individual goals, while at the same time respecting the joint goal of the interaction and the individual goals of the communication partner. If an actor in such a communication process perceives a violation of his/her individual goals, the interaction may end in disagreement and impasse.

The present paper seeks to analyse such mixed-motive communication processes through the application of techniques from Sentiment Analysis. The authors' point of view is that each turn in the course of such an interaction (in our case, written, asynchronous, electronic communication) can be seen as an opinionated document containing evaluative, polar statements about the different dimensions of the interaction process (i.e. the interaction topic, personal evaluations of the communication partner, etc.). We expect a difference in the polarity distributions between successful and failing interactions, especially in the form of a "foreshadowing" of failure. Since reasonably well-constructed Sentiment Analysis applications are used in an automated manner, this detection mechanism could enable computer systems to recognize failing interaction at an early stage, and potentially intervene in order to prevent said failure. Apart from the different polarity distribution, we expect interactions to differ in their sentiment expression with respect to "2nd order outcomes", e.g. the subjectively experienced quality of the interaction, social relationship formation process as well as the degree of trust established between the communicating parties.

As a main exponent of such communication processes, we will look into the area of negotiations, in our case Business to Business negotiations conducted asynchronously in an electronic manner using a Negotiation Support System (NSS). Therefore, we will present a brief overview on the communicational influence on negotiation outcomes, then outline details of the application of Sentiment Analysis methods before introducing four variations of sentiment assessment we applied in the course of our research. These methods, and probably solutions integrating multiple of the methods are to be evaluated using a dataset of experimental negotiations created in December 2013. The main research goals that are to be followed in the course of this paper are:

To which degree are methods of Sentiment Analysis applicable to complex communication interactions?

How are sentiment-based assessments of negotiation interactions linked to common outcome variables of negotiations, such as success or failure of the negotiation (i.e. negotiations resulting concluding in agreement with a final contract or negotiations resulting in an impasse where there is no outcome), substantive outcomes (individual and joint utilities) as well as satisfaction of the negotiators?

3.2. Analysis – Communication and Negotiation Outcomes

The influence of negotiators' communication behaviour on negotiation outcome variables is one that has been widely discussed in negotiation literature. In most cases, a sub-construct and its facilitation through communication methods are analysed such as the cognitive or the behavioural role of communication. There exists a large body of research on the affective element of communication, such as the conveyance of positive or negative emotions (e.g. Liu et al. 2010, Hines et al. 2009, Martinovski 2010).

These communicative dimensions have commonly been linked to the economic as well as the relational outcomes of the negotiation process. It has even been argued that the communicational content in early phases may have a distinctive influence on negotiation outcomes (e.g. Lewicki et al. 2010, usage of affective persuasion in Adair and Brett 2005, also Simons 1993).

Duckek (2010) developed a model that links effects of communication quality to relational as well as to substantive outcomes of the negotiation process. The model evaluates communication quality as a result of grounding, coherency of the communication process and relational communication. Applied in the context of electronic negotiations it has been shown that failing negotiations are characterised by a lowered mutual understanding between the negotiators, less friendly communication and a tendency to avoid compromises. Conversely, a higher negotiation quality results in increased satisfaction and an increased level of trust between the negotiators.

Liu et al. (2010) distinguish three negotiation communication dimensions, namely clarity, responsiveness and comfort. Clarity encompasses the negotiators' understanding of the negotiation situation, facilitated by the degree to which information is exchanged, and the resulting negotiators' ability to identify trade-offs and integrative potential in the negotiation situation. Similar to preceding research on information sharing (e.g. Adair et al. 2004) Liu et al. report that a higher level of communication clarity increases joint gains in negotiation situation as well as the satisfaction of the negotiators with the negotiation process and outcome. The second dimension identified is responsiveness, which encompasses engaging in integrative behaviour, communication of concern and more generally, communication of one's own reflection on the partner's perspective. Likewise, a higher responsiveness tends to yield higher joint gains and higher rates of satisfaction.

The last dimension, similarly linked to joint gains and satisfaction is comfort, consisting mostly of the emotional state of the negotiators and the affective communication they interchange. This dimension is especially interesting from a Sentiment Analysis point of view, as will be laid out in the following chapter. Liu et al distinctively point out the negative effects of a low-comfort situation on negotiation outcomes and negotiator satisfaction (as also argued in van Kleef 2009), consistent with previous findings such as Hines et al. (2009) who argue that displaying positive emotions can be predictive of negotiation success (see also Martinovski 2010). Also, the display of emotions such as happiness and anger can distinctively alter the negotiation partner's concession behaviour, depending for example on the general integrativeness of the negotiation task, the substantive (i.e. concession) behaviour that accompanies the displaying of emotions and whether the recipient of the emotional reaction deems it to be appropriate in the given situational context (van Kleef et al. 2004).

It is, furthermore, important to discuss the role of the affective dimension of communication when we switch from a face-to-face scenario, where the negotiators can directly see and talk synchronously to each other to an electronic situation where the negotiators merely communicate in an asynchronous dislocated, and – as is the case in this study – written manner, without the possibility to see or hear each other. There is an extensive body of discussion on how social interactions are shaped by the medium through which they are conveyed. The common course of the debate sees two opposing positions, which have been subsumed by Walther with the terms “cues filtered out” and “cues filtered in” (Walther and Parks 2002).

The “cues filtered out”-perspective is theoretically rooted in the Social Presence Theory (Short et al. 1976). The basic notion introduced by this theory is that interpersonal communication is conveyed via different communication channels. These channels can be distinguished as verbal and non-verbal channels. Whilst verbal channels convey the factual content of an utterance, non-verbal channels provide the listener with additional information, such as gestures, facial expressions, or the tone of voice. The fewer channels are available, the lower the likelihood of creating an interpersonal relationship. Communication becomes de-personalized, since the “social presence” of the individual decreases (e.g. Kiesler et al. 1984). According to this approach, electronic communication, especially in a written-only, asynchronous scenario such as the one used in this study, would not allow for the conveyance of affective communication.

However the counter-perspective, known as “cues filtered-in” argues that even though there are fewer communication channels in the notion of Short et al., the importance of the information transported via these channels increases and becomes more salient for the interpretation of an utterance. Additional ways to transmit social cues are developed and imposed on the remaining channels (such as, for example, inflectives or emoticons in internet communication). Social Information Processing Theory (Walther 1992) furthermore states that although social relationship development is more difficult in a reduced-channel scenario, it nevertheless is possible to the same degree as in a face-to-face-situation – the only factor that increases is the time needed for development.

In the context of electronic negotiations, the latter notion is for example confirmed in an exploratory manner by Griessmair and Köszegi (2009). According to their findings, emotion is carried in a less explicit manner via the asynchronous negotiation message but there is an implicit emotional layer to electronic negotiation communication which is even conveyed by factual statements. Nevertheless, the explicit linguistic manifestation of these statements remains important for their interpretation (cf. Martinovski (2010)). Finally, there are differences in the development of affective communication patterns between successful and failing negotiations, which again emphasizes the crucial role of communication for negotiation success. Electronic and face-to-face negotiations show similarities in the linguistic traits Sokolova et al. (2006).

Albeit the decision-theoretic perspective on negotiations (i.e. the factual, rational quality of offers exchanged and concessions made), communication of offers plays a crucial role concerning negotiation outcomes, identification of integrative potentials, and negotiator satisfaction with the negotiation process as well as with the negotiation outcomes.

There exist different attempts to formalize and simplify communication analysis, similar to the method described in this paper. In fact, most of the manual methods used, do exactly fulfil this task. A common example in the context of (electronic) negotiations is Content Analysis (e.g. Srnka and Köszegi 2007). Negotiators' utterances are separated into single "units of thought" and then manually classified into a predefined category scheme; the exact form of this scheme often depends on the research question that is to be answered by the analysis process. There exists an attempt to automate the process of content analysis using Machine Learning techniques (Nastase et al. 2007), but with rather unsatisfactory results, most likely due to the high amount of classes used in the classification problem.

Automatic prediction of negotiation success based on communicational content has also been tried in recent years, with varying success. Twitchell et al. (2013) manually code data from divorce negotiations into integrative and distributive speech acts. This coded data is then used to train a Machine Learning scheme to distinguish between success and failure of a negotiation, reaching an accuracy of up to 85%.

Sokolova and colleagues use a linguistic approach to analyse differences between successful and failing negotiations, first focusing on modals, pronouns, mental verbs and simple positively and negatively connotated verbs as well as expressions of negation (Sokolova and Szpakowicz 2007), and later on determining an informativeness rating for the message based on the usage of words of degree, scalars and comparatives (Sokolova and Lapalme 2012). Their findings report significant improvements in classification accuracy over the baseline.

3.3. Common Sentiment Analysis Approaches

Sentiment analysis (also referred to as Opinion Mining) has been an emerging research field during the past ten years. The aim of Sentiment Analysis is the *analysis of opinionated texts*, i.e. documents of any kind that convey the subjective opinion of the writer and are designed to be subjective and evaluative. The research field has obviously received great attention in recent years with the emergence of the Web 2.0 and the massive increase in

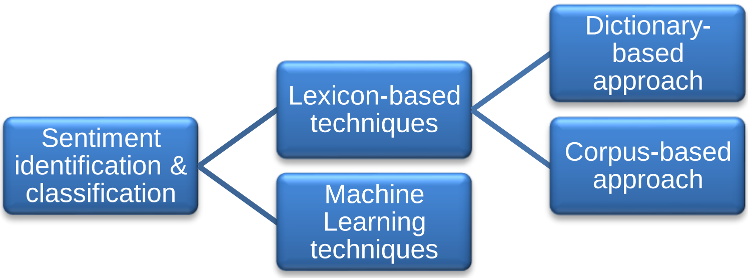

Figure 8: Elementary sentiment classification approaches

publicly available, user-generated, opinionated documents.

The original and most common domain of Sentiment Analysis is customer reviews on products in online shops such as amazon.com (e.g. Kanayama and Nasukawa 2012). In addition, Sentiment Analysis has been used in a wide range of different domains, such as movie, hotel and restaurant reviews (e.g. Ganu et al. 2013), blog posts (e.g. Zhang et al. 2009), tracking of political opinions and determination of election results (e.g. Lu and Zhai 2008), stock market development determination (Das and Chen 2007), e-mail communication (Mohammad and Yang 2011) and brand sentiment tracking via Twitter (Mostafa 2013).

One of the core tasks in Sentiment Analysis is polarity classification of texts or text fragments, commonly into the two simplified dimensions *positive* and *negative* (Liu 2012, Pang and Lee 2008). As Figure 8 shows, existing methods to conduct this task can roughly be classified into two subfields. Firstly, there are the methods that apply (and sometimes generate) a specific *Sentiment Lexicon* for the evaluation of terms and phrases occurring in the document to be classified. The generation of these sentiment lexica is typically distinguished into two approaches, the dictionary-based approach and the corpus based-approach. Secondly, there are the methods that rely on a trained Machine Learning model. Since both of these methods will be applied in the course of this paper, the following section is dedicated to explain them in further detail.

The type of method to be applied on a specific sentiment classification problem mainly depends on the granularity of the classification task, i.e. types of granularity, classification on document level, sentence level, and aspect level (Liu 2012). We will focus on classification on sentence level and on aspect level here, since we expect a negotiation document to contain a multitude of differing opinions on certain aspects of the negotiation.

Machine Learning-based techniques model sentiment classification as a typical supervised Text Classification problem. Most commonly, it is applied on sentence-level granularity. Starting from a predefined set of classes (typically positive, negative, and neutral), human coders assign these classes to a training set of sentences for the chosen domain of application. This set is then used to create a classification model using common Text Classification techniques for preparation of the dataset and for dimensionality reduction and feature vector creation such as stemming, lemmatization, stopword filtering etc. (for an extensive overview on these methods, see for example Manning et al. 2008, Feldman and Sanger 2007, Sebastiani 2002).

Apart from the supervised learning methods, some researchers use lexicon-based techniques combined with different scoring methods for sentences (Hu and Liu 2004). Lexicon-based methods rely on sentiment evaluation using and sometimes also constructing a *Sentiment lexicon*, i.e. a lexicon of words that are considered to indicate positive or negative expressions in the domain the lexicon is generated for. Before scoring or classification steps can be performed, a sentiment lexicon has thus to be constructed. Whilst there are general-purpose sentiment lexica (e.g. Hu and Liu 2004; Wiebe et al. 2005; Baccianella et al. 2010), domain-specific lexica are deemed to be better for polarity assessment accuracy.

Liu (2012) distinguishes three different approaches to create a sentiment lexicon: The corpus-based approach, the dictionary-based approach and the manual approach which, due to its labour intensity, is very rarely used alone, but rather in combination with one of the other approaches.

The *dictionary-based approach* is an automated way to generate a sentiment lexicon, which is based on synonym and antonym-searches in dictionaries. Starting from a short list of seed words with a given polarity (defined by the researcher), synonym and antonym lists are obtained from online dictionaries such as WordNet (Fellbaum 1998). Similarly, terms that tend to co-occur with seed words can be obtained through online searches (Turney and Littman 2003 use AltaVista's NEAR-Operator to perform this search). Resulting synonyms and antonyms are then assigned the respective polarity of their seed word, and are used in the next iteration. After a sufficient amount of iterations, the process stops. The benefit of this method is that no large dataset is required to construct the sentiment lexicon. However, due to the nature of the approach (i.e. – starting with a very unspecific list of seed words), the resulting sentiment lexica tend to be rather domain-unspecific.

In contrast to the dictionary-based approach, the *corpus-based approach* provides the possibility to create a sentiment lexicon with a domain-specific focus. It relies on an initial corpus of documents from the respective domain from which sentiment words are extracted. In this way, it is possible to transfer existing sentiment lexica to a specific domain (Liu 2012). The extraction of sentiment words follows a set of rules defined by the researcher. Popescu and Etzioni (2010) propose a feature-driven approach which first identifies features that are potential targets of the sentiments, and then seeks out adjectives occurring in the context of those features. A similar approach is taken by Hu and Liu (2004). Evaluation of these adjectives can be conducted manually in a human coding process or automatically using the data of existing sentiment lexica.

3.4. Application of the Approaches to Negotiation Data

We are designing and developing a program able to automatically annotate negotiation statements with respective sentiment expressions drawn from a sentiment lexicon developed for this specific context. Therefore, we used a large corpus of electronic B2B-negotiations that were conducted during the past six years in student experiments at the University of Hohenheim, using the Negotiation Support System Negoisst (Schoop et al. 2003, Schoop 2010). The complete dataset consisted of 2495 negotiation messages from 182 completed negotiations, all taken from the same experimental case, a joint venture negotiation between two companies. After we extracted the negotiation data from the respective experimental databases, a manual cleaning phase was carried out, in which we filtered negotiations that were obviously conducted in an unserious manner or – contrary to the experimental specifications – not conducted in English. The resulting negotiation messages corpus consisted of 2459 messages from 173 negotiations, of which about 75% ended successfully and about 25% failed.

In the next step, we tried to minimize the effect of the experimental case on the lexicon generation, which of course is largely attributed to aliases of the negotiators, items from the agenda or common terms that are specific for this negotiation (such as 'joint venture'). Therefore, we heuristically replaced names of persons, locations and companies with a generic tag using the Named Entity Recognition toolkit Stanford NER (Finkel et al. 2005). Additionally, a filtering list consisting of 165 terms that were subjectively assessed as being overly specific for a generalizable negotiation sentiment lexicon was created manually. In the later process feature candidates were ignored if they occurred in this list. Furthermore, we removed numerals from the negotiation texts. In the last preprocessing step, we parsed all negotiation messages using the Stanford Parser (Toutanova et al. 2003). which models linguistic relationships between two terms as *Typed Dependencies*. These *Typed Dependencies* (see de Marneffe et al. 2006 for further information) are used by our program to identify feature and sentiment candidates in the following.

The extraction of Features and Sentiments was conducted in an iterative process. First, we extracted the most frequent nouns in the corpus, to obtain an initial feature list. The minimum threshold for a noun to become a feature was experimentally decided to be 300 occurrences in the corpus. We also decided to include nouns with a direct grammatical relationship to a possessive pronoun exists into the feature set. The idea was to obtain specific terms that relate to the negotiators' actions and the negotiators' individual characteristics during the course of the negotiation (e.g. "my offer", "your behaviour" etc.). The threshold for these pronoun-noun-combinations was experimentally set to 15 occurrences.

In the next step, we expanded the feature list by synonyms of the extracted words, in order to ensure a certain degree of generalizability from the raining corpus. We used WordNet (Fellbaum 1998) and its Java-Interface JAWS to fulfil this task. However, to obtain meaningful synonym lists from WordNet, the extracted features had to be annotated with their correct word sense. Doing this in an automatic manner is a rather difficult task, and although many heuristics for automatic word sense disambiguation exist (see for example Navigli 2009), the problem itself remains unsolved. Therefore, word senses of the features were distinguished manually.

We then used the feature list to obtain a first collection of sentiment word candidates. For this, we obtained all adjectives and adverbs that were modifying words occurring in our feature list from the corpus. Furthermore, we obtained verbs from the corpus that occurred in negation constructs (such as e.g. "not accept"). The polarity assessment of the sentiment candidates was conducted using two existing sentiment dictionaries, namely those constructed by Hu & Liu (2004) and Wiebe et al. (2005). If a sentiment candidate was found in one of the two dictionaries, its polarity was set accordingly. Conflicts between the two lexica (i.e. a term has a positive polarity in one dictionary and a negative in the other one) were resolved manually.

In the second iteration, the sentiment list created in the first iteration was used to identify rare features, i.e. features that occur rarely in the corpus but in combination with common sentiment expressions. We thus obtained all dependencies between adjectives with a previously identified polarity and nouns (and adverbs and nouns respectively) and added the nouns that had not been in the feature list before.

After the two iterations, we obtained a sentiment lexicon consisting of 726 features and 762 sentiment expressions. A rather similar approach to generate a sentiment lexicon is also presented by Liu (2012). Lastly, the obtained features were manually grouped into one of seven different categories (Feature Generalization similar to Kim and Hovy 2007), in order to generalize the semantic information carried by the features.

The application of the lexicon created in this way will be done according to 3 different evaluation variations:

First, since we used the Stanford Parser to parse the messages, we want to exploit the typed dependencies, i.e. automatically identified direct grammatical relations between single terms in a sentence. Therefore, the first variation evaluates every feature-adjective-dependency where the adjective occurs in the sentiment lexicon according to its polarity. Valence shifting (Kennedy & Inkpen 2006) is performed by using negation relationships, preceding adverbs for intensification and diminishing (e.g. this is a *very* good offer" "your argument is really ridiculous" etc.) as well as adverbial modifiers, again identified via the typed dependencies the adjective occurs in.

The second variation directly operates on the parsing tree, not on the typed dependencies. Sentiment words and feature words are collected on the leaf level and then propagated upwards through the parse tree. Sentiment words are assigned to each feature they meet on a node. Similarly, negating leaves are identified and propagated, modifying the first polar expression they encounter on a node. If no sentiment word is encountered, the sentence is marked as neutral.

The third variation does not rely on parsing relationships and only operates on the part-of-speech-tags assigned to single terms in a sentence. Sentiments and feature words are identified checking each adjective and noun in the sentence. Lastly each feature obtains a polarity score based on the evaluation function given by Liu (2012):

$$score\ (a_i, s) = \sum_{sw_j \in s} \frac{sw_j.so}{dist(sw_j, a_i)}$$

with a_i being the i-th aspect (feature) in sentence s, $sw_j.so$ being the semantic orientation of sentiment word j in s – represented by +1 for a positive polarity and -1 for a negative polarity, and the denominator weighing in the distance of the sentiment word to the feature in the sentence.

In a fourth variation, we also employed a Machine Learning approach to sentiment classification, this time on sentence granularity. Our collected dataset consists of roughly 25000 single sentences of electronic negotiations. Two human coders subjectively judge those sentences as positive, negative, or as neutral in a negotiation. Based on this set of manually labelled data, we will be training a Machine Learning model using RapidMiner and its java interface for the application of the model on our experimental data. By comparing different learning models, the most accurate one can be used in the latter classification process. The initial preparation of the data consists of tokenization of the sentences, stemming and lowercasing of the terms used, the generation of uni- and bigrams from the single word tokens, calculation of TF-IDF-scores of the respective n-grams and, lastly, a feature selection process based on the information gain criterion, selecting the top 5000 n-grams to generate the final classification model. For detailed information on data preparation and word vector generation steps, see for example Manning et al. (2008). Table 4 gives a brief summarization of the different variations applied.

Variation	Outline
Typed Dependencies	Exploitation of feature-adjective-relationships identified by the Stanford Parser. Valence shifting via negation relationships Intensification and diminishing of sentiments via adverbs modifying the respective adjective
Stanford Parsing Tree	Propagation of sentiment words along the grammatical parsing tree of the sentence. Sentiment-Feature assignment when a sentiment meets a feature at a node of the tree.
Part-of-speech method	Identification of sentiments and features only by Part-of-speech-tags (i.e. all adjectives and substantives). Polarity scoring via Liu's evaluation function (2012)
Machine Learning	No sentiment lexicon used. Instead, assessment of polar sentences by human coders. ML-Classifier based on this data will label unknown sentences.

Table 4: Overview of the four variations applied

3.5. Conclusion and Outlook

In this paper, we presented our ongoing research on the application of Sentiment Analysis techniques to mixed-motive communication processes, in our case, electronic negotiations. The contributions during the course of our research in this context encompasses an adaption of a sentiment lexicon for electronic negotiation processes. Furthermore, we seek to contribute to a better understanding of communication processes in electronic negotiations, and how exactly aspects of these communication steps influence the overall result of the negotiation as well as negotiators' assessment and satisfaction with the negotiation. The gained knowledge marks a step towards pro-active communication support in the context of electronic negotiations. A system may use the discussed Sentiment Analysis techniques in an ongoing negotiation and provide feedback or act as a warning mechanism for the negotiators, when the negotiation is on the brink of failure.

Further steps include the application of the variations to experimental data gathered in an international negotiation experiment in December 2013. We seek to relate the evaluation of our methods to common negotiation outcome variables, the most obvious one being the distinction between successful and failing negotiations. In addition, we seek to focus on efficiency measures such as common substantive-level measurements (e.g. contract imbalance, joint utility - Tripp and Sondak 1992), as well as common post-negotiation assessment of the negotiators such as Quality of Communication Experience (Liu et al. 2010) and Process and Outcome satisfaction (Curhan et al. 2006).

A further challenge in the evaluation of the sentiment assessment is the question of aggregation of the sentiment data to message, and finally, negotiation level. This is mostly due to the specific type of interaction, consisting of multiple documents written over time by two different actors. We will have to compare different aggregation dimensions (e.g. all negotiation communication, communication separated by actors, etc.) as well as the exact process of aggregation, and whether the scoring results of the different variations should be integrated to obtain different perspectives on the evaluation of the negotiation process. A separation by aspect categories, as defined earlier can enhance the semantic information of the simple counting of positive/negative statements (e.g. Hu and Liu 2004). Lastly, common phase distinctions of negotiations have to be regarded. While for face-to-face negotiations, different phase models (with differing communicative characteristics) exist (e.g. Olekalns et al. 2003), these structures do not seem to be as prevalent and clear-cut for electronic negotiations (Köszegi et al. 2011).

Lastly, a limitation to be regarded is contextual influence on the negotiation situation, especially from the substantive level, i.e. the general integrativeness of the negotiation situation, power asymmetries, the quality of alternative solutions (Best Alternatives To Negotiated Agreements, or BATNAs) – all of which may contribute to a rather strategic usage of negative or positive expressions in a negotiation situation as well as a higher/lowered tolerance by the recipient for such expressions.

3.6. References

Adair, W.L. and Brett, J.M.(2005) The Negotiation Dance: Time, Culture, and Behavioral Sequences in Negotiation, Organization Science, 16(1) 33–51.

Adair, W., Brett, J., Lempereur, A., Okumura, T., Shikhirev, P. and Tinsley, C.H., Lytle, A.L.(2004) *Culture and Negotiation Strategy*, Negotiation Journal, 20 87–100.

Baccianella, S., Esuli, A. and Sebastiani, F.(2010) *SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining*, In Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10), pp. 2200–2204.

Bavelas, J.B., Kenwood, C. and Phillips, B.(2002) *Discourse Analysis*, In Handbook of Interpersonal Communication, (Eds, Knapp, M. L. and Daly, J. A.), SAGE Publications, Thousand Oaks, CA, pp. 102–130.

Curhan, J.R., Elfenbein, H.A. and Xu, H.(2006) *What Do People Value When They Negotiate? Mapping the Domain of Subjective Value in Negotiation*, Journal of Personality and Social Psychology, 91(3) 493–512.

Das, S. and Chen, M.(2007) Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web, Management Science, 53(9) 1375–1388.

De Marneffe, M.-C., Maccartney, B. and Manning, C.D.(2006) *Generating typed dependency parses from phrase structure parses*, Proceedings of the International Conference on Language Resources and Evaluation 2006 449–454.

Duckek, K.(2010) Ökonomische Relevanz von Kommunikationsqualität in elektronischen Verhandlungen, 1st Edition, Gabler, Wiesbaden.

Feldman, R. & Sanger, J.(2007) The text mining handbook: Advanced approaches in analyzing unstructured data, Cambridge University Press, Cambridge, New York.

Fellbaum, C. (ed.) (1998) WordNet: An Electronic Lexical Database, MIT Press, Cambridge, MA.

Finkel, J., Grenager, T. and Manning, C.D.(2005) *Incorporating Non-local Information into Extraction Systems by Gibbs Sampling*, Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL 2005) 2005 363–370.

Ganu, G., Kakodkar, Y. and Marian, A.(2013) Improving the quality of predictions using textual information in online user reviews, Information Systems, 38(1) 1–15.

Griessmair, M. and Köszegi, S.T.(2009) *Exploring the Cognitive-Emotional Fugue in Electronic Negotiations*, Group Decision and Negotiation Journal, 18 213–234.

Hines, M.J., Murphy, S.A., Weber, M. and Kersten, G.E.(2009) *The Role of Emotion and Language in Dyadic E-Negotiations*, Group Decision and Negotiation, 18(3) 193–211.

Hu, M. and Liu, B.(2004) *Mining and Summarizing Customer Reviews*, Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2004.

Kanayama, H. and Nasukawa, T.(2012) *Unsupervised lexicon induction for clause-level detection of evaluations*, Natural Language Engineering, 18(1) 83–107.

Kennedy, A. and Inkpen, D.(2006) *Sentiment Classification of Movie Reviews using Contextual Valence Shifters*, Computational Intelligence, 22(2) 110–125.

Kiesler, S., Siegel, J. and McGuire, T.(1984) *Social psychological aspects of computer-mediated communication*, American Psychologist, 39 1123–1134.

Kim, S.-M. and Hovy, E.(2007) *Crystal: Analyzing predictive opinions on the web*, Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning 2007.

Köszegi, S.T., Pesendorfer, E.-M. and Vetschera, R.(2011) Data-Driven Phase Analysis of E-Negotiations: An Exemplary Study of Synchronous and Asynchronous Negotiations, Group Decision and Negotiation, 20(4) 385–410.

Komorita, S.S. and Parks, C.D.(1995) *Interpersonal Relations: Mixed-Motive Interaction*, Annual Review of Psychology, 46 183–207.

Lewicki, R.J., Barry, B. & Saunders, D.M.(2010) Negotiation, 6th Edition, McGraw-Hill, New York, NY.

Liu, B.(2012) Sentiment Analysis and Opinion Mining, Morgan & Claypool, San Rafael, CA.

Liu, L.A., Chua, C.H. and Stahl, G.K.(2010) Quality of Communication Experience: Definition, Measurement, and Implications for Intercultural Negotiations, Journal of Applied Psychology, 95(3) 469–487.

Lu, Y. and Zhai, C.(2008) *Opinion Integration Through Semi-supervised Topic Modeling*, Proceedings of the International World Wide Web Conference 2008 121–130.

Manning, C.D., Raghavan, P. & Schütze, H.(2008) Introduction to information retrieval, Cambridge University Press, Cambridge.

Martinovski, B.(2010) *Emotion in Negotiation*, In Handbook of Group Decision and Negotiation, (Eds, Kilgour, D. Marc and Eden, C.), Springer, Dordrecht, pp. 65–86.

Mohammad, S. and Yang, T.(2011) *Tracking sentiment in mail: How genders differ on emotional axes*, Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis 2011 70–79.

Mostafa, M.(2013) More than words: Social networks' text mining for consumer brand sentiments, Expert Systems with Applications, 40(10) 4241–4251.

Nastase, V., Koeszegi, S. and Szpakowicz, S.(2007) *Content Analysis Through the Machine Learning Mill*, Group Decision and Negotiation, 16(4) 335–346.

Navigli, R.(2009) *Word Sense Disambiguation: A Survey*, ACM Computing Surveys, 41(2) 10:1 - 10:69.

Olekalns, M., Brett, J.M. and Weingart, L.R.(2003) *Phases, Transitions and Interruptions: Modeling Processes in Multi-Party Negotiations*, The International Journal of Conflict Management, 14(3/4) 191–211.

Pang, B. and Lee, L.(2008) *Opinion Mining and Sentiment Analysis*, Foundations and Trends in Information Retrieval(2) 1–135, viewed 4 April 2011.

Schoop, M.(2010) *Support of Complex Electronic Negotiations*, In Advances in Group Decision and Negotiation, (Eds, Kilgour, D. Marc and Eden, C.), Springer Netherlands, Dordrecht, pp. 409–423.

Schoop, M., Jertila, A. and List, T.(2003) Negoisst: a negotiation support system for electronic business-to-business negotiations in e-commerce, Data & Knowledge Engineering, 47(3) 371–401.

Sebastiani, F.(2002) *Machine learning in automated text categorization*, ACM Computing Surveys, 34(1) 1–47.

Short, J., Williams, E. & Christie, B.(1976) The social psychology of telecommunications, Wiley, London.

Simons, T.(1993) Speech Patterns and the Concept of Utility in Cognitive Maps: The Case of Integrative Bargaining, Academy of Management Journal, 36(1) 139–156.

Sokolova, M. and Lapalme, G.(2012) How Much Do We Say? Using Informativeness of Negotiation Text Records for Early Prediction of Negotiation Outcomes, Group Decision and Negotiation, 21(3) 363–379.

Sokolova, M., Shah, M. and Szpakowicz, S.(2006) *Comparative Analysis of Text Data in Successful Face-to-Face and Electronic Negotiations*, Group Decision and Negotiation, 15(2) 127–140.

Sokolova, M. and Szpakowicz, S.(2007) *Strategies and language trends in learning success and failure of negotiation*, Group Decision and Negotiation, 16(5) 469–484.

Srnka, K.J. and Köszegi, S.T.(2007) From Words to Numbers: How to transform Qualitative Data into Meaningful Quantitative Results, Schmalenbach Business Review, 59 29–57.

Toutanova, K., Klein, D., Manning, C.D. and Singer, Y.(2003) *Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network*, Proceedings of HLT-NAACL 2003 252–259.

Tripp, T.M. and Sondak, H.(1992) An Evaluation of Dependent Variables in Experimental Negotiation Studies: Impasse Rates and Pareto Efficiency, Organizational Behavior and Human Decision Processes, 51 273–295.

Turney, P.D. and Littman, M.L.(2003) *Measuring praise and criticism: Inference of Semantic Orientation from Association*, ACM Transactions on Information Systems, 21(4) 315–346.

Twitchell, D.P., Jensen, M.L., Derrick, D.C., Burgoon, J.K. and Nunamaker, J.F.(2013) *Negotiation Outcome Classification Using Language Features*, Group Decision and Negotiation, 22(1) 135–151.

van Kleef, Gerben A.(2009) *How Emotions Regulate Social Life: The Emotions as Social Information (EASI) Model*, Current Directions in Psychological Science, 18 184–188.

van Kleef, Gerben A., Dreu, C. de and Manstead, A.(2004) *The interpersonal effects of anger and happiness in negotiations*, Journal of Personality and Social Psychology, 86(1) 57–76.

Walther, J.B.(1992) Interpersonal effects in computer-mediated interaction: A relational perspective, Communication Research, 19(1) 52–90.

Walther, J.B. and Parks, M.R.(2002) *Cues Filtered Out, Cues Filtered In - Computer-Mediated Communication and Relationships*, In Handbook of interpersonal communication, (Eds, Knapp, M. L. and Daly, J. A.), SAGE Publications, Thousand Oaks, CA, pp. 529–563.

Wiebe, J., Wilson, T. and Cardie, C.(2005) *Annotating Expressions of Opinions and Emotions in Language*, Language Resources and Evaluation, 39(2-3) 165–210.

4. Study II: Feature Constructed Prediction of E-Negotiation Outcomes

Feature-Constructed Prediction of E-Negotiation Outcomes based on their Communication Content

4.1. Introduction and Motivation

Negotiation research has always recognized the value of communication in order to improve individual as well as joint outcomes. Whilst strong influences of the concession making process and the displaying of emotions have been extensively analysed, studies that focus on the (communicative) vehicle with which negotiators formulate offers and transmit their emotional state to their counterparts are rare in the current negotiation research body.

In text-based electronic negotiations, transcripts of negotiator's communication are readily available for analysis. Furthermore, they are the single remaining channel to transmit relational cues, since the negotiators do not possess aural or visual access to each other. This notion is especially interesting because messages then are one of the very few remaining ways to transmit perceptions of conflict in the negotiation situation, be it on the cognitive-substantive or on the affective-relational level. In the context of electronic Negotiation Support Systems (NSSs), the positive effects of passive, regulating and structuring communication support have been studied multiple times (Schoop 2010). The main question our research poses is whether communication support in e-negotiations can be elevated towards a more proactive level, *using* negotiators' communication data and explications of a potential conflict.

In order to facilitate such a proactive way of supporting the negotiators, such a system would have to be able to automatically assess how critical the conflicting situation in a negotiation is. Therefore, and as a first step towards supporting e-negotiators' communication proactively in ongoing negotiation situations, this paper seeks to provide an answer to the following research question:

To what degree is it possible to automatically classify negotiation message sequences according to success or failure of the negotiation?

Based on a corpus of negotiation data collected over several years, we trained a Machine Learning scheme in order to classify negotiations based on their communication content.

We used feature-constructed Text Classification incorporating knowledge from previous work in this field as well as the introduction of techniques drawn from Sentiment Analysis. Our final models show predictive power that significantly differs from the baseline.

The remainder of the paper is structured as follows: First, we give a theoretical introduction in order to introduce the domain and to identify potentially decisive elements in negotiators' communication. Afterwards, we briefly outline the methodological approach taken in the paper – Predictive Analytics – along with a detailed description of how we arrived at our specific data representation. We try to present the steps undergone to create the model as detailed as possible in order to encourage other researchers to apply these steps into similar domains of electronic interaction and to ensure reproducibility of our approach. Lastly, results of the classification process are presented and discussed.

4.2. Theoretical Background

When negotiators communicate, they exchange series of offers, counteroffers, persuade each other and exchange information, with the ultimate goals of reaching an outcome that is a) beneficial for themselves on an individual level and b) reasonably good enough a compromise for both negotiators to be acceptable (Bichler et al. 2003). A multi-attribute negotiation is thus an archetype of a mixed-motive communication situation, where individual goals of single negotiators may collide with each other and hamper the process towards the joint goal that both negotiators have (Komorita & Parks 1995). It is now the complex communication task of the negotiators to resolve this clash of goals in order to reach an agreement and to steer away from impasse. Usually this involves the solution of conflicts on the cognitive-substantive level as well as on the relational-emotional level with the main goal being to establish a framework of trust and shared understanding as a foundation of a successful negotiation process. Therefore, specific communication strategies are employed in order to reduce the cognitive and communicational complexity of the task (Te'eni 2001).

Different perspectives on relationship formation in Computer-Mediated Communication have argued that relationship formation is severely impeded if communication is reduced to asynchronous, written messages (e.g. Short et al. 1976, Lea & Spears 1992), and due to increased ambiguity, it is more difficult to reach mutual understanding (Daft et al. 1987) and thus more different to achieve agreement in a negotiation conducted via such a lean medium. Typically, these argumentations ground on the assumption that lean media provide only little opportunity to exchange relational cues, especially when visual or aural channels are unavailable. However, specifically Social Information Processing Theory

(SIP, Walther 1992) addresses these shortcomings of lean media in that it acknowledges them but introduces a temporal component: The basic argumentation is that when only few channels are available to transmit relational cues, the relative importance of the remaining cues increases drastically. Furthermore, new forms of relational cues are devised which can take explicit (via specific terms, words, symbols etc.) as well as implicit (via message structure, timing etc.) manifestations (Walther 2011). As for the ultimate goals of mutual understanding and relationship-building, SIP suggests that given enough time these goals can be reached just as well as in face-to-face situations.

In line with these argumentations, i.e. the exchange of relational cues as well as the usage of communication strategies by the negotiators, we argue that there should be substantive differences between negotiations where an agreement is reached and those that end in impasse. Recent research also explored this topic regarding the linguistic aspects of successful and failing negotiations. Sokolova and colleagues used INSPIRE (Kersten & Noronha 1999) negotiations in order to identify different linguistic markers being indicative of agreement and failure such as politeness markers (Sokolova et al. 2004), the usage of personal pronouns, modal verbs and negations (Sokolova & Szpakowicz 2005), as well as utterances hinting at exchange of factual information between the negotiators (Sokolova & Lapalme 2012). Expressions of assent and negations have also been analysed in this context (Hine et al. 2009, Huffaker et al. 2011). Twitchell et al. (2013) used similar linguistic features for the training of a classification model on divorce negotiations, but unfortunately did not elaborate on the features they found to be the most distinctive ones.

Brett et al. (2007) use Linguistic Inquiry and Word Count (LIWC) (Pennebaker et al. 2001) in order to code eBay dispute resolution data and find that usage of causal explanations, words associated with firmness, negative emotions and commands are discriminative for settlement likelihood, whilst the found no influence of positive emotion words and suggestions on the negotiation result. Having established that crucial communicational cues may be encoded in negotiators' messages, the next step is to investigate their nature. In order to do this, we use the classic separation in cues related to the substantive level and cues related to the relational level of the negotiation.

On the cognitive-substantive level, negotiators on the one hand communicate their actual offers as instances of the negotiation agenda. Positions and strategies of the negotiators are reflected in their offers and the size and frequency of their concessions. Whilst this classical decision-theoretic perspective on negotiations has been subject to extensive research, the communicational manifestations of these positions have only been rarely regarded in literature. Negotiators employ constant feedback about their satisfaction in evaluative statements regarding the offer behaviour of the counterpart (Te'eni 2001). Furthermore, they communicate in order to justify, support and frame their own offers, and provide contextual information and might even reveal their preferences in order to enable the identification of trade-offs. Lastly, negotiators can suggest these trade-offs using logrolling statements.

The second influential dimension in negotiations is on the relational-emotional level. In joint communication tasks, displays of affect play an important role in reducing communicative complexity and ambiguity in order to create a shared understanding (Te'eni 2001, Schoop et al. 2014). More specifically in negotiations, communication of emotions hence serves as an important influence of the outcome of negotiations, since it directly influences the concession behaviour of the counterpart as well as their decision to continue to negotiate or to end the negotiation in impasse. In existing research there are two perspectives on the effects of emotion communication on the actual outcome. Firstly, there is the perspective based on theories of emotional contagion (Hatfield et al. 1994) which state that negotiators tend to adapt the emotional state displayed by their counterpart and alter their behaviour accordingly. In case of positive emotions this behavioural adaption should lead to a willingness to employ cooperative and integrative negotiation strategies (Forgas 1998) as well as an increased likelihood to finish a negotiation successfully, i.e. with an agreement rather than an impasse. Accordingly, displaying negative emotions should lead to rather competitive behaviour and, in the worst case to an escalation of conflict and finally, impasse and failure of the negotiation (Forgas 1998).

However, recent research in the negotiation sector shows a picture that is not fully explicable using an emotional contagion perspective. It has been shown that negotiators that display negative emotions such as anger are perceived to have rather low limit regarding their acceptable offers. In order to avoid impasse recipients of negative reactions may be induced to greater concessions. Thus, the strategic usage of negative emotions can be beneficial regarding individual outcomes (van Kleef et al. 2004a, van Kleef et al. 2006, van Kleef & Côté 2007).

An additional factor regarding affective displays in e-negotiations is group behaviour. Theories such as the Social Identity Model of Deindividuation Effects (Lea & Spears 1992) argue that in online communication scenarios, the behaviour of the communicator is affected by deindividuation processes, which enhance group perception effects. In negotiation terms this means: If the counterpart is perceived as an in-group member, there is a tendency to react more positively to negative affective displays (Lelieveld et al. 2013) as well as a tendency towards more integrative behaviour (de Dreu et al. 2016).

Conversely, if the counterpart is perceived as an out-group member, negotiators increasingly focus on their individual gains (Rothbart & Hallmark 1988), are subject to communication biases due to negative attributions of the counterpart (Thompson & Nadler 2002) and may react to negative affect displays with retaliation and increasing of the conflict until the breakdown of the negotiations.

Interestingly, the seminal papers on displays of emotion and affect in e-negotiations rarely make statements regarding whether or not an agreement can be achieved at all (Griessmair & Köszegi 2009, Hine et al. 2009 and Huffaker et al. 2011 being rare examples of explicit comparisons of agreements with disagreement). This seems to be largely attributed to study designs and does, furthermore, not seem to be in the focus of the respective research. What seems to be general consensus is that displays of emotion and affect do have an important influence on negotiation success, although the direction of the effect remains indistinguishable without additional contextual information. Nevertheless, in communication data from e-negotiations, statements of positive or negative valence and especially their interplay with additional features of the negotiation task at hand may serve as an important indicator of negotiation success (Gibbons et al. 1992, Hine et al. 2009).

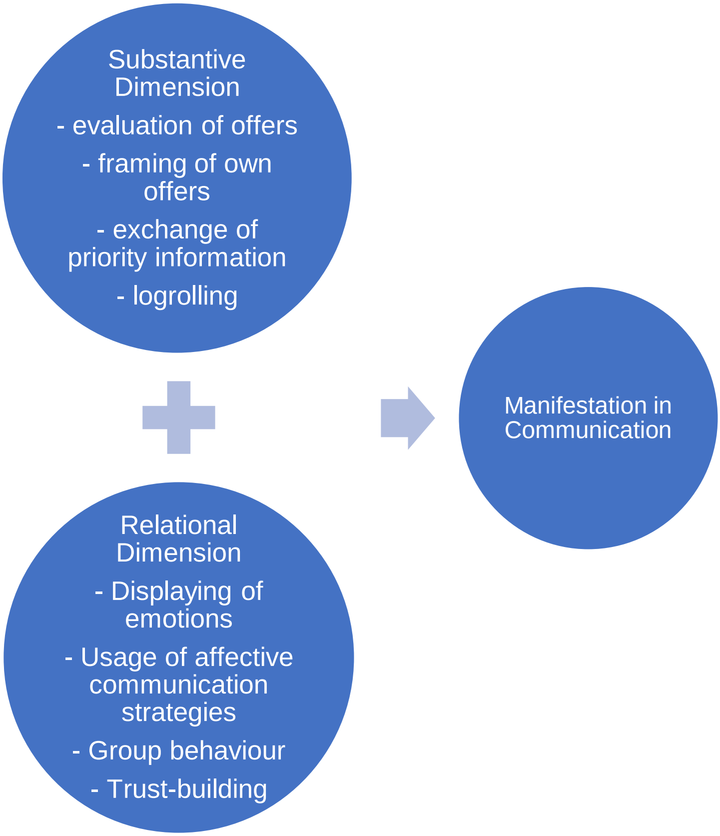

Figure 9: Manifestation of Negotiation Dimensions in Communication

All in all, we obtain a picture where evaluative statements made in negotiations are of strong importance regarding the outcome of the negotiation – an overview of which can be obtained from Figure 9. Recent advances in data analysis techniques shows a promising field with the exact focus of such evaluative statements that may carry a specific valence and may relate to an outcome category: Sentiment Analysis. Sentiment Analysis (or Opinion Mining) originated in the last ten years as a subfield of data mining and information retrieval. Facing the increasing amount of publicly available opinionated text in the web 2.0, specific Text Mining methods were devised to analyse these types of documents.

Sentiment Analysis thus became an important technique in research and practice and has been used to analyse movie and product reviews (e.g. Chaovalit & Zhou 2005, Zhung et al. 2006), track political or brand opinions (e.g. Tumasjan et al. 2010, Murthy 2015) and even online discussions (Somasundaran & Wiebe 2009). Commonly, Sentiment Analysis is used either in the classical data mining sense, to identify specific patterns of opinions in texts and to assess evaluations on different product features, or it is used in a Predictive Analytics manner, to support the training of a Machine Learning scheme and then to assess the valence or opinion specified in a given text. In the latter case, we have an instance of a Text Classification problem and sentiments expressed in the text as well as construction of features may be used in order to create an appropriate representation of a data set that provides sufficient classification accuracy. In the following, we decided to incorporate these notions in order to represent our negotiation documents, because we argue that evaluative statements used in negotiations to indicate conflict can be used in a sentiment-like manner to classify negotiation success. Liu (2012) describes these evaluative statements as either factual evaluations which may be of a specific valence or opinionated subjective statements, similarly carrying a specific valence. An example of the first case could be an evaluative statement regarding a received offer, since it does not fit one’s preferences. In the second case, an example could be any affective reaction referring to the relationship or trust between the negotiation partner; a comment on the negotiation partner’s personality etc. The following section will define how we apply an existing methodology from the realm of Predictive Analytics in IS research to our Text Classification problem in the e-negotiation domain.

4.3. Constructing a Feature-Based Classifier for E-Negotiation Messages

Viewing negotiation classification as a Predictive Analytics problem – more specifically Text Classification, we roughly follow the methodological schema for Predictive Analytics in IS given by Shmueli and Koppius (2011) in order to structure and explain our model generation process (Figure 10):

Figure 10: Methodological Steps for Predictive Analytics Processes in Information Systems (Shmueli & Koppius 2011)

Goal Definition

Our goal is to determine the (dichotomous) outcome of negotiations based on the communication exchanged by the negotiators. It is therefore a typical classification task. For simplification purposes, we for now represent the class decision as mutually exclusive, meaning there is no probability distribution for documents to belong to a specific class but rather a hard class decision.

Data Collection & Preparation

The data set we use stems from a series of negotiation experiments conducted at the University of Hohenheim using the Negotiation Support System Negoisst (Schoop et al. 2003, Schoop 2010). All negotiation data stems from bilateral negotiations, where negotiators exchanged a series of messages to negotiate according to a specific case study over multiple days in a strictly alternating fashion. Additionally, negotiators had decision supporting means available based on a linear-additive preference model, showing them the quality of offers exchanged according to their preferences as a value between 0 and 100. All in all, the negotiation data set amounts up to 646 negotiations of which about 82% were successful and 18% were failing. The average length of a negotiation document is 2299 words. The corpus we obtained is not particularly large for a Predictive Analytics task, but for the negotiation domain it is one of the largest collections available. Furthermore, different negotiation cases were used over the data set, so the corpus should be sufficient to avoid model biases due to overfitting to a single case.

Nevertheless, during the data preparation phase, we applied Named Entity Recognition using Stanford NER (Finkel et al. 2005) to the data set, thus unifying negotiators' aliases as well as names of persons, locations and companies, in order to rule out other potential side-effects. Negotiations where the outcome was not defined e.g. due to negotiators not finishing the experiment in time were omitted from further analysis.

Lastly, since the final goal is to achieve adequate classification quality in ongoing negotiation, we generated three representations of our negotiation data. The first representation contains the complete negotiation messages exchanged, the second representation consists of the first half of the negotiation data and the third representation represents the first three quarters of textual data. This allows us to draw conclusions on how the classification quality will develop when a complete set of negotiation data is unavailable.

Exploratory Data Analysis, Data Set Representation & Feature Construction

The crucial step in our classification task is the choice of variables to include into the classification process. In typical Text Classification processes, a term-document matrix is generated from n-grams in the documents. It records the frequencies of words or n-grams in each document. Due to the massive variety to express oneself in language, these plain bag-of-words-representations – which were argued to provide a suitable means to represent negotiation data (Sokolova et al. 2004) – tend to yield very sparsely populated high-dimensional matrices, which are difficult to be processed by common Machine Learning schemes since the probability of distortion or blurring effects is very high (Forman 2008). Therefore, various means of dimensionality reduction can be applied, which are subsumed under the term “Feature Selection”. Usually, three selection types are distinguished: Filter methods, wrapper methods and embedded methods (Sebastiani 2002, Forman 2008), all of which rely on statistical/mathematical techniques in order to determine an ideal subset of features that excel at characterizing the specificities and differences of each class. Note that these methods do not introduce theoretical considerations, but are purely of mathematical nature.

Alternately it is possible to construct feature sets using abstract classes and dictionaries that assign specific features to these classes (Sebastiani 2002). This way, abstraction from simple words is possible, while at the same time retaining a maximum of information contained in the messages. Again, automated methods of clustering such as Latent Dirichlet Allocation (LDA, Blei et al. 2003) can be applied here as well as manual construction of sets according to theoretic knowledge in order to incorporate domain information (Guyon & Elisseeff 2003). The latter is the approach conducted in the course of this paper – automated topic modelling on negotiation messages is difficult to apply because of two main reasons. Firstly, the topic models get massively distorted by terms that are specific for a given negotiation case, such as the names of the issues or common lines of argumentation. Secondly, the algorithms have difficulties to extract topics of meaning at all, since typically all negotiation documents address a lot (in fact, almost all) potential topics that can come up in the course of a negotiation.

Furthermore, for negotiation data it has been shown that retaining the semantic domain of the words used improves classification results over simple frequency-based approaches (Sokolova et al. 2005, Sokolova & Szpakowicz 2005).

In Sentiment Analysis, the term *Feature* is typically used in a different manner from common Text Classification literature. Here, a *Feature* is viewed as a potential target of sentiment expressions, a specific *aspect* of the item/topic that is evaluated by sentiment statements. Usually, these aspects can be extracted from a corpus of training data via the collection of noun words that occur frequently (Liu 2012). This is the first of two notions from Sentiment Analysis that is used in our research. We started with this approach on a subset of our corpus and identified 726 frequent nouns as potential aspect words.

In order to reduce the dimensionality of our data set and to increase the informativeness of single features, we used this set of frequent nouns as a foundation for the creation of semantic categories (or aspect categories, as referred to by Liu 2012). This form of feature construction is used to obtain a more concise and aggregated representation of our data set, which can support classification accuracy (Riloff et al. 2006). Similar to feature construction/subsumption approaches from Sentiment Analysis (e.g. by Riloff et al. 2006, Zaidan et al. 2007), or qualitative research (and by Twitchell et al. 2013 for their classification model), for example in the context of Grounded Theory (Corbin & Strauss 1990) we performed open coding on the feature nouns in order to generate the categories to group them into arriving at a simplified knowledge-based domain representation (Guyon & Elisseeff 2003). The semantic categories that we identified and that are henceforth used as our main features of the negotiation are denoted in table 5. Note that these feature categories are not explicitly neutral in their valence. Especially the categories Problemsfeature and Integrativefeature may carry predefined implicit positive and negative volition. This is intentionally so; we tried to capture these types of features with the categories in order to separate inherently neutral aspects of the negotiation from aspects that may already indicate success or failure (analogous to the notion of Zhang & Liu 2011).

The second notion of Sentiment Analysis is the application of a sentiment lexicon in order to distinguish utterances of positive valence from utterances of negative valence. Hence, we used the frequent noun list created in the step beforehand to generate a sentiment lexicon on our negotiation corpus in an alternating process described in deeper detail in Körner & Schoop (2014). Our initial feature list was extended with features that predominantly occurred in the context of possessive pronouns (e.g. “our relationship”, “your offer”) since we believe these to be recurring targets of sentiment expressions. In two iterations frequent sentiment words related to these features were identified using existing sentiment lexica (Hu & Liu 2004 and Wiebe et al. 2005) and those again were used to expand the feature list, resulting in 726 features and 762 sentiments (Körner & Schoop 2014). Lastly, we assigned semantic categories for sentiment expressions of positive and negative valence, separating adjectives from verbs of positive and negative volition (as is also done in Sokolova & Szpakowicz 2005).

We took a multitude of additional variables to be represented as semantic categories under consideration as well. Intensifying adverbs are commonly used in Sentiment Analysis in order to modify valences of sentiment expressions on a more detailed granularity (Polanyi & Zaenen 2006). Statements of a certain valence represent a stronger sentiment when accompanied by an intensification adverb such as “very” or “really”. Furthermore, the usage of intensifications is indicative of a higher verbal immediacy which is supposed to represent more powerful language styles (Gibbons et al. 1992). Similarly, indications of firmness in the language of negotiators has been associated with settlement likelihood (Brett et al.

2007). Therefore, we compiled a list of the most common intensification words in the English language extracted from Kennedy (2003) and included them as an additional semantic category.

The usage of negation expressions in has similarly been assessed by negotiation as well as by Sentiment Analysis literature. In Sentiment Analysis, negation expressions fulfil a modifying role ("valence shifters") when co-occurring with sentiment expressions (Kennedy & Inkpen 2006). Positive sentiments co-occurring with negations become negative sentiments and vice versa. The inclusion of negation modifiers into the analysis is, therefore, in general supposed to be supportive to analysis quality and accuracy (Polanyi & Zaenen 2006). In negotiation research, negation usage has been analysed multiple times with mixed results. Whilst Sokolova & Szpakowicz (2005) found negation statements non-influential for negotiation outcomes, Hine et al. (2009) report significant impacts of negations in the last half of the negotiation. Due to their double-role regarding the improvement of Sentiment Analysis results as well as their somewhat unclear status in the negotiation field, we included a category containing statements of negation. For this purpose, we used the list of negations from Linguistic Inquiry and Word Count (Pennebaker et al. 2015).

Another notion from Sentiment Analysis that we took under consideration is that of an "opinion holder" (Liu 2012), i.e. the person that expresses an opinion in a given document. This notion is interesting for example in texts, where at first opinions of other persons are recapitulated and afterwards an own, original opinionated statement on a given topic is made. Including the notion of different opinion holders seems purposeful at well for the negotiation domain, since multiple opinions may be exchanged and recapitulated, and also opinions held by parties external to the core negotiation may be discussed since they are influential to the negotiation outcome. Therefore, we included personal and possessive pronouns into our analysis. We split the personal and possessive pronouns according to their potential opinion holders into multiple semantic categories as can be seen in table 5. Note that this also may capture potential group behaviour as known from negotiation theory as discussed in chapter 3.2.

Especially in combination with personal pronouns, modal verbs that indicate obligations or commands such as "must", "will" etc. have also been suggested as being potentially of distinctive nature in negotiations (Sokolova & Szpakowicz 2005, Brett et al. 2007). According to the suggestions given in the respective studies, we constructed a category of modal verbs indicating obligation as well.

Lastly, it would be interesting to take a look into logrolling statements, which would be an interesting manifestation of decisional conflict management behaviour and has been found as distinctive for integrative negotiations multiple times (e.g. Thompson 1990, Harinck & Druckman 2017). However, it is difficult to capture an idealized logrolling statement using term lists, since there are abundant formulating possibilities in natural language that allow for logrolling statements. These formulations rarely follow a specific pattern that can be grasped with concrete words such as “IF you give me X, THEN you can have Y of me”. For this rather practical reason, logrolling statements were omitted from further analysis, although it would surely be interesting to look into different verbal manifestations of logrolling statements in order to capture them as semantic categories.

Table 5 contains an integrated overview of all the features that were identified and discussed in the previous paragraphs. We distinguish two main levels of feature categories, semantic and syntactic feature levels. The difference between these groups is that while syntactic features have an explicit, linguistically defined scope (such as pronouns or intensifiers), semantic categories encompass domain-specific knowledge, which is inherently incorporated into the categories themselves.

Set of Words	Respective Category
Semantic Level	
Feature relating to offer Exchange	Discussionfeature
Features relating to agenda and issues	Issuesfeature
Features relating to individual goals and demands	Demandsfeature
Features relating to joint goals and compromises	Integrativefeature
Features relating to Problems and Errors during the process	Problemsfeature
Features relating to expressions of internal feelings	Internalfeature
Features relating to the relation- and partnership between the negotiators	Relationshipfeature
Positive Sentiment Adjective	Posdictsentiment
Negative Sentiment Adjective	Negdictsentiment
Verbs of positive volition	Posvolverb
Verbs of negative volition	Negvolverb
Syntactic Level	
Adverbs of intensification	Intensificationmodifier
Expressions of negation	Negationmodifier
Personal Pronoun 1st person singular	PerspronI
Personal Pronoun 2nd person singular/plural	PerspronYou
Personal Pronoun 3rd person singular	PerspronIt
Personal Pronoun 1st person plural	PerspronWe
Possessive Pronoun 1st person singular	PosspronMy
Possessive Pronoun 2nd person singular/plural	PosspronYour
Possessive Pronoun 1st person plural	PosspronOur
Modal Verbs of obligation	ObligationModals

Table 5: Feature Category Overview

Variable Selection

These constructed semantic categories are now used to represent our negotiation data set – the terms in the negotiation documents are simply replaced by pseudo-n-grams of our semantic categories. Other words are omitted from the classification process in order to not distort the classification and to obtain semantically meaningful classification results. This document representation using only semantic categories is now used in a classical text-classification fashion: Using RapidMiner 5.3, we created a uni- and bigram representation of the documents using term frequencies[4]. Then we employed a Chi² automated feature selection (Sebastiani 2002), reducing the data set to the top 150 discriminative n-grams. Based on this representation method we trained several Naïve Bayes classifiers using a holdout set in order to determine, which assembly of semantic categories would potentially yield the best classification results. In order to obtain a reasonably good representation, we used a greedy approach to search in the potential combinations of semantic categories: Starting with our basic domain representation using the compiled domain features derived from the frequent substantives, we then combined this representation with each of the remaining categories iteration-wise. The best-performing category – that also provided an overall increase in classification quality would then be added permanently to our category set. Then the next iteration the remaining categories would be added one by one and so on until no category addition could bring an increase in performance anymore.

Furthermore, we found that several bigrams had impacts on the classification process that were semantically completely meaningless. We, therefore, devised several rules for bigram filtering in order to omit them from the representation of the data set and thus from the classification process. These rules can be obtained from table 6.

Filtering rules for bigrams
1. Two successive feature categories must never form a bigram
2. Intensification words may not form a bigram with a pronoun
3. Two successive negations may never form a bigram
4. Two successive pronouns may never form a bigram

Table 6: Filtering Rules for Bigrams

4.4. Results

As stated above, our negotiation data consists of approximately 82% successful and 18% failing negotiations. Regarding our classification goal, this is a rather large and distorting skew in the data sets. For purposes of generating and evaluating a model we therefore decided to generate a balanced data set with equal distributions between the two classes

---

4 Other representations, such as binary term occurrences or a TF-IDF-representation were also tried, but not performing as well as a pure term frequency-approach

as has been suggested as a remedy for this kind of skew effects in Mao & Lebanon (2006). We retained all failing negotiations and randomly selected an equivalent amount of successful negotiations from our original set.

Interestingly, our category selection process described in the last section developed differently for our three data sets. Table 7 shows the categories that were removed for half, three quarters, and full negotiations.

First Half	Three Quarters	Full Negotiations
Positive Sentiment Adjectives	Positive Sentiment Adjectives	Negative Sentiment Adjectives
Verbs of negative Volition	Modal Verbs of Obligation	Modal Verbs of Obligation
Intensifications	Possessive Pronouns	Personal Pronouns
		Possessive Pronouns

Table 7: Category Filtering During the Model Training Phase

Especially the contribution of our sentiment categories to classification quality is noteworthy. Whilst we found little direct contribution of positive sentiments for half and three quarter negotiations, for full negotiations positive sentiments were found to be contributing towards classification quality. Instead, negative sentiment adjectives were somewhat surprisingly dropped from the selection process. It seems that the classifier posed greater importance on negative utterances during the beginning of the negotiation process instead of the end. This picture is however not completely consistent, since verbs of negative volition had also been dropped for half negotiation classification. Viewing positive sentiments as expressions of positive emotions, this behaviour is consistent with the findings of Hine et al. (2008) who found no differences between successful and unsuccessful negotiations in positive emotion expressions *in the first half* of the negotiation process. In the last half of the negotiation, they found positive emotions to contribute significantly towards negotiation success. For negative sentiment expressions our results differ from Hine et al. (2008) who found similar behaviour for negative emotions than for positive emotions. It also seems that contextual information as expressed with personal and possessive pronouns decline in their contribution to classifier performance for full negotiations. For incomplete negotiation data, personal and possessive pronouns were still contributing. These results, again, are similar to the seminal research by Sokolova & Lapalme (2012) who reported only slight to no increases in accuracy with the inclusion of personal pronouns. However, in our case, inclusion of pronouns did improve classification quality.

Our classification results were obtained applying 10-fold Cross-Validation on a balanced subset of 234 negotiations. The baseline scores for classification are those obtained using a ZeroR classifier that marks all negotiation documents as successful (Sebastiani 2002). We applied several of the most commonly used classifier instances in RapidMiner in order to cross-check our results and determine the classifier working best. Classifiers chosen were Naïve Bayes (Maron & Kuhns 1960, Maron 1961), k-Nearest Neighbour (Cover & Hart 1967), Logistic Regression (Cox 1958) (using the myKLR implementation available in RapidMiner), Neural Networks, Support-Vector Machines (SVM, Cortes & Vapnik 1995) and Decision Trees (Quinlan 1986). The parameters of the classifiers were chosen as follows: For the k-Nearest-Neighbour-Classifier, we modified k stepwise from 1 to 13. The distance metrics we tried were Euclidean distance, Chebychev distance and cosine similarity because of its natural appropriateness in document classification problems (Manning et al. 2008). Logistic regression and Support Vector Machine results were obtained using a Radial Basis Function (RBF)-kernel, Decision Trees were used with Information Gain (IG) as node splitting criterion. We performed grid-based parameter tuning for Logistic Regression, Neural Network, SVM and Decision Trees, resulting in different parameter sets for our three groups of negotiation data. These parameter sets are listed in table 8.

Classifier	Parameters First Half	Parameters Three Quarters	Parameters Full
k-Nearest Neighbour	k:3 distance measure: cosine similarity	k: 13 distance measure: euclidean distance	k: 11 distance measure: chebychev distance
Logistic Regression	C = 2 γ = 0.00048828125	C = 2048 γ = 0.00048828125	C = 0.125 γ = 0.0078125
Neural Network	Learning Rate = 0.1 Momentum = 0.58	Learning Rate = 0.26 Momentum = 0.26	Learning Rate = 0.9 Momentum = 0.1
Support Vector Machine	C = 2 γ = 0.0078125	C = 128 γ = 0.00048828125	C = 0.5 γ = 0.001953125
Decision Tree	Criterion: gain ratio Minimum split size = 4 Minimum leaf size = 4 Minimum gain = 0.001 Confidence = 0.1	Criterion: gain ratio Minimum split size = 4 Minimum leaf size = 2 Minimum gain = 0.001 Confidence = 0.1	Criterion: information gain Minimum split size = 8 Minimum leaf size = 2 Minimum gain = 0.001 Confidence = 0.1

Table 8: Classifier Parameters for Classification

Table 9 shows an overview on the classifiers' performance measures on the different data sets. We report classification accuracy, weighted mean precision, weighted mean recall and F-Score (β = 1). Best performances for each data set and measure are indicated in bold – significant improvements over the baseline (tests are discussed in the following section) are marked with * next to the classifier name.

Since the data set contains a 50-50 distribution of successful negotiations, the baseline scores for a trivial classification (all negotiations are classified as successful) are an accuracy and precision of 50%. Overall, most of the performances obtained from classification show an improvement over the baseline scores. Regarding full negotiation samples, there are 3 classifiers outperforming the others, namely SVM, Logistic Regression and Naïve Bayes, with SVM being the top performing classifier, similar to Sokolova & Lapalme (2012). The absolute performance values are somewhat lower than the results reported there as well as by Twitchell et al. (2013), however, it is difficult to actually compare the results because of the differing baselines (0.55 is the accuracy baseline in the Sokolova & Lapalme and 0.60 in the Twitchell et al study compared to our 0.5), it is likely that the classifiers constructed via our method perform slightly worse.

Half Negotiations	**Naïve Bayes**	**kNN**	**Logistic Regression**	**Neural Network**	**SVM***	**Decision Tree**
Accuracy	60.18	51.05	58.44	56.34	**60.54**	51.58
Precision	60.83	50.50	59.98	56.37	**61.66**	53.73
Recall	60.08	50.91	58.59	56.28	**60.41**	52.05
F-Score	60.45	50.70	59.28	56.32	**61.02**	52.88
Kappa	0.202	0.042	0.170	0.126	**0.209**	0.042
Three Quarters	Naïve Bayes*	kNN	Logistic Regression*	Neural Network*	SVM	Decision Tree*
Accuracy	**65.65**	53.17	64.40	63.15	63.53	61.38
Precision	**66.52**	53.47	64.71	63.35	63.57	65.57
Recall	**65.68**	53.22	64.48	63.15	63.66	61.58
F-Score	**66.10**	53.34	64.59	63.25	63.61	63.51
Kappa	**0.313**	0.064	0.289	0.263	0.273	0.230
Full Data	Naïve Bayes*	kNN	Logistic Regression*	Neural Network	SVM*	Decision Tree*
Accuracy	68.62	64.04	69.08	65.29	**69.91**	63.93
Precision	69.03	64.58	70.39	65.35	**71.79**	64.47
Recall	68.68	64.20	69.25	65.24	**70.05**	63.92
F-Score	68.85	64.39	69.82	65.29	**70.91**	64.19
Kappa	0.373	0.283	0.384	0.305	**0.400**	0.279

Table 9: Classification Result Overview

A further difference is that our various classification models perform less steady on the data set, with kNN (which is also among the top performing methods in Sokolova & Lapalme), Neural Nets and Decision Trees falling short of the classification quality that Naïve Bayes, SVM and Logistic Regression show. This may be a direct result of our comparatively low sample size for a classification task of that complexity. In order to test for an actual improvement over baseline performance despite the variances in the classification performances over the single folds, we applied a Friedman test on the F-Score performances over the folds as exemplified in Japkowicz & Shah (2011). Assuming a significance level of $p < 0.05$, for all three data sets (half, three quarters, and full), the results indicated a significant difference in performance between the classifiers, giving $\chi^2(6) = 23.815$, *p = 0.001* for full negotiations, $\chi^2(6) = 29.521$, *p = 0.000048* for the three-quarter set and $\chi^2(6) = 15.328$, *p = 0.*018 for the classifiers on half of the negotiation transcripts.

Therefore, for all three cases we conducted Wilcoxon signed-rank tests in post hoc analysis, each time comparing the classifiers to baseline performance. Hence, Bonferroni correction lowered the significance level to $p < 0.0083$. For the full negotiation data set, we detected significant performance differences between the baseline and Naïve Bayes *($Z = -2.803$, $p = 0.0051$),* Logistic Regression *($Z = -2.701$, $p = 0.0069$), SVM ($Z = -2.803$, $p = 0.0051$)* and Decision Tree *($Z = -2.803$, $p = 0.0051$),* Neural Networks and kNN yielded no significant improvement over the baseline, showing that their performance was too erratic over the folds. For three quarters of the negotiation data we detected significant performance differences between the baseline and Naïve Bayes *($Z = -2.701$, $p = 0.0069$),* Logistic Regression *($Z = -2.810$, $p = 0.0049$),* Neural Networks *($Z = -2.805$, $p = 0.0050$)* and Decision Tree *($Z = -2.805$, $p = 0.0050$),* whilst kNN and SVM failed to yield significant differences in performance. For half negotiation transcripts, classification performance became substantially more erratic, with only SVM *($Z = -2.703$, $p = 0.0069$)* performing significantly different from the baseline, whilst all other classifiers showed too much variance over the folds to provide a consistent and significant improvement over the baseline.

Interestingly, the development of the classifiers when the available negotiation data is reduced is declining stronger than in comparable publications, especially kNN and Decision Tree suffered strongly from the reduction – in fact for half negotiations they lose almost all of their predictive ability – whilst Naïve Bayes, SVM and Regression performed better on average, but for all classifiers except SVM, classification behaviour became too erratic facing a strong reduction in transcript length. Still, the performance of SVM on half negotiation shows that some predictive ability was retained, although the results are not as encouraging as in Twitchell et al. (2013) and Sokolova & Lapalme (2012).

Seeking for a more concise understanding of the classification decisions, an interpretation of the resulting models can be performed. This is especially important in order to counter-check whether the classifiers actually extracted purposeful knowledge. This process, however, tends to have a subjective bias since it is less guided by quantitative measures (Vellido et al. 2012). Furthermore, some of the outputs are too complex to be interpreted as is. Whilst it is theoretically possible to investigate the support vectors and weightings of SVM and Logistic Regression, in this case this is not meaningful since we used a radial kernel – which makes the relation between weightings of features and their actual distance to the separating hyperplane non-trivial, i.e. a higher weight does not necessarily mean a better distinctive quality of a feature for the classification decision (van Belle & Lisboa 2013). Nevertheless, it is possible to compare some of the classifier in- and outputs against previously established theoretical knowledge. First of all, we therefore decided to check the outputs of the chi² feature selection process. Table 10 shows an overview of the top 20 features used to classify full negotiations. The scores are normalized on a 0-1 scale and they do not contain information about the direction – i.e. all features that contribute to the class decision have a positive value, regardless of whether they are indicative of negotiation *success* or negotiation *failure.*

Feature n-gram	Weighting
discussionfeature_negationmodifier	0.987
negationmodifier	0.899
posdictsentiment_discussionfeature	0.588
negvolverb_discussionfeature	0.542
demandsfeature_intensificationmodifier	0.540
negationmodifier_integrativefeature	0.515
negationmodifier_issuesfeature	0.514
negvolverb	0.474
integrativefeature_negationmodifier	0.470
negationmodifier_demandsfeature	0.435
negationmodifier_relationshipfeature	0.418
intensificationmodifier_intensificationmodifier	0.398
problemsfeature_negationmodifier	0.378
posdictsentiment_posdictsentiment	0.375
problemsfeature	0.371
problemsfeature_posdictsentiment	0.356
internalfeature_negationmodifier	0.353
issuesfeature_posdictsentiment	0.336
intensificationmodifier	0.316
negationmodifier_posdictsentiment	0.307

Table 10: Distinctive Features According to the Chi²-Method

The weightings show a strong reliance on the usage of negations during the negotiation process. Again, these results are consistent with what Hine et al. (2008) reported, who identify negations as being especially discriminative during the last half of the negotiation. Regarding the semantic features we constructed the weightings suggest a reliance on utterances related to the discussion process and offer exchange ("discussionfeature") which are part of 3 of the 4 top n-grams, once paired with negations and twice with evaluative utterances, one positive and one negative. This suggests an important role of explicit statements on the quality of the offer exchange as a predictor for success or failure of the negotiation.

Another point of view emerges when looking at the Decision Tree model for full negotiations. The Decision Tree produces an output model which – compared to the other classifiers – is rather easy to interpret and understand, with the most distinctive features being placed closer to the root of the tree and the features containing less information for the class decision being placed near the leaves (or being omitted from the model altogether). Figure 11 shows the Decision Tree that was generated by RapidMiner. Consistently with the chi² weightings, the Decision Tree model poses great importance on negations, again not only occurring in context, but also appearing without context. Somewhat surprisingly, the aforementioned discussion feature does not play an important role in the classification process at all. It only is used as a measure occurring as a unigram in the middle branch of the tree. This effect might be due to the differing evaluation criteria between the chi² selection and the Decision Tree model generation. Furthermore, the chi² selection does not accurately account for the effect of a small amount of data points. So while it might be that the presence of the bigram “discussionfeature_negationmodifier” accurately detects one of the two classes, this is most likely true *only for a few negotiations* – the Decision Tree most likely excluded the feature because it did not provide much information concerning *the whole dataset* of negotiations.

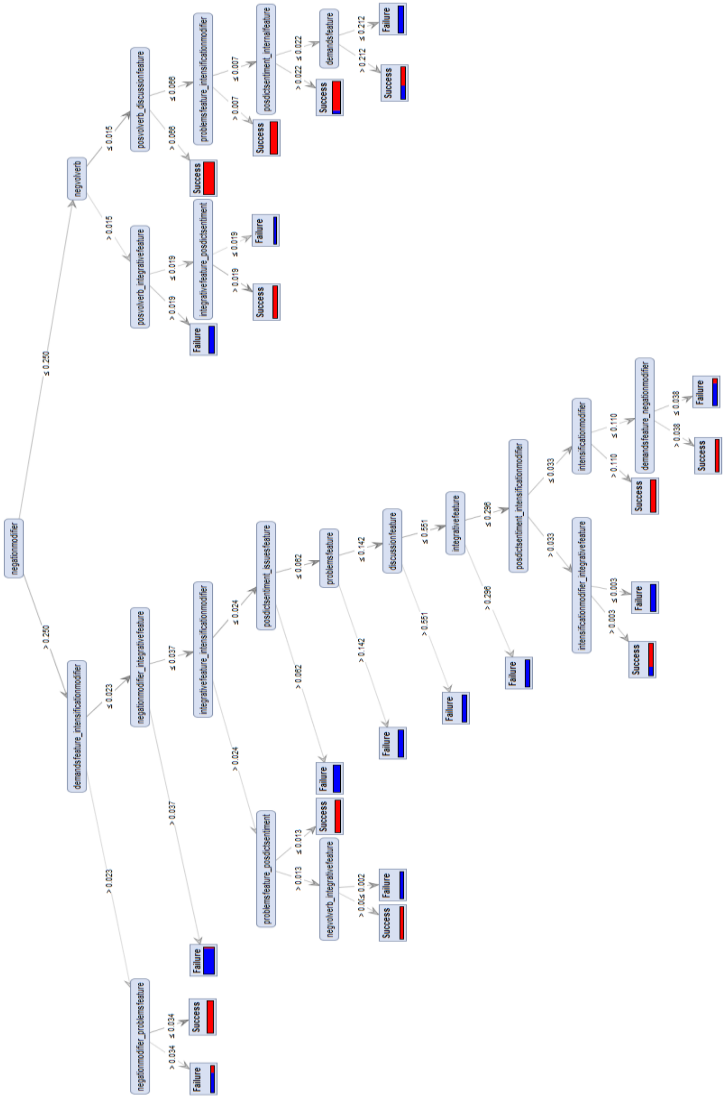

Figure 11: Decision Tree for Complete Negotiations

All in all, more than half of the top 20 bigrams according to chi² are not contained in the tree which retrospectively seems to justify our selection of a large percentage of the overall features created in order to account for these diversity effects. This in fact supports the notion, that the classification decision for negotiation success indeed is a complex one which calls for a rather complex and granular representation (without leaving the frame of what a Machine Learning classifier can actually achieve). On a further theoretical note, most of the bigrams used by the classifiers consist of combinations between our previously defined semantic level features and syntactic level features – which is again consistent with the expectation that the categories on the syntactic level modify and differentiate our semantic level categories and thus introduce novel information regarding the classification decision.

4.5. Conclusion and Outlook

In this paper we used Sentiment Analysis techniques in order to construct a feature-based Text Classification scheme that can reasonably well distinguish between successful and failing negotiations.

The theoretical contributions of our paper are twofold. Firstly, we sought to add to the existing knowledge via our description of the process with which our classification model was created, especially for a complex communication domain such as electronic negotiations. We acknowledge that it is difficult to assess the “body of theory” in research doing Predictive Analytics (Markus 2014) and therefore propose our classification model and the process of the model creation as a legitimate theoretical artefact. We tried to be as accurate as feasible in the description of our process, in order to encourage researchers to adapt the approach of domain knowledge incorporation to similar domains. This research directly builds on previous work on negotiation outcome classification and in this sense seeks to contribute towards a deeper understanding of the linguistic differences between successful and failing negotiations.

Secondly, assessing the predictability of specific phenomena is regarded as one of the main aspects of Predictive Analytics research contributions (Shmueli & Koppius 2011). We showed that we can reasonably well predict negotiation outcomes using a simplified representation of negotiation messages that incorporates knowledge about the domain structure using an approach similar to the topic modelling notion in Sentiment Analysis. We assessed this degree of predictability for the domain of electronic negotiation success, obtaining an accuracy that is above the respective baseline for full negotiations, with some degree of decrease in quality as the negotiation transcripts are shortened, somewhat contrarily to what is argued in Sokolova & Lapalme (2012) and, respectively Simons (1993). Nonetheless, the performances reported in parts remain significantly better than the baseline results and confirm that it is at least to a degree possible to detect negotiation success or failure at an early stage.

The notion of negotiation success predictability is also of practical value: Incorporating this model in an electronic Negotiation Support System could help in the alleviation of conflict and misunderstandings via an early detection of conflicting situations in the negotiation. This detection mechanism can now be used as a foundation for proactive communication support – for example via intervention in the negotiation process, further diagnostic information gathering about the source of the conflict, or the suggestion of introducing a mediating party to the negotiation. Potentially, negotiations on the brink of failure could be brought back onto the path towards a successful outcome, which avoids further seeking for negotiation partners or renegotiation and hence reduces information costs as well as transaction costs for the complete negotiation process (Schoop et al. 2008).

A simplified representation of a negotiation as the one used in this paper is of course not without information loss. First and foremost, the representation we chose did not include a separation of the two negotiating parties, which may make it difficult for our model to detect asymmetric conflict situations (i.e. where one party experiences and communicates conflict while the other party is fine) as well as effects of reciprocal reactions to previous statements such as for example convergence of linguistic styles of the negotiators (Huffaker et al. 2011). The second dimension we collapsed for our method of representation is the temporal component, which makes it difficult to track conflict development over time. These two dimensions allow potential for a lot of future research, it is for example thinkable to model the negotiation steps as a time series of an interacting dyad and employ sequential analysis methods (Kenny et al. 2006). This would also allow to capture transitions of emotional-affective states of the negotiators, which provide information about conflict development (Filipowicz et al. 2011, also note the work of Smith et al. 2005 on the topic of interaction sequences). However, as the representations become more complex, the requirements regarding the dataset size and quality increase considerably, in order to prevent blurring of the classification results, so at this point this seems rather difficult to achieve from a practical perspective. A further step in this direction could also be to perform a more fine-grained analysis of the negotiation transcripts, for example on the level of single sentences in order to obtain a more concise picture of the conflict development. One could for example use the sentence-level evaluation as a means to weigh utterances according to their importance for the negotiation result. It is thinkable to separate explicit evaluative statements from factual information exchanges using weighting functions. Statements with strong subjective evaluations could then receive a higher weighting since they overshadow other utterances that might point into different directions.

Another promising research direction is the introduction of more complex feature types – for example one could try to further differentiate the valence of the sentiments into different emotional categories. This would allow the differentiating notion about the rather differing effects of the various facets of negative emotions in negotiations (see Lelieveld et al. 2012 and Lelieveld et al. 2013). Furthermore, up to this point there is no detection of strategic displays of emotions and (of the same problem category), misinformation or deception in the negotiation process. Whilst this is a rather complex Natural Language Processing-task, there has been some work on it recently (Zhou et al. 2004, van Swol et al. 2014), which would be interesting to be integrated into the classification scheme.

4.6. References

Bichler, M., Kersten, G. and Strecker, S. (2003). "Towards a structured design of electronic negotiations" *Group Decision and Negotiation* 12 (4), 311–335.

Blei, D. M., Ng, A. Y. and Jordan, M. I. (2003). "Latent Dirichlet Allocation" *Journal of Machine Learning Research* 3, 993–1022.

Bouckaert, R.R. (2003), “Choosing between two learning algorithms based on calibrated tests”, in Fawcett, T. and Mishra, N. (Eds.), *Proceedings of the 20th International Conference on Machine Learning, Washington DC, 21.-24.08.*, AAAI Press, Menlo Park, CA, pp. 51–58.

Brett, J. M., Olekalns, M., Friedman, R. A., Goates, N., Anderson, C. and Lisco, C. C. (2007). "Sticks and Stones: Language, Face, and Online Dispute Resolution" *Academy of Management Journal* 50 (1), 85–99.

Chaovalit, P. and Zhou, L. (2005), “Movie Review Mining: A Comparison between Supervised and Unsupervised Classification Approaches”, in *Proceedings of the 38th Hawaii International Conference on System Sciences, Big Island, Hawaii, 03.-06.01.2005*, IEEE.

Cortes, C. and Vapnik, V. (1995). "Support-Vector Networks" *Machine Learning* 20 (3), 273–297.

Cover, T. and Hart, P. (1967). "Nearest neighbor pattern classification" *IEEE Transactions on Information Theory* 13 (1), 21–27.

Cox, D.R.(1958). The Regression Analysis of Binary Sequences. Journal of the Royal Statistical Society 20(2), 215–242.

Daft, R. L., Lengel, R. H. and Trevino, L. K. (1987). "Message Equivocality, Media Selection, and Manager Performance: Implications for Information Systems" *MIS Quarterly* 11 (3), 355–366.

De Dreu, Carsten K. W., Gross, J., Méder, Z., Giffin, M., Prochzkova, E., Krikeb, J. and Columbus, S. (2016), “In-group defense, out-group aggression, and coordination failures in intergroup conflict”, *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 113 No. 38, pp. 10524–10529.

Dennis, A. R., Fuller, R. M. and Valacich, J. S. (2008). "Media, Tasks, and Communication Processes: A Theory of Media Synchronicity" *Management Information Systems Quarterly* 32 (3), 575–600.

Filipowicz, A., Barsade, S. and Melwani, S. (2011). "Understanding Emotional Transitions: The Interpersonal Consequences of Changing Emotions in Negotiations" *Journal of Personality and Social Psychology* 101 (3), 541–556.

Finkel, J., Grenager, T. and Manning, C. D. (2005). "Incorporating Non-local Information into Extraction Systems by Gibbs Sampling" *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL 2005)* 2005, 363–370.

Forgas, J.B. (1998), “On feeling Good and Getting Your Way: Mood Effects on Negotiator Cognition and Bargaining Strategies”, *Journal of Personality and Social Psychology*, Vol. 74 No. 3, pp. 565–577.

Forman, G. (2008). "Feature Selection for Text Classification", in H. Liu and H. Motoda (eds.), *Computational methods of feature selection*, Chapman & Hall/CRC, Boca Raton.

Gibbons, P., Bradac, J. J. and Busch, J. D. (1992). "The Role of Language in Negotiations: Threats and Promises", in L.L. Putnam and M.E. Roloff (eds.), *Sage Annual Reviews of Communication Research: Communication and Negotiation*, pp. 156–175, SAGE Publications, Thousand Oaks, CA.

Griessmair, M. and Köszegi, S. T. (2009). "Exploring the Cognitive-Emotional Fugue in Electronic Negotiations" *Group Decision and Negotiation* 18, 213–234.

Corbin, J. and Strauss, A. (1990), “Grounded Theory Research: Procedures, Canons, and Evaluative Criteria”, *Qualitative Sociology*, Vol. 13 No. 1, pp. 3–21.

Guyon, I. and Elisseeff, A. (2003). "An introduction to variable and feature selection" *Journal of Machine Learning Research* 3, 1157-1182, from http://dl.acm.org/citation.cfm?id=944919.944968.

Harinck, F. and Druckman, D.(2017). Do Negotiation Interventions Matter? Resolving Conflicting Interests and Values. Journal of Conflict Resolution 61(1), 29–55.

Hatfield, E., Cacioppo, J.T. & Rapson, R.L. (1994), *Emotional Contagion*, Cambridge University Press, Cambridge, MA.

Hine, M. J., Murphy, S. A., Weber, M. and Kersten, G. E. (2009). "The Role of Emotion and Language in Dyadic E-Negotiations" *Group Decision and Negotiation* 18 (3), 193–211.

Hu, M. and Liu, B. (2004). "Mining and Summarizing Customer Reviews" Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2004.

Huffaker, D. A., Swaab, R. and Diermeier, D. (2011). "The Language of Coalition Formation in Online Multiparty Negotiations" *Journal of Language and Social Psychology* 30 (1), 66–81.

Japkowicz, N. and Shah, M. (2011), *Evaluating Learning Algorithms: A Classification Perspective*, Cambridge University Press, Cambridge, MA.

Kennedy, G. (2003). "Amplifier Collocations in the British National Corpus: Implications for English Language Teaching" *TESOL Quarterly* 37 (3), 467–487.

Kennedy, A. and Inkpen, D. (2006). "Sentiment Classification of Movie Reviews using Contextual Valence Shifters" *Computational Intelligence* 22 (2), 110–125.

Kenny, D.A., Kashy, D.A. & Cook, W.L. (2006), *Dyadic Data Analysis*, The Guilford Press, New York, London.

Kersten, G. E. and Noronha, S. J. (1999). "WWW-based negotiation support: design, implementation, and use" *Decision Support Systems* 25 (2), 135–154.

Körner, M. and Schoop, M. (2014). "Sentiment-Based Assessment of Electronic Mixed-Motive Communication - A Comparison of Approaches", in L. Brooks, D. Wainwright, & D. Wastell (eds.), *Proceedings of the UK Academy for Information Systems Conference,* Oxford, UK, 08.04-09.04.

Komorita, S. S. and Parks, C. D. (1995). "Interpersonal Relations: Mixed-Motive Interaction" *Annual Review of Psychology* 46, 183–207.

Lea, M. and Spears, R. (1992). "Paralanguage and social perception in computer-mediated communication" *Journal of Organizational Computing* 2, 321–341.

Lelieveld, G.-J., van Dijk, E., van Beest, I. and van Kleef, Gerben A. (2012). "Why Anger and Disappointment Affect Other's Bargaining Behavior Differently: The Moderating Role of Power and the Mediating Role of Reciprocal and Complementary Emotions" *Personality and Social Psychology Bulletin* 38 (9), 1209–1221.

Lelieveld, G.-J., van Dijk, E., van Beest, I. and van Kleef, Gerben A. (2013). "Does Communicating Disappointment in Negotiations Help or Hurt? Solving an Apparent Inconsistency in the Social-Functional Approach to Emotions" *Journal of Personality and Social Psychology* 105 (4), 605–620.

Liu, B. (2012), *Sentiment Analysis and Opinion Mining*, Morgan & Claypool, San Rafael, CA.

Manning, C.D., Raghavan, P. & Schütze, H. (2008), *Introduction to information retrieval*, Cambridge University Press, Cambridge.

Mao, Y. and Lebanon, G.(2007). Isotonic Conditional Random Fields and Local Sentiment Flow, In Schölkopf, B., Platt, J.C., and Hoffman, T. (eds.), Advances in Neural Information Processing Systems 19: Proceedings of NIPS 2006, MIT Press.

Markus, M. L. (2014). "Maybe not the king, but an invaluable subordinate: A commentary on Avison and Malaurent's advocacy of 'theory light' IS research" *Journal of Information Technology* 29, 341–345.

Maron, M. E. and Kuhns, J. L. (1960). "On Relevance, Probabilistic Indexing and Information Retrieval" *Journal of the ACM* 7 (3), 216–244.

Maron, M. E. (1961). "Automatic Indexing: An Experimental Inquiry" *Journal of the ACM* 8 (3), 404–417.

Murthy, D. (2015), “Twitter and elections: are tweets predictive, reactive, or a form of buzz”, *Information, Communication & Society*, Vol. 18 No. 7, pp. 816–831.

Pennebaker, J.W., Booth, R.J., Boyd, R.L. and Francis, M.E. (2015), *Linguistic Inquiry and Word Count: LIWC 2015*, Pennebaker Conglomerates, Austin, TX.

Pennebaker, J.W., Francis, M.E. and Booth, R.J. (2001), *Linguistic Inquiry and Word Count: LIWC 2001*, Erlbaum Publishers, Mahwah, NJ.

Polanyi, L. and Zaenen, A. (2006), “Contextual Valence Shifters”, in Shanahan, J.G., Qu, Y. and Wiebe, J. (Eds.), *Computing Attitude and Affect in Text: Theory and Applications: The Information Retrieval Series*, Springer, Dordrecht, pp. 1–10.

Quinlan, J. (1986). "Induction of Decision Trees" *Machine Learning* 1 (1), 81–106.

Riloff, E., Patwardhan, S. and Wiebe, J.(2006). Feature Subsumption for Opinion Analysis, In Jurafsky, D. and Gaussier, É. (eds.), Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, Sydney, 22.-23.07.2006, pp. 440–448.

Rothbart, M. and Hallmark, W. (1988). "In-group-out-group differences in the perceived efficacy of coercion and conciliation in resolving social conflict" *Journal of Personality and Social Psychology* 55 (2), 248–257.

Schoop, M. (2010). "Support of Complex Electronic Negotiations", in D.M. Kilgour and C. Eden (eds.), *Advances in Group Decision and Negotiation*, pp. 409–423, Springer Netherlands, Dordrecht.

Schoop, M., Jertila, A. and List, T. (2003). "Negoisst: a negotiation support system for electronic business-to-business negotiations in e-commerce" *Data & Knowledge Engineering* 47 (3), 371–401.

Schoop, M., Köhne, F., Staskiewicz, D., Voeth, M. and Herbst, U. (2008). "The antecedents of renegotiations in practice—an exploratory analysis" *Group Decision and Negotiation* 17 (2), 127–139.

Schoop, M., van Amelsvoort, M., Gettinger, J., Körner, M., Köszegi, S.T. and van der Wijst, P.(2014). The Interplay of Communication and Decisions in Electronic Negotiations: Communicative Decisions or Decisive Communication? Group Decision and Negotiation 23, 167–192.

Sebastiani, F. (2002). "Machine learning in automated text categorization" *ACM Computing Surveys* 34 (1), 1–47.

Shmueli, G. and Koppius, O. R. (2011). "Predictive Analytics in Information Systems Research" *MIS Quarterly* 35 (3), 553–572.

Short, J., Williams, E. & Christie, B. (1976), *The social psychology of telecommunications*, Wiley, London.

Simons, T. (1993), “Speech Patterns and the Concept of Utility in Cognitive Maps: The Case of Integrative Bargaining”, *Academy of Management Journal*, Vol. 36 No. 1, pp. 139–156.

Smith, P.L., Olekalns, M. and Weingart, L.R. (2005), “Markov Chain Models of Communication Processes in Negotiation”, *International Negotiation*, Vol. 10, pp. 97–113.

Sokolova, M. and Lapalme, G. (2012). "How Much Do We Say? Using Informativeness of Negotiation Text Records for Early Prediction of Negotiation Outcomes" *Group Decision and Negotiation* 21 (3), 363–379.

Sokolova, M., Nastase, V. and Szpakowicz, S. (2005). "Feature Selection in Electronic Negotiation Texts", *Proceedings of Recent Advances in Natural Language Processing,* Borovets, Bulgaria, pp. 518–524.

Sokolova, M. and Szpakowicz, S. (2005). "Analysis and Classification of Strategies in Electronic Negotiations", in B. Kégl and G. Lapalme (eds.), *Advances in Artificial Intelligence - Proceedings of the 18th Conference of the Canadian Society for Computational Studies of Intelligence*, pp. 145–157.

Sokolova, M., Szpakowicz S. and Nastase V. (2004). "Using language to determine success in negotiations: A preliminary study" *Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science)* 3060

Somasundaran, S. and Wiebe, J. (2009). "Recognizing Stances in Online Debates", in K.-Y. Su, J. Su, & J. Wiebe (eds.), Proceedings of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing, Singapore, 02.-07.08.2009, pp. 226–234.

Te'eni, D. (2001). "Review: A Cognitive-Affective Model of Organizational Communication for Designing IT" *Management Information Systems Quarterly* 25 (2), 251–312.

Thompson, L.L.(1990). An Examination of Naive and Experienced Negotiators. Journal of Personality and Social Psychology 59(1), 82–90.

Thompson, L. L. and Nadler, J. (2002). "Negotiating via Information Technology: Theory and Application" *Journal of Social Issues* 58 (1), 109–124.

Tumasjan, A., Sprenger, T.O., Sandner, P.G. and Welpe, I.M. (2010), “Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment”, in Cohen, W.W. and Gosling, S. (Eds.), *Proceedings of the Fourth International Conference on Weblogs and Social Media, Washington, DC, 23.-26.05.2010*, AAAI Press, Menlo Park, CA, pp. 178–185.

Twitchell, D. P., Jensen, M. L., Derrick, D. C., Burgoon, J. K. and Nunamaker, J. F. (2013). "Negotiation Outcome Classification Using Language Features" *Group Decision and Negotiation* 22 (1), 135–151.

van Kleef, Gerben A., De Dreu, Carsten K. W., Pietroni, D. and Manstead, A. (2006). "Power and emotion in negotiation: Power moderates the interpersonal effects of anger and happiness on concession making" *European Journal of Social Psychology* 36, 557–581.

van Kleef, Gerben A., De Dreu, Carsten K. W. and Manstead, A. (2004). "The interpersonal Effects of Emotions in Negotiations: A Motivated Information Processing Approach" *Journal of Personality and Social Psychology* 87 (4), 510–528.

van Kleef, Gerben A. and Côté, S. (2007). "Expressing Anger in Conflict: When It Helps and When It Hurts" *Journal of Applied Psychology* 92 (6), 1557–1569.

Van Swol, Lyn M. and Braun, M. T. (2014). "Communicating Deception: Differences in Language Use, Justifications, and Questions for Lies, Omissions, and Truths" *Group Decision and Negotiation* 23, 1343–1367.

Walther, J. B. (1992). "Interpersonal effects in computer-mediated interaction: A relational perspective" *Communication Research* 19 (1), 52–90.

Walther, J.B. (2011), “Theories of Computer-Mediated Communication and Interpersonal Relations”, in Knapp, M.L. and Daly, J.A. (Eds.), *The SAGE Handbook of Interpersonal Communication*, SAGE Publications, Thousand Oaks, CA, pp. 443–479.

Wiebe, J., Wilson, T. and Cardie, C. (2005). "Annotating Expressions of Opinions and Emotions in Language" *Language Resources and Evaluation* 39 (2-3), 165–210.

Zaidan, O.F., Eisner, J. and Piatko, C.D.(2007). Using "Annotator Rationales" to Improve Machine Learning for Text Categorization, In Sidner, C.*, et al.* (eds.), Proceedings of the 2007 Conference of the North American Chapter of the Association for Computational Linguistics, Rochester, NY, 22.04.-27.04., pp. 260–267.

Zhang, L. and Liu, B. (2011). "Identifying noun product features that imply opinions", in D. Lin, Y. Matsumoto, & R. Mihalcea (eds.), *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies,* Portland, USA, 19.-24.06.2011, pp. 575–580.

Zhou, L., Burgoon, J. K., Nunamaker Jr., J. F. and Twitchell, D. P. (2004). "Automating Linguistics-Based Cues for Detecting Deception in Text-based Asynchronous Computer-Mediated Communication" *Group Decision and Negotiation* 13, 81–106.

Zhuang, L., Jing, F. and Zhu, X.-Y. (2006), “Movie Review Mining and Summarization”, 6.-11.11.2006, Arlington, Virginia.

5. Study III: Micro-Level Sentiment Assessment of Negotiators' Utterances

Micro-Level Sentiment Assessment of Negotiators' Utterances

Abstract. Assessment of opinionated utterances in electronic negotiations are important to detect potential negotiation failure automatically. These opinionated statements can be analysed using Sentiment Analysis. A key component in Sentiment Analysis is the distinction between opinionated, subjective statements and factual, objective statements. Based on manual coding of 28,667 single sentences from electronic negotiation data, we trained several classification models that fulfil this task. We show that sentiment assessment of negotiation statements is feasible on the micro-level and present according results.

Keywords: Electronic Negotiations, Sentiment Analysis, Subjectivity Assessment, Text Mining

5.1. Introduction and Motivation

Research on negotiations in various contexts has acknowledged the important influence of communication between the negotiators on the outcome of the negotiation. The utterances the negotiators exchange act as a stepwise exchange of relational and substantive communication and thus provide a framework for shaping the development of a relationship characterized by mutual trust which in turn enable the exploration of the decision space and allow for the identification of joint potentials. Specifically, in interorganisational settings, more and more negotiations are conducted electronically, which, according to seminal theories on Computer-Mediated Communication introduces potentials for the amplification of conflict, misunderstandings, and impediments of trust development. Hence, Negotiation Support Systems seek to alleviate these negative effects via the provision of supportive mechanisms that aim to clarify the communication situation and to enable an atmosphere of mutual trust. There have been various suggestions for these systems to take a more proactive stance towards the negotiators via monitoring of negotiation development and offering direct support for the specific situation without being explicitly prompted (Curhan & Pentland 2007, Kersten & Lai 2007, Druckman et al. 2012). In order to achieve this, a Negotiation Support System needs to be able to assess a given negotiation situation with reasonable accuracy and also to interpolate the potential development regarding the negotiation outcome.

Recent advances in the context of Predictive Analytics have identified its approaches as being of extreme practical relevance, regarding its potential to provide additional information in ambiguous decision-making contexts and thus increasing business value. Similarly, the topic is of considerable interest in recent IS research where its potential to contribute to the assessment of existing theories as well as to the generation of new theories has been recognized (Shmueli and Koppius 2011).

In a related context, Sentiment Analysis (or Opinion Mining) has gained popularity as a means to evaluate large unstructured data sets automatically, particularly user-generated content in the web (e.g. Chesley et al. 2006, Mostafa 2013, Thelwall et al. 2012). Originating from an Information Science context, methods of Natural Language Processing and Text Mining have been used to facilitate such automated analysis. Detection methods of sentiments in texts are used to evaluate the author's attitude toward its domain, e.g. a specific movie, political parties, events, brands etc.

The aim of this paper is to contribute to the application of Sentiment Analysis techniques in written communication between human negotiators in business contexts. Our rationale is to view e-negotiation documents (or transcripts) as an interchange of opinionated statements that reflect the conflict that is to be resolved in the negotiation process. We argue that the detection of the polarity in subjective negotiation statements can provide a system with indicators of negotiation success, and – more interestingly – of potential negotiation breakdowns. These indicators can then be used in a predictive manner in order to assess the likelihood of failure while the negotiation is still ongoing. Such a detection method could be implemented in the context of asynchronous B2B Negotiation Support Systems (NSSs) (Lim and Benbasat 1992/93, Schoop 2010) and be used as an entry point for further diagnostics of problems in negotiations and subsequently suggestions for resolving the specific problems (e.g. via introduction of a mediator) and thus preventing unnecessary impasse in negotiations or reduce the likelihood of renegotiations – which ultimately results in shorter timespans until a final agreement is reached and thus more efficient negotiations. In line with this we present our research question that will be assessed in the content of this paper:

To what extent is it possible to detect subjective statements in e-negotiation transcripts with reasonable accuracy?

In order to answer this research question, the remainder of this paper is structured as follows: First, we will discuss the theoretical foundations that communication research in (e-)negotiation literature provides – thus motivating that subjective assessments as well as expressions of emotions in negotiation are influential regarding the negotiation outcome. Afterwards, we will give a brief overview on Sentiment Analysis methods and will provide reasoning how we adopted techniques from Sentiment Analysis in order to detect subjective statements on a large corpus of experimental negotiation data.

5.2. Theoretical Foundations

5.2.1. Communication in Negotiations

Communication between the negotiators is a crucial aspect of each negotiation process. Negotiators communicate to exchange offers and preference information in order to achieve mutually beneficial outcomes, which at the same time satisfy their own individual needs. Since the negotiators are typically in a state of mutual dependency (Bichler et al. 2003), their communication thus adopts a mixed-motive stance (Komorita and Parks 1995), where negotiators must carefully judge how to achieve the most for themselves, while not being overly conflictive – which may lead to a breakdown of the negotiations. Thus, they use communication to persuade and to demand, to provide rationales for their demands and to evaluate the state of the negotiation through affective responses towards their counter-part's actions and behaviours (cf. Habermas 1984, Spangle and Isenhart 2003). Negotiation communication is typically characterized by this dual nature of not only resolving but also creating tension throughout the negotiation process.

The way negotiators communicate through the negotiation process can have strong impacts on the negotiation outcome, economically as well as on the intangible, socio-psychological level. Appropriate communication behaviour can increase economic outcomes on a joint and on an individual level (Morris et al. 2002). Vice versa, inappropriate communication, such as overly conflicting behaviour may result in lower gains or even premature impasses, i.e. a breakdown of the negotiation (Adair et al. 2004). On the psychological level, the manner of communication affects negotiators' satisfaction with their outcome and the negotiation process in general (Barry et al. 2004) and can impede long-term relationships even in successful negotiations, since the willingness to cooperate with the same partner in the future is negatively affected. Especially statements that clearly show negative affect have been under consideration of recent research in negotiation. Whilst the classical perspective assumed showing positive affect in general to have a positive effect and showing negative affect to have negative consequences, this assumption has been shown to be insufficient. Recent research shows that tactile usage of negative emotions can indeed increase individual gains, when the recipient deems the sender's negative reaction to be appropriate given the context of the specific negotiation (van Kleef et al. 2010). Success of a negotiation is then characterized by a "breaking out" of the spiral of negative communication at some point, and an eventual shift back to positive communication towards the end of the negotiation (Griessmair & Köszegi 2009).

In Computer-Mediated Communication (CMC), these effects also have a tendency to be enhanced, since specific communication biases may occur in negotiations that facilitate misunderstandings and lower levels of trust between the negotiators (Thompson and Nadler 2003, Thompson 2005). Furthermore, a riskier negotiation and communication behaviour is encouraged when negotiators only communicate in an asynchronous, written fashion: Due to their perceived anonymity they feel less social accountability for their actions which can lead to abrasive communication behaviour ("flaming", e.g. Johnson et al. 2009) and can impede negotiation processes drastically (Kiesler and Sproull 1992; Friedman and Currall 2003). Especially in written communication without the ability to transmit visual or aural social cues, the relative importance of the cues conveyed via the written statements is greatly increased and thus, these cues become especially viable as subject of research activity (Walther 1992; Walther and Burgoon 1992).

It is known that communication plays an important role regarding the outcome of the negotiation. Furthermore, the relationship between negotiation communication and outcomes is of complex nature, displaying positive affect may not necessarily be good and displaying negative reaction may not necessarily be bad (van Kleef et al. 2004). Nevertheless, it seems interesting to study evaluative utterances that negotiators make, reflecting the cognitive and affective states the negotiators are in, indicating the degree of severity of the conflict and thus hinting at potential success or failure of the negotiation, or strongly influencing socio-psychological outcomes of the negotiation, i.e. a more differentiated view of conclusion, where quality, satisfaction and/or fairness perceptions about process and outcome are evaluated by the negotiators[5]. Communication in negotiations has been described as consisting of repeating micro-level patterns that shape and change the negotiation process (Putnam and Roloff 1992). Therefore, we propose techniques from Machine Learning and Sentiment Analysis in the remainder of this paper in order to detect and assess such evaluative statements according to their subjectivity and valence.

5 Note that we focus on viewing the outcome of the negotiation as dichotomous which is a deliberate simplification regarding the overall goal of the research (early detection of potential failure) and as such a common sacrifice in external validity (see for example Kersten and Zhang 2003, Twitchell et al. 2013).

5.2.2. Sentiment Analysis and Subjectivity Detection

Approaches using automatic processing methods to make statements about negotiation data have yielded promising results in the past. Especially data mining and Machine Learning techniques have been employed in negotiation research with promising results (e.g. Kersten and Zhang 2003, Nastase et al. 2007, Sokolova and Lapalme 2012, Twitchell et al. 2013) regarding prediction of negotiation outcomes based on negotiation data. However, none of the aforementioned studies explicitly tries to separate subjective utterances from objectively stating facts. Sentiment Analysis and subjectivity detection provide us with a methodological framework to include this step into classification processes in negotiations, potentially allowing us to make clearer distinctions between utterances that reflect cognitive states and evaluations of the negotiators.

Given a specific subjective document, the core task of Sentiment Analysis is to evaluate the opinion expressed by the author of said document automatically. As such, Sentiment Analysis has a wide range of applications in the context of user-generated content in the web, for example product reviews by customers on e-commerce platforms, tracking of political opinions or stock market predictions from sentiment expressions on social media (Tumasjan et al. 2010, Bollen et al. 2011) and tracking developments of online discussions (Pang and Lee 2008, Mostafa 2013). The vast majority of these applications is focused specifically on the polarity of the opinion expressed by the author, i.e. the task is to determine whether a document expresses a positive or a negative opinion towards a specific entity (Liu 2012). On the methodological level, Sentiment Analysis can thus be understood as an interdisciplinary field, combining techniques from Sociolinguistics, Computational Linguistics, Natural Language Processing, Machine Learning and Information Retrieval and Text Mining in order to fulfil its task.

In its simplest application, Sentiment Analysis can be viewed as a Text Classification problem. A set of documents with known opinion orientation (e.g. created through manual coding) is represented as a word vector and – after applying a multitude of different preprocessing methods such as part-of-speech tagging, stemming, n-gram-generation and feature selection (e.g. Forman 2008) used to train a Machine Learning classifier. Evaluation of different performance metrics on test sets yields information about how accurate an unknown document can then be classified by the trained model. Newer approaches also employ dictionary-based detection of sentiment words in order to identify appropriate features for the classification process (Hu and Liu 2004, Esuli and Sebastiani 2006). In this case, lexica with sentiment words previously annotated with polarities are used. Whilst there exist general purpose-lexica (Hu and Liu 2004, Wiebe et al. 2005, Baccianella et al.

2010), researchers can also employ different methods on corpora of documents in order to create domain-specific lexica, which in theory should provide higher levels of accuracy in determining appropriate sentiment words and in assessing their polarity (Kanayama and Nasukawa 2006). Lastly, indication words that may shift the valence of sentiment expressions such as intensifying adverbs and negation statements are also taken into account to increase classification accuracy (e.g. Polanyi and Zaenen 2006, Taboada et al. 2011, Wilson et al. 2011).

However, written texts often consist of a mixture of factual information and opinionated statements that reflect the author's position regarding the topic of the text or that directly relate to the factual in-formation given beforehand. Also, an author may display mixed opinions on different aspects of the domain s/he is writing about. In these cases, classification on the document level may be oversimplifying since it cannot capture these complex relationships (Liu 2012). Hence, more fine-grained methods are employed such as classification on the level of single sentences in the document or regarding specific aspects of the domain the text is about.

Here, another subtask of Sentiment Analysis arises: In order to be more accurate in judging multifaceted opinionated texts, one must first separate this factual information from the opinionated statements. In Sentiment Analysis this is commonly known as Subjectivity Classification (or Subjectivity Detection, Wiebe et al. 1999). Subjectivity classification becomes extremely important in situations where authors present factual information along with their opinions and when multiple aspects are discussed at once. Classical approaches, especially when new domains are encountered use manually coded data in order to train a classifier to determine subjective statements – so the method to fulfil the task is rather similar to the determination of polarities, although the goal categories for the classifiers differ.

Since negotiation transcripts can be considered an archetype of complex, multifaceted opinionated documents, in order to classify negotiation transcripts according to their valence and eventually according to the outcome of the negotiation, a respective model should first separate factual information from evaluative statements that the negotiators make, which is the main goal of this paper and which will be described in the following chapter.

5.3. Research Methodology

The data set that was used for the research in this paper stems from several B2B-negotiation experiments that were conducted between students of different universities in

Europe, mostly from Germany, Austria, and the Netherlands. Participants were required to act as representatives from two companies that negotiate a joint venture contract that involved distributive as well as integrative issues. Negotiations were conducted using the NSS Negoisst (Schoop et al. 2003, Schoop 2010), providing decision support functionalities as well as structuring of the message exchange, i.e. limiting the negotiators to strictly alternating communication. The data set amounts up to 2459 messages over 173 negotiations, 75% of which were successful and 25% were failing, i.e. ended in impasse.

These negotiation messages were automatically split into 28,667 sentences – which were then put into randomized order and given to two human coders, an approach common in methodologies such as Content Analysis (Srnka and Köszegi 2007) in order to cross-validate the subjective judgements of the single coder thus counteracting methodological problems in data-driven development of constructs (Shmueli and Koppius 2011). We tried to obtain a notion of whether an assessment of subjectivity vs. objectivity is generalizable in the sense that it can be done in an objective fashion (previous research on this topic points in the same direction e.g. Wiebe et al. 1999), therefore we used independent coders, which makes it possible to test the validity and identifiability of our defined classes. Reliability scores of the schemes can be obtained by calculating the kappa measure (Cohen 1960), which is defined as

$$\kappa = \frac{p_0 - p_c}{1 - p_c}$$

where p_0 is the proportion of classified units the coders agreed on (i.e. made the same class decision) and p_c is the proportion where agreement occurs by chance. The result is a quality measure that accounts for chance agreement of the human coders.

Furthermore – as Machine Learning classifiers tend to become very unstable when dealing with a high number of categories – we tried to keep the coding scheme as simple as possible. Hence, coders were given the simplistic 2-step scheme shown in table 11.

Classification dimensions	Subjectivity Classification	Polarity Classification
classes	Fact	Positive
	Opinion	Negative
	Uncertain	Strong positive
		Strong negative
		No orientation ("Neutral")
		Mixed Orientation
		Uncertain

Table 11: Coding Scheme for Negotiator Utterances

Coding for each unit thus involved first a judgement on whether a factual or an opinionated statement is made (similar to the classes described in Yu and Hatzivassiloglou 2003) and then, in the next step the valence of the statement was categorized into positive, negative, neutral, or mixed. We decided to opt for this hierarchical classification scheme as opposed to a collapsed scheme (with categories such as *objective*, *positive, neutral*, see Wilson et al. 2009), since contrary to most Sentiment Analysis approaches in negotiation data factual statements may have polarities as well, either explicitly or with an implicit emotional undertone (Griessmair and Köszegi 2009). Therefore a synonymous understanding of the categories *fact/objective* and *neutral* as is sometimes applicable in Sentiment Analysis (Wilson et al. 2009) is not possible in our case. This distinction has also been put under scrutiny for classical Sentiment Analysis approaches, where the concept of "polar facts" has been introduced as potentially influencing the learning of sentiment expressions (e.g. Ruppenhofer & Rehbein 2012, Toprak 2010)

The reason for the inclusion of an "uncertain" category which can diminish result quality due to "lazy coding" in difficult cases, is again our future interest with the coded data, we wanted the distinction between subjectivity and objectivity to be as clear as possible when using these manual codes as an input for a Machine Learning scheme.

We briefly considered classifying opinion holders, i.e. entities expressing the opinion shown in the unit, since the notion exists in Sentiment Analysis and especially for negotiations may yield additional information, e.g. in assessing the source of a conflicting situation or evaluating diverging conflict perceptions of the negotiators. Due to a lack of data in our overall set which only amounts to 173 negotiations, we omitted this classification since we did not consider it feasible to purposefully use an opinion holder classification with such few data points. Furthermore, a generalized opinion holder identification in negotiation is complex since aside from pronouns, indication of opinion holders would only rely on names and labels that are specific to our training data set, which contradicts the aim to develop a negotiation classifier that is as domain-independent as possible.

Cohen's kappa was used to evaluate intercoder consistency (Brennan and Prediger 1981), the resulting values were 0.74 for subjectivity classification and 0.63 for polarity classification – both kappa values can be considered decent. Since the categories "Strong negative" and "Strong positive" were only sparsely populated (0.0078% and 0.0013% of all coded units), we decided to group them together with the "Negative" and "Positive" category. The resulting kappa value for the collapsed scheme was 0.66, which still can be considered a moderate rate of agreement.

In the next step, we tried to create a "stereotypical" training set for our Machine Learning scheme. We omitted all sentences where the coders disagreed or that were coded as uncertain, leaving us with 21,788 coded sentences and a two-class classification problem regarding subjectivity assessment. We split the dataset up in an 80-20 ratio, creating a training set (17,430 units) and a test set (4,358 units) for later evaluation. All subsequent filtering and classification steps were performed using RapidMiner (Hofmann and Klinkenberg 2013). We used the training set as a corpus to extract a base lexicon containing terms that are specifically distinctive for fact and opinion categories. The corpus first was POS-tagged and stemmed, then we applied chi²-feature selection to identify the most distinctive adjectives, verbs and nouns in our corpus – we base further classifications mainly on word tokens out of these three classes, since previous research indicates that these should hold the majority of informational content regarding subjectivity detection (e.g. Wiebe et al. 2005). We decided experimentally on using a list of the 150 most distinctive features over our three parts of speech together. Additionally, we kept intensifying adverbs using the list of common intensifiers in Kenny and Inkpen (2003), as well as personal and possessive pronouns, as they play an important role in identifying opinionated speech (Pennebaker et al. 2003, Tausczik and Pennebaker 2009). Several different Machine Learning models were subsequently trained on the training set using 10-fold Cross-Validation. The different classifiers applied are: Naïve Bayes (Maron and Kuhns 1960), Decision Trees (Quinlan 1986), K-Nearest Neighbour (kNN, Cover and Hart 1967), Logistic Regression (Cox 1958), and Support Vector Machines (SVM, Cortes and Vapnik 1995). In the case of SVM, we also performed parameter tuning using a grid-based approach as recommended in Hsu et al. (2003) and applied a similar approach to parameters of Logistic Regression, Decision Trees and kNN. Parameter Tuning yielded best performances for SVM with C=512, γ=0.000122 (2^-13), radial kernel, Logistic Regression with C=32, γ=0.000122 (2^-13), Decision Trees with Minimum Gain=0.000122, Confidence=0.125 and lastly kNN with k=9 and cosine similarity as distance measure. Note that this course of action implies that we also use our training set as a validation set, an approach which has been described as sound in Japkowicz and Shah (2011). We decided experimentally on a unigram and bigram representation of each sentence – and each feature was represented using the TF-IDF-score. Lastly, before the actual training of the model, we again applied chi^2-feature selection, working only with the top 300 unigrams and bigrams.

The same process we described above for the subjectivity classification was subsequently repeated twice for the respective polarity decisions. Since we expected the polarity evaluation for factual statements to be based on different aspects than the polarity evaluation for subjective statements, we compiled a training and test set consisting only of subjective statements and one consisting only of factual statements. The subjective training and test set contained 5896 and 1478 units, whilst the fact training and test sets consisted of 11513 and 2881 units. Parameter tuning for the subjective set yielded best performances for SVM with C=128, γ=0.00003051757 with a radial kernel, Logistic Regression with C=512, γ=0.0001227031, Decision Trees with Minimum Gain=0.001 and Confidence=0.1 and kNN with k=3 and Cosine Similarity as distance measure. For the fact set we obtained values for SVM with C= 2048, γ=0.00048828125 with a radial kernel, Logistic Regression with C=128, γ= 0.00048828125, Decision Trees with Minimum Gain=0.001 and Confidence=0.1, and lastly kNN with k=3 and Cosine Similarity as distance measure.

The models that were created in this manner on the training set were subsequently applied to the test set, from where the performance measures in the following chapter are derived.

5.4. Results and Discussion

This section discusses the results of the aforementioned model performances. It is divided according to the single steps described above, starting with the results of the subjectivity classification and continuing with the polarity classification results, which in turn are split into sentences that had been coded as subjective and as not subjective.

5.4.1. Subjectivity Classification

We start with taking a look at word tokens that have been identified as distinctive between subjective and factual statements. Table 12 shows these most distinctive features for the subjectivity vs. objectivity decision identified via chi²-feature selection over all verb, noun, and adjective stems (similar to the analysis provided in Das & Chen 2007).

Feature	Type	Chi²-Weighting
think	Verb	1857.1
hope	Verb/Noun	912.0
understand	Verb/Noun	527.8
regard	Noun	444.1
pleas	Adjective	431.8
feel	Verb	423.9
seem	Verb	416.7
fair	Adjective	361.7
be	Verb	302.3
believ	Verb	300.4
suggest	Verb/Noun	299.9
kind	Adjective	260.2
dear	Adjective	204.9
propos	Verb/Noun	193.8
see	Verb	170.7
glad	Adjective	168.7
appreci	Verb	166.4
want	Verb	163.3
happi	Adjective	163.3
agreement	Noun	162.2

Table 12: Chi²-Weightings for Negotiator Utterances

Note that, since we applied stemming before the feature selection, some of the word stems can represent multiple grammatical forms of a specific word (e.g. understand/understanding, pro-pose/proposal/proposition), hence the ambiguous annotations in the "type"-column. As can be seen, the majority of the distinctive features are verbs that are linked to expressions of evaluation, thoughts and personal beliefs (think, hope, understand, feel, see, seem, appreci(ate), want) which is consistent with our expectations due to existing findings (Wiebe et al. 2005). The second important class of features contains stems of evaluative adjectives (pleas(ed), fair, glad, happi) which should also be expected in statements of subjective nature. Also it is interesting to note that almost all of the features in the list indicate membership of a sentence to the opinion class, while only very few features explicitly distinguish the fact class – aside from the verb "be", the only features that belong to this class indicate greeting formulas, which, (by coding convention) our manual coders put into the fact category (i.e. kind, regard(s), dear). This indicates that the distinction between subjective and objective statements is made by explicitly separating subjective statements from objective statements and not vice versa, i.e. the fact-class appears to be fuzzier in nature – an effect which also is hinted on by Yu and Hatzivassiloglou (2003). In our negotiation data set this effect could be enhanced, since it seems plausible that factual information that is given in the course of the negotiation is always directly related to the specific negotiation agenda and domain – whilst general subjective evaluations are similar to each other across different negotiation agendas. Since we used a filtering list that excludes terms that are overly specific of the experimental case in order to provide generalizability of the model, we potentially have destroyed parts of the indicators of factual information.

We report the results for the subjectivity classification in terms of giving the performance of the trained model on the test set. These performances are shown in table 13. All classifiers exhibited a performance on the test set that is on the same level of the performances in the validation runs on the training set, indicating that the generated models are very stable. The baseline scores have been determined using a trivial majority classifier that classifies every unit encountered as "fact".

Classifier	Accuracy	Precision	Recall	F-Measure	Kappa	RMSE
Baseline	61.45	30.73	50.00	38.07	0.000	0.487
Naïve Bayes	79.14	80.92	74.85	77.77	0.530	0.457
SVM	84.33	84.69	81.89	83.27	0.658	0.356
Regression	83.87	84.16	81.41	82.76	0.648	0.442
Decision Tree	83.85	84.34	81.22	82.75	0.647	0.356
kNN	78.25	79.77	74.38	76.98	0.518	0.400

Table 13: Subjectivity Classification Results

Performance scores for precision, recall and the F-measure are reported with a focus on the less populated class (i.e. "opinion"). Furthermore, we denote Cohen's kappa in order to rule out an overly optimistic assessment of accuracy measures, since the kappa statistic is less influenced by skew in the data set (Cohen 1960 Japkowicz and Shah 2011). Lastly, we give the root mean squared error (RMSE) as an assessment of classifier stability. Top performances for each criterion are highlighted in bold (cf. table 13).

As can be seen from table 13, all trained classification models perform much better on the test set than the trivial baseline classifier. In order to verify the consistency of this performance, we tested the models again on 10 random subsamples of our test set. The performances showed that indeed all classifiers scored significantly above the baseline (Two-matched samples t-test, $p < .0001$ for all classifiers) whereas the SVM, Logistic Regression and Decision Tree classifiers showed a significantly stronger result than kNN and Naïve Bayes ($p < 0.0001$). Among themselves SVM, Logistic Regression and Decision Trees did not differ significantly ($p > 0.1$ in all cases).

So, conclusively the results show that subjectivity detection can be performed on negotiation sentences with reasonable accuracy. Furthermore, the results we obtain are consistent over different classifiers, with SVM, Decision Trees and Logistic Regression performing strongest. Since especially SVM and regression models are known for performing well in this Text Classification tasks (e.g. Manning et al. 2008, Tang et al. 2009, Sokolova & Lapalme 2012) this indicates stability and reliability of the results. Also – although it is difficult, if not impossible to directly compare performance measures on differing data sets – the quality of the results seem to be on a comparable level of existing research on subjectivity detection (e.g. Yu and Hatzivassiloglou 2003, Maas et al. 2011).

5.4.2. Polarity Classification of subjective utterances

As for polarity classification we provide a brief preliminary discussion of distinctive features given in table 14. This table shows Chi²-feature selection results for sentences that had been classified by both coders as subjective.

Feature	Type	Chi²-Weighting
thank	Verb	617.4
happi	Adjective	475.7
glad	Adjective	371.3
dear	Adjective	323.3
am	Verb	302.7
t	Negation Particle	265.1
negoti	Noun	257.0
disappoint	Adjective/Verb	253.0
m	Verb Particle	195.8
agreement	Noun	193.6
sorri	Adjective	181.5
appreci	Verb	141.1
afraid	Adjective	123.9
nice	Adjective	117.5
are	Verb	112.8
insult	Adjective/Noun	111.8
cooper	Adjective/Verb/Noun	108.9
serious	Adjective	103.0
good	Adjective	99.8
pleasur	Noun	99.2

Table 14: Chi²-Weightings for the Polarity of Subjective Utterances

Compared to the selection results on the subjectivity decision, there are substantially more adjectives that have been identified as distinctive – all of which clearly hint at specific valences of utterances (happi, glad, disappoint(ed), sorri, afraid, nice, insult(ed), cooper(ative), serious and good). Interestingly, the verb "to be" – usually a classical stopword which is filtered out before analysis – occurs in three times in this list, twice in first person singular (am and the particle m such as in "I'm") and once in pluralized form (are). This is most likely due to regular co-occurrences of "to be" with positive sentiment utterances and, therefore, is in the list as an indicator of a positive sentiment. The reason for this is presumably rooted in sociolinguistics: People tend to use more direct expressions for positive emotions, while negative statements are oftentimes more complex verbal constructs, where behaviours such as hedging is applied in order to remain polite to the counterpart. Another often applied behaviour is a tendency towards negation constructs in negative utterances (i.e. formulations such as "not very good" are more common as just using the straightforward "bad") – which is also reflected in the feature list in the form of the particle t occurring in negation constructs such as aren't or won't. Lastly, we also detect a tendency towards positive utterances, which is also consistent with the notion that the variety in formulating things positively is lower than in the negative case.

The polarity classification problem for subjective utterance is a bit different from the subjectivity classification problem, mainly because in this case, we have three possible classifications – positive, neutral and negative – as possible sentence evaluations. Furthermore, the distribution of units across the classes is different from the subjectivity classification problem, our training set contained about 76% neutral statements, whilst the positive (14%) and negative (10%) classes were much more sparsely populated. Whilst this distribution corresponds to comparable annotation studies in literature (e.g. Wiebe et al. 2005), it makes the classification problem a rather asymmetric one and has some consequences regarding the evaluations of the performance measures. The main effect of such an asymmetric distribution is that accuracy as a performance measure loses most of its informative value, since a trivial classifier that classifies all utterances as neutral will obtain a performance of 76% in terms of accuracy, albeit being of no practical value. Thus, it is recommended to put stronger emphasis on agreement statistics (such as kappa) that seek to correct for high performance values obtained by chance as a direct result of an asymmetric class distribution (Japkowicz and Shah 2011).

Table 15 again shows the classifier performances on the test set.

Classifier	Accuracy	Precision	Recall	F-Measure	Kappa	RMSE
Baseline	75.83	25.28	33.33	28.75	0	0.731
Naïve Bayes	77.66	63.77	71.27	67.31	0.498	0.470
SVM	84.29	79.05	63.57	70.47	0.537	0.465
Regression	85.31	79.65	66.56	72.52	0.578	0.587
Decision Tree	81.18	75.55	53.82	62.86	0.395	0.404
kNN	82.13	74.17	59.45	66.00	0.468	0.39

Table 15: Performance of Subjective Polarity Classification

Here, Logistic Regression emerged as the strongest performer in overall averaged terms, nonetheless, the differences between individual classifier performances are relatively low. The overall performance values are lower than in the subjectivity detection problem, which most likely is a result of the combination of the ternary nature of the classification task as well as the strong asymmetry in class distribution. Hence, these values should be taken with care when assessed for their practical value. The main question one has to pose is whether it is more desirable to not detect a polar statement when there is one or whether to detect a polar statement where there is none. Regarding the original motivation of the paper, we seek to achieve an accurate detection of potential sources of conflict during a negotiation process, hence the desired value for this decision is a high recall for our sparsely populated classes, here especially negative polar statements are of interest. In order to further evaluate the inner workings of our classifiers, we take a look at their performances on individual classes, in terms of class-specific precision and recall, which are given in table 16. Here, in the order positive – negative – neutral, precision and recall values are given.

Classifier	PR_POS	REC_POS	PR_NEG	REC_NEG	PR_NT	REC_NT
Baseline	0	0	0	0	75.83	100
Naïve Bayes	52.38	75.12	48.62	57.89	90.32	80.80
SVM	79.17	55.61	72.29	39.47	85.68	95.62
Log. Regression	79.75	61.46	72.22	42.76	86.98	95.45
Decision Tree	87.64	38.05	56.94	26.97	82.07	96.43
kNN	72.03	50.24	66.23	33.55	84.25	94.55

Table 16: Subjective Polarity Classification Performance on Individual Classes

Here, common distortion effects from the class distribution are visible. All classifiers achieve reasonably high performances on the neutral classes, for the classes of positive and negative valence, the performances are considerably lower. Especially Decision Tree and kNN suffer from a strong bias towards the majority class, which results in very low recall values for positive and negative instances – in the case of Decision Tree this is even combined with a lowered precision on the negative class – which means that this class is not covered in a very accurate fashion by Decision Tree.

Interestingly, Naïve Bayes behaves differently from all the other classifiers, obtaining better scores for the recall values at the cost of classifier precision, i.e. any instance that actually belongs to the positive/negative class is more likely to be assigned as such by Naïve Bayes whilst on the other hand, Naïve Bayes is also more likely to erroneously assign a neutral instance to these two classes. Given the above statement on desired performances, one could argue that this is a behaviour in favour of Naïve Bayes, also counting in that Naïve Bayes obtains strong precision values on the neutral class – it seems that Naïve Bayes is least affected by the unequal class distribution. Nevertheless, the likelihood of false positives on the positive/negative valences is much higher for Naïve Bayes than for the other classifiers. Also, given that Logistic Regression performs rather solidly on the single classes resulting in high average performances and top scores in the F-Measure, the decision on a single best classification model remains rather inconclusive. At the end of the day, the decision comes down to whether the detection of a *higher percentage* of statements of valence is preferred over a *more accurate* detection of statements of valence, which is the typical precision-recall trade-off in Machine Learning problems.

5.4.3. Polarity Classification of factual statements

Regarding polarity classification for factual statements, the skew in the data is even stronger than in the subjective statement case. About 90% of the statements were identified by the coders as neutral, as opposed to only 4% positive and 6% negative statements. Averaged classification performances can be obtained from table 17

Classifier	Accuracy	Precision	Recall	F-Measure	Kappa	RMSE
Baseline	88.93	29.64	33.33	31.38	0	0.330
Naïve Bayes	84.21	56.46	72.84	63.61	0.428	0.396
SVM	90.56	68.90	56.59	62.14	0.443	0.430
Log. Regression	91.53	75.11	58.36	65.68	0.485	0.570
Decision Tree	90.87	76.43	49.16	59.83	0.365	0.282
kNN	90.70	71.23	54.71	61.89	0.423	0.285

Table 17: Polarity Classification Performance on Factual Statements

Here, we see that although absolute accuracy values of the classifiers have increased, the overall performance is worse than in the case of detecting valences of subjective statements. This is not surprising on a technical level considering the skewed class distribution, and also on a theoretical level, since it has been deemed difficult to assign valences to factual statements (Liu 2012) – which are often presented in a neutral tone since their main purpose is conveyance of information rather than stating an opinion, whereas positive or negative connotations of the facts stated are often implicit and hidden in the context or the exact choice of words. As in the subjective polarity classification case, Logistic Regression obtains the most solid performances across all classes, although its average recall is not particularly good. In general, all classifiers suffer from a very low recall over all classes, except once again Naïve Bayes, which in turn sacrifices a good score in precision. In this state, it is questionable whether valence evaluation of factual statements is a) purposeful and b) provides additional information regarding the overall goal of evaluating negotiation success. We tried to combine the advantages of the different classifiers by using Meta-classification procedures such as Voting and Stacking procedures, but found no improvement in classification accuracy. The main effect in combining the different classifiers was that the skewing effect accumulated over the meta-classification steps and performance converged into the direction of the baseline.

5.5. Conclusion

In this paper we theorized that subjectivity and polarity assessment of negotiation utterance is desirable on the micro-level in order to assess negotiators' communication and to draw conclusions about the negotiation process. As a first step towards a holistic classification approach for negotiation texts and transcripts, we trained several models that classify utterances according to their subjectivity. We showed that the subjectivity detection models perform significantly above the baseline – and on a comparable level to subjectivity detection in different domains. The research done in the context of this paper thus marks an initial step towards a more fine-grained automatic assessment of sentiment expressions in negotiator communication and thus contributes to the promising results of recent research on the topic of classifying negotiation outcomes based on negotiation communication data (specifically Sokolova and Lapalme 2012 and Twitchell et al. 2013) – our proposed differentiation between subjective and factual statements in negotiation thus can be seen as an interesting refinement step of these methods.

The application of Predictive Analytics in IS research is commonly seen as serving multiple different purposes: Whilst using prediction models obviously enables researchers to assess the predictability of a specific task, it can also contribute to a deeper understanding of the classification domain – via the unveiling of previously unknown patterns – and even result in the generation of new theory (Shmueli and Koppius 2011). Our research is closely related to the research area analysing the influence of emotional expressions in negotiations. As stated beforehand in section 5.2, this area is still subject to ongoing discussion and there is no prevalent theory that explains the effects of emotions in negotiations. We hope to contribute insights on this as well. On a methodological level, the most prevalent approach to analyse communication in negotiation research is content analysis, which involves a great deal of manual labour and is generally rather cumbersome. There have been attempts to employ machine learning techniques in order to automate this task but the results have been rather discouraging (Nastase et al. 2007) because the number of categories in order to provide a great level of detail is too high to still guarantee good classification performance. Nevertheless, Machine Learning methods could contribute not as a replacement, but as a simpler variation of content analysis with a lower granularity while at the same time, requiring considerably less manual effort – so the approach presented in the paper could be used to train simple content analysis models. On a further note, Sentiment Analysis is still a very young research area – thus relatively little is known about its application to complex do-mains such as e-negotiations where multiple actors exchange documents discussing a variety of different aspects. Our research aspires to contribute to methodological knowledge regarding this issue. Lastly, there are also practical aspects: Negotiations, as an important managerial task are oftentimes conducted using asynchronous electronic media such as e-mail. Various theories on Computer-Mediated Communication suggest that written asynchronous communication might not be an appropriate choice for such a complex task due to high levels of ambiguity in communication and therefore a higher likelihood of misunderstandings leading to impasse and renegotiations. A tool that is capable of conflict detection in e-negotiation could improve negotiators' knowledge about the specific negotiation situation by raising their awareness of potential impasse – and therefore preventing negotiations from failing that would have failed otherwise, which in turn would result in lower times to satisfying agreements, less renegotiations and therefore lowered costs.

However, there are several limitations to the study that have to be noted in order to accurately evaluate its contribution. First of all, generalizability of the domain-specific classifier might be an issue. We used negotiation data samples from a single negotiation case that has been negotiated many times by student participants in experiments. While training the classifying model, we manually generated a list of words that were deemed overly specific to the negotiation case – these words were taken as stop-words and excluded from the training of the classifier. Nevertheless, we cannot fully guarantee generalizability to other negotiation tasks – this will of course be subject to additional testing. Secondly, our data sample consisted of student negotiations under experimental conditions which may not accurately reflect "real" negotiation scenarios that occur in practice. However, samples with trained students – as all of our participants were (all participants receive a 90-minute introduction and briefing in advance of the experiment) – have been shown to be surprisingly reliable (Herbst and Schwarz 2011). Whilst real data from practice of course is desirable in the light of validity, obtaining this data in an amount which would be purposeful for the research attempted in the paper should unfortunately be close to impossible. B2B negotiation data is oftentimes seen by companies as being too sensible to leave the sphere of influence of the respective company. Regardless, the amount of the data used is reasonable compared to other research on classification in negotiation (Sokolova et al. use a considerably bigger dataset which had been collected over years using the NSS Inspire; Twitchell et al. 2013 use a small set of 20 divorce negotiations for their research).

Further research steps will involve classifying the polarity of the subjective utterances and then evaluating performance metrics for using the two classifiers in succession. The most crucial task that will be tackled in future research is to aggregate these sentence-level classifications to document level (i.e. a single negotiation message) and subsequently to full negotiation transcripts. Furthermore, the final prediction model has to be tested on partial negotiation transcripts in order to determine whether negotiation outcome can be accurately determined when the negotiation is yet unfinished – so that the model could predict potential negotiation failure while the negotiation is still on-going. Recent results indicate that classifiers retain most of their accuracy when the first half of the full negotiation transcript is available (Sokolova and Lapalme 2012).

Lastly, these automatically generated classifiers will then be benchmarked against classifiers based on a sentiment lexicon generated in previous research on the topic (Körner and Schoop 2014) in order to determine the best-performing model with regard to classifying negotiation outcomes. An-other step in further research could include identifying feature classes for potential targets of opinionated statements in negotiations: Any negotiation document involves multiple different attitudes of the author regarding different facets of the negotiation – for example the offer quality or the communication behaviour of the counterpart, the negotiation in general, the speed of progression in the negotiation etc. etc. A classifier that can differentiate between these different aspects of the negotiation in the form of opinion target categories should be able to provide more accurate assessment of the negotiation situation and the conflict level.

5.6. References

Adair, W., Brett, J., Lempereur, A., Okumura, T., Shikhirev, P. and Tinsley, C.H., Lytle, A.L. (2004). "Culture and Negotiation Strategy" Negotiation Journal 20, 87–100.

Baccianella, S., Esuli, A. and Sebastiani, F. (2010). "SentiWordNet 3.0: An Enhanced Lexical Re-source for Sentiment Analysis and Opinion Mining", Proceedings of the Seventh

International Conference on Language Resources and Evaluation (LREC'10), pp. 2200–2204.

Barry, B., Fulmer, I. S. and van Kleef, Gerben A. (2004). "I Laughed, I Cried, I Settled: The Role of Emotion in Negotiation", in M.J. Gelfand and J.M. Brett (eds.), The Handbook of Negotiation and Culture, pp. 71–94, Stanford University Press, Palo Alto, CA.

Bichler, M., Kersten, G. and Strecker, S. (2003). "Towards a structured design of electronic negotiations" Group Decision and Negotiation 12 (4), 311–335.

Bollen, J., Mao, H. and Zeng, X. (2011). "Twitter mood predicts the stock market" Journal of Computational Science 2 (1), 1–8.

Chesley, P., Vincent, B., Xiu, L. and Srihari, R. K. (2006). "Using Verbs and Adjectives to Automatically Classify Blog Sentiment", Proceedings of the 2006 AAAI Spring Symposium Series, Palo Al-to, CA, 27.-29.03.2006.

Cohen, J. (1960). "A Coefficient of Agreement for Nominal Scales" Educational and Psychological Measurement 20, 37–46.

Cortes, C. and Vapnik, V. (1995). "Support-Vector Networks" Machine Learning 20 (3), 273–297.

Cover, T. and Hart, P. (1967). "Nearest neighbor pattern classification" IEEE Transactions on Information Theory 13 (1), 21–27.

Cox, D.R.(1958). The Regression Analysis of Binary Sequences. Journal of the Royal Statistical Society 20(2), 215–242.

Curhan, J.R. and Pentland, A.(2007). Thin Slices of Negotiation: Predicting Outcomes From Conversational Dynamics Within the First 5 Minutes. Journal of Applied Psychology 92(3), 802–811.

Das, S. and Chen, M.(2007). Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web. Management Science 53(9), 1375–1388.

Druckman, D., Filzmoser, M., Gettinger, J., Köszegi, S.T., Mitterhofer, R. and Vetschera, R.(2012). 2.0² G eNerationS - Avenues for the Next Generation of Pro-active Negotiation Support, In Almeida, A., Morais, D., and Daher, S. (eds.), Proceedings of the 2012 Conference on Group Decision and Negotiation, Recife, 20.-24.05., pp. 82–84.

Forman, G. (2008). "Feature Selection for Text Classification", in H. Liu and H. Motoda (eds.), Computational methods of feature selection, Chapman & Hall/CRC, Boca Raton.

Friedman, R. A. and Currall, S. C. (2003). "Conflict Escalation: Dispute Exacerbating Elements of E-Mail Communication" Human Relations 56, 1325–1347.

Griessmair, M. and Köszegi, S. T. (2009). "Exploring the Cognitive-Emotional Fugue in Electronic Negotiations" Group Decision and Negotiation 18, 213–234.

Habermas, J. (1984), The theory of communicative action, Beacon Press, Boston.

Herbst, U. and Schwarz, S. (2011). "How Valid is Negotiation Research Based on Student Sample Groups? New Insights into a Long-Standing Controversy" Negotiation Journal 27 (2), 147–168.

Hine, M. J., Murphy, S. A., Weber, M. and Kersten, G. E. (2009). "The Role of Emotion and Language in Dyadic E-Negotiations" Group Decision and Negotiation 18 (3), 193–211.

Hofmann, M. & Klinkenberg, R. (2013), RapidMiner: Data Mining Use Cases and Business Analytics, Chapman & Hall/CRC, London.

Hsu, C.-W., Chang, C.-C. & Lin, C.-J. (2003), A Practical Guide to Support Vector Classification, Department of Computer Science, National Taiwan University, Taiwan

Hu, M. and Liu, B. (2004). "Mining and Summarizing Customer Reviews" Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2004.

Japkowicz, N. & Shah, M. (2011), Evaluating Learning Algorithms: A Classification Perspective, Cambridge University Press, Cambridge, MA.

Johnson, N. A., Cooper, R. B. and Chin, W. W. (2009). "Anger and flaming in computer-mediated negotiation among strangers" Decision Support Systems 46, 660–672.

Kanayama, H. and Nasukawa, T. (2006). "Fully Automatic Lexicon Expansion for Domain-oriented Sentiment Analysis", in D. Jurafsky and É. Gaussier (eds.), Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, Sydney, 22.-23.07.2006, pp. 355–363.

Kennedy, G. (2003). "Amplifier Collocations in the British National Corpus: Implications for English Language Teaching" TESOL Quarterly 37 (3), 467–487.

Kersten, G.E. and Lai, H.(2007). Negotiation Support and E-negotiation Systems: An Overview. Group Decision and Negotiation 16, 553–586.

Kersten, G. E. and Zhang, G. (2003). "Mining Inspire Data for the Determinants of Successful Internet Negotiations" Central European Journal of Operational Research 11 (3), 297–316.

Kiesler, S. and Sproull, L. (1992). "Group Decision Making and Communication Technology" Organizational Behavior and Human Decision Processes 52, 96–123.

Kiritchenko, S., Matwin, S., Nock, R. and Famili, A. F. (2006). "Learning and Evaluation in the Presence of Class Hierarchies: Application to Text Categorization", in L. Lamontagne and M. Marchand (eds.), Advances in Artificial Intelligence, pp. 397–408, Springer, Berlin, Heidelberg

Komorita, S. S. and Parks, C. D. (1995). "Interpersonal Relations: Mixed-Motive Interaction" Annual Review of Psychology 46, 183–207.

Lim, L.-H. and Benbasat, I. (1992/93). "A theoretical perspective of negotiation support systems" Journal of Management Information Systems 9, 27-44.

Liu, B. (2012), Sentiment Analysis and Opinion Mining, Morgan & Claypool, San Rafael, CA.

Maas, A. L., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y. and Potts, C. (2011) "Learning word vectors for sentiment analysis", in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, pp. 142-150.

Maron, M. E. and Kuhns, J. L. (1960). "On Relevance, Probabilistic Indexing and Information Retrieval" Journal of the ACM 7 (3), 216–244.

Morris, M. W., Nadler, J., Kurtzberg, T. and Thompson, L. L. (2002). "Schmooze or Lose: Social Friction and Lubrication in E-Mail Negotiations" Group Dynamics: Theory, Research, and Practice 6 (1), 89–100.

Mostafa, M. (2013). "More than words: Social networks' text mining for consumer brand sentiments" Expert Systems with Applications 40 (10), 4241-4251.

Nastase, V., Koeszegi, S. and Szpakowicz, S. (2007). "Content Analysis Through the Machine Learning Mill" Group Decision and Negotiation 16 (4), 335–346.

Pang, B. and Lee, L. (2008). "Opinion Mining and Sentiment Analysis" Foundations and Trends in Information Retrieval 2, 1–135.

Pennebaker, J. W., Mehl, M. R. and Niederhoffer, K. G. (2003). "Psychological Aspects of Natural Language Use: Our Words, Our Selves" Annual Review of Psychology 54, 547–577.

Polanyi, L. and Zaenen, A. (2006). "Contextual Valence Shifters", in J.G. Shanahan, Y. Qu, & J. Wiebe (eds.), Computing Attitude and Affect in Text: Theory and Applications: The Information Retrieval Series, pp. 1–10, Springer, Dordrecht.

Putnam, L.L. & Roloff, M.E. (eds.) (1992), Sage Annual Reviews of Communication Research: Communication and Negotiation, SAGE Publications, Thousand Oaks, CA.

Quinlan, J. (1986). "Induction of Decision Trees" Machine Learning 1 (1), 81–106.

Ruppenhofer, J. and Rehbein, I.(2012). Semantic frames as an anchor representation for sentiment analysis, In Balahur, A.*, et al.* (eds.), Proceedings of the 3rd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis, Jeju, Republic of Korea, 12.07., pp. 104–109.

Schoop, M., Jertila, A. and List, T. (2003). "Negoisst: A Negotiation Support System for Electronic Business-to-Business Negotiations in E-Commerce" Data and Knowledge Engineering 47(3), 371-401.

Schoop, M. (2010). "Support of Complex Electronic Negotiations", in D.M. Kilgour and C. Eden (eds.), Advances in Group Decision and Negotiation, pp. 409–423, Springer Netherlands, Dordrecht.

Shmueli, G. and Koppius, O. R. (2011). "Predictive Analytics in Information Systems Research" MIS Quarterly 35 (3), 553–572.

Sokolova, M. and Lapalme, G. (2009). "A systematic analysis of performance measures for classification tasks" Information Processing and Management 45, 427–437.

Sokolova, M. and Lapalme, G. (2012). "How Much Do We Say? Using Informativeness of Negotiation Text Records for Early Prediction of Negotiation Outcomes" Group Decision and Negotiation 21 (3), 363–379.

Spangle, M.L. & Isenhart, M.W. (2003), Negotiation - Communication for Diverse Settings, SAGE Publications, Thousand Oaks, CA.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K. and Stede, M. (2011). "Lexicon-Based Methods for Sentiment Analysis" Computational Linguistics 37 (2), 267–307.

Tang, H., Tan, S. and Cheng, X. (2009). “A survey on sentiment detection on reviews” Expert Systems with Applications 36, 10760-10773.

Tausczik, Y. and Pennebaker, J. (2010). "The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods" Journal of Language and Social Psychology 29 (1), 24–54.

Thelwall, M., Buckley, K. and Paltoglou, G. (2012). "Sentiment strength detection for the social web" Journal of the American Society for Information Science and Technology 63 (1), 163-173

Thompson, L.L. (2005), The mind and heart of the negotiator, Pearson Education, Upper Saddle River.

Thompson, L. L. and Nadler, J. (2002). "Negotiating via Information Technology: Theory and Application" Journal of Social Issues 58 (1), 109–124.

Toprak, C., Jakob, N. and Gurevych, I.(2010). Sentence and Expression Level Annotation of Opinions in User-Generated Discourse, In Hajic, J., Carberry, S., and Clark, S. (eds.), Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, Uppsala, 11.-16.06.2010.

Tumasjan, A., Sprenger, T. O., Sandner, P. G. and Welpe, I. M. (2010). "Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment", in W.W. Cohen and S. Gosling (eds.), Proceedings of the Fourth International Conference on Weblogs and Social Media, Washington, DC, 23.-26.05.2010, pp. 178–185.

Twitchell, D. P., Jensen, M. L., Derrick, D. C., Burgoon, J. K. and Nunamaker, J. F. (2013). "Negotiation Outcome Classification Using Language Features" Group Decision and Negotiation 22 (1), 135–151.

van Kleef, Gerben A., De Dreu, Carsten K. W. and Manstead, A.(2004). The interpersonal Effects of Emotions in Negotiations: A Motivated Information Processing Approach. Journal of Personality and Social Psychology 87(4), 510–528.

van Kleef, Gerben A., De Dreu, Carsten K. W. and Manstead, A. (2010). "An Interpersonal Approach to Emotion in Social Decision Making: The Emotions as Social Information

Model", in M.P. Zanna (ed.), Advances in Experimental Social Psychology, pp. 45–96, Academic Press, Burlington.

Walther, J. B. (1992). "Interpersonal effects in computer-mediated interaction: A relational perspective" Communication Research 19 (1), 52–90.

Walther, J. B. and Burgoon, J. K. (1992). "Relational Communication in Computer-Mediated Interaction" Human Communication Research 19 (1), 50–88.

Wiebe, J., Bruce, R. F. and O'Hara, T. P. (1999). "Development and use of a gold-standard data set for subjectivity classifications", in R. Dale and K. Church (eds.), Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics, Maryland, pp. 246–253.

Wiebe, J., Wilson, T. and Cardie, C. (2005). "Annotating Expressions of Opinions and Emotions in Language" Language Resources and Evaluation 39 (2-3), 165–210.

Wilson, T., Wiebe, J. and Hoffmann, P. (2009). "Recognizing Contextual Polarity: An exploration of features for phrase-level sentiment analysis" Computational Linguistics 35 (3), 399–433.

Yu, H. and Hatzivassiloglou, V. (2003). “Towards Answering Opinion Questions: Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences”, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing, Sapporo, Japan, pp. 129–136.

6. Study IV: Enriching Negotiation Transcripts with Sentence-Level Information

Enriching Negotiation Transcripts with Sentence-Level Information to Improve Negotiation Outcome Classification

6.1. Introduction and Motivation

The prediction of negotiation outcome based on the negotiation process is a fascinating, yet challenging problem. In traditional, face to face-settings, a human actor may develop a sense for the negotiation state at a given time, by interpreting the variety of social cues transmitted between him and his negotiation partner. After processing this interpretation, the human actor can adapt his strategic and tactical behaviour when he feels the need to do so, e.g. when his interpretation suggests that the negotiation process may end in an impasse. In electronic communication scenarios, this interpretive ability of a human actor is limited. Here, he is faced with fewer transmission opportunities for social cues and thus may be susceptible to misinterpretations of a given negotiation situation – which can lead to "accidental" failures of the negotiation which could have been prevented in a face-to-face scenario. Curiously, the abilities of Negotiation Support Systems to support a human actor in this respect have been rather limited. Traditional Decision Support functionalities in Negotiation Support Systems help to assess the utility a negotiator would obtain from different offers, based on a previously specified preference model. Furthermore, they focus on structuring and regulation of the communication, e.g. through the implementation of a negotiation protocol which lessens the likelihood of chaotic message exchange and thus ensures a clearer picture of what the current situation in a negotiation is. However, the interpretation of the negotiation situation as a whole is still left to the human actor, although his interpretation abilities may be hampered given the diminished number of cues available to deduct the negotiation state from while communicating via an electronic medium.

In the decision-making perspective as well as in the communication-oriented approach, researchers, therefore, have argued multiple times for more proactive forms of negotiation support, where a Negotiation Support System interacts with the user during the negotiation process, and supplies them with suggestions for potential offers or communication advice given the current negotiation situation (Kersten & Cray 1996, Braun et al. 2006, Kersten & Lai 2007, Vetschera et al. 2014). Interestingly, although research acknowledges this need for proactive support on all dimensions of the negotiation process, most of the actual implementations for proactive tools focus almost solely on the decision-theoretic perspective. In communication as well as in socio-emotional facets of negotiation research, advances towards proactive intervention rarely exist. In itself, this is not much of a surprise, given that a major problem in supporting a negotiator with this is usually that computerized systems struggle even more than human actors in evaluating components of the complex interaction process that a multi-attributive negotiation is. Kersten and Lai (2007) argue that any proactive Negotiation Support System needs the ability to monitor the negotiation process and to predict further likely actions by the negotiators in order to decide when to step in and offer support. Such an *entry point* may be more easily deducible when monitoring offer exchanges and concession paths from a decision-theoretic point-of-view.

With the continuing advance of Predictive Analytics, there may exist potentials for a system to deduce aspects from a given negotiation situation and to detect such entry points, without being limited to a decision-theoretic view only but with the ability to analyse negotiators' communication. Systems like IBM's Watson Project (IBM 2017) or Crystal (Crystal 2017) indicate that proactive means to support communication processes are a feasible option, and similar approaches could be applied in the context of electronic negotiation support. This paper provides a means to find such crucial entry points in incorporating methods from Machine Learning and Sentiment Analysis in order to predict negotiation success or failure automatically. To this end, we trained classification models that integrate sentence-level information from negotiation texts into a negotiation-transcript-level prediction of the negotiation outcome in order to answer the research question whether an integrated view on negotiators' communication on different levels of granularity can provide an improvement in negotiation result prediction quality.

The paper is structured as follows: First, we will provide theoretical background on previous research in classification and prediction of negotiations, then we will briefly introduce Sentiment Analysis as a means to classify negotiations (see previous work for a more elaborate discussion of this topic) – and later on focus on how sentence-level classifications can be employed as additional information for the more coarsely grained negotiation level (which we will also refer to as document level, staying in terms of Sentiment Analysis for the technical parts). Lastly, we will provide the classification models trained based on the theoretical foundations and discuss classification results.

6.2. Theoretical Background

6.2.1. Classification of Negotiation Data in Previous Research

Applying Data Mining methods to negotiation data has been repeatedly performed in previous research. The most common application instances focus on the decision-theoretic aspects, i.e. the analysis and prediction of factual offer exchanges based on negotiators' preference structures. Examples include the prediction of a negotiation partner's offers based on their concession behaviour and strategic choices earlier in the negotiation (Carbonneau et al. 2008, Lee & Ou-Yang 2009).

Kersten and Zhang (2003) used Logistic Regression, Decision Trees and Rule sets and Neural Networks to determine indicators of successful negotiations in negotiation metadata, and found that the timing of offers has a significant impact on negotiation success. Furthermore, they used the metadata to predict negotiation outcome and reached prediction accuracies of up to 75% (with 53% being the baseline value) on completed negotiations. Furthermore, they emphasize that the exchange of messages containing only relational content does not differ between successful and failing negotiation, hence they argue for a more concise analysis of communicational content of messages with the goal of negotiation outcome prediction.

A longer stream of research devoted to this goal has been conducted by Sokolova and colleagues (Sokolova et al. 2006, Sokolova & Szpakowicz 2007, Sokolova & Lapalme 2012). Starting with a comparison of communication data exchanged between face-to-face and electronic negotiations they experiment with different variations to predict negotiation outcomes. The first notable variation includes representation of frequent word patterns that represent strategic steps as classification features, where it was shown that these patterns result in a slight improvement over simply using the top 500 frequent words as representation of the negotiation data set (Sokolova & Szpakowicz 2007). The second main representation in this line of research bases classification on an evaluation of how informative the exchanged messages are – measured in the amount of terms expressing degrees, scalar words and comparative expressions in the negotiation data (Sokolova & Lapalme 2012). The results presented there indicate that negotiation outcome prediction based on communication means is indeed a possible and feasible task for a Negotiation Support System. Lastly, they also train models using only the first half of each negotiation and report a surprisingly low decrease in accuracy, hinting that reasonably good prediction of negotiation outcome need not be conducted *ex post* but also *during* an ongoing negotiation process (Sokolova & Lapalme 2012).

The notion of using communication data to identify and predict strategic steps during the negotiators' interaction is also prevalent in Smith et al. (2005), who employ Markov Chains to model strategic interaction sequences and to predict subsequent strategic steps at a given point in time. In a more recent approach, Twitchell et al. (2013) score speech acts in 20 divorce negotiations on an integrative/distributive scale and employ these scores in an overall negotiation classification model based on a two-step Machine Learning process, which is similar to the approach taken in the course of this paper.

6.2.2. Sentiment Analysis and Negotiations

Sentiment Analysis or the synonymously used term Opinion Mining has established itself over the last 15 years as an important stream of research in the field of Predictive Analytics.

In general, Sentiment Analysis is defined as the task of analysing "*people's opinions, sentiments, evaluations, appraisals, attitudes and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes"* (Liu 2012) As can already be seen from this definition, Sentiment Analysis encompasses a wide range of tasks as well as potential applications. It thus has been applied to a steadily increasing number of varying domains, including the tracking of political opinions and election prediction (Kim & Hovy 2007), brand sentiment tracking (Mostafa 2013), analysis of reviews on the web for products, restaurants, hotels and movies (Pang et al. 2002, Turney 2002).

Whilst the goals of applying Sentiment Analysis can be manifold, the most important ones can be distinguished into two streams (Pang & Lee 2008). The first stream uses Sentiment Analysis with a *summarization goal*, i.e. to reduce opinionated documents to a summary that captures the core opinions expressed, either on a general level, or broken down to different aspects of the item discussed in the text. The second stream, which is sometimes intertwined with the first one is a *classification goal*, i.e. to automate the analysis of opinionated documents, so that a computerized system can classify the direction of the opinion expressed given a transcript of the document. These classifications can range from detecting whether a positive or negative opinion is expressed to prediction of review scores in the form of star ratings or percentage scores and also the aspects of the item under review can be split up and several differing scores can be calculated to account for mixed opinions in a review (for example a camera where the resolution is praised but the price is perceived as a reason to give a negative overall opinion).

Tasks belonging to the classification stream usually involve a form of Machine Learning conducted over an exemplary set of documents from a given domain. To this end, the original textual data undergoes extensive preparation in order to transform it in a numeric representation that is both usable as input for Machine Learning classifiers and is an appropriate representation of the classification domain. For common Sentiment Analysis tasks these preparation steps involve the identification of polarized language, i.e. terms, utterances, or sentences that contain a subjective emotional evaluation, the identification of features or aspects of the classification domain, e.g. specific functions or attributes of the product to be evaluated, and decisions on how polarized language and features are to be represented, and whether and how they should be set in context to each other in the numeric representation. The data is subsequently converted into said representation and used to train the classifiers. Provided the classifiers reach satisfactory performance levels, they can henceforth be used to classify previously unseen data.

A secondary output of this classification process is rather of descriptive nature, showing which types of attributes are more important for the classification decision than others. This can allow further interesting insights on the classification domain itself. Similarly, some classification models allow for reasoning about the classification of specific single instances, providing information as to why a specific document was classified into a certain class. Provided that the common pitfall of data-driven approaches – to be of self-fulfilling nature – is circumvented, these prediction models can even identify patterns and regularities in the data that can contribute on further theory development regarding the problem domain (Shmueli & Koppius 2011).

This study assumes the position that the set of messages exchanged during an electronic negotiation can be viewed as a stream of opinionated documents, and thus provide a valid area to apply techniques of Sentiment Analysis. The main reasoning for this approach is rooted in the theoretical works on Computer-Mediated Communication (CMC). Being a complex, mixed-motive task, the conduction of negotiations over a lean medium such as electronic mail or a text-based Negotiation Support System poses severe challenges on the negotiators. In theoretical works on negotiation, communication serves an important role at the factual as well as the socio-emotional level. Offers are posed to each other, supported by factual information and verbally framed in order to reach individual as well as joint goals (Tutzauer 1992, Griessmair & Köszegi 2009). Furthermore, a core achievement of “good” communication in negotiations is to provide a framework in which information can be exchanged, commonalities can be identified and trust- and relationship-building between the negotiators take place – ideally resulting in a form of agreement on an outcome that is mutually beneficial (Schoop et al. 2010).

As classical CMC theories suggest, the reduction of channels to transmit social cues makes relationship-building and the formation of mutual trust more difficult, which in turn may impede negotiation outcomes (Lea & Spears 1992). Additionally, lean media introduce additional complexity to the negotiation task, as they increase the likelihood of misunderstandings (Daft et al. 1987). Hence, negotiators develop specific communication strategies as they adapt to this change in social perception. These strategies seek to reduce task and communication complexity and commonly include feedback loops (i.e. letting the counterpart know your evaluation of the situation), justification and persuasion to clarify one's own position, and affective-relational strategies (Te'eni 2001). Ultimately, these strategies correspond to exchanges of negotiators' perception of the negotiation situation via the written messages – negotiators communicate their state of satisfaction with the negotiation progress, trust development, offer quality and the potential to jointly resolve the negotiation task. As the number of channels through which such indications can be expressed is reduced (i.e. there are no visual or aural channels available), the written exchange becomes the sole focus based on which the negotiation situation is evaluated by the negotiators, constructing a mental perception of the likelihood of negotiation success given the specific context of the negotiation. These assessments are directly woven into the messages exchanged, either explicitly through lexical choice or implicitly through meta-attributes such as message and argumentation structure (Walter 1992). Hence, we suggest that these evaluations can be used as indicators for negotiation success and that methods of Sentiment Analysis provide a means with which these indicators can be captured and – integrated with knowledge from the negotiation domain – used to represent negotiation transcripts and to predict negotiation outcomes.

Sokolova and Szpakowicz (2007) already suggest to introduce features representing emotions, expressions of affect and opinions into the data. More specifically, Sentiment Analysis is briefly mentioned as a potential means to include information about emotions and subjective opinions into negotiation transcripts.

Previous research (Study II, this thesis) has employed a sentiment-based representation of negotiation data, with promising results but also with room for improvement. This study specifically seeks to integrate different layers of analysis, enhancing the negotiation data at hand with micro-level information based on sentence subjectivity classification described in the following chapters.

6.2.3. Micro-level Sentiment Assessment to Improve Document Classification

Early work in the beginnings of Sentiment Analysis already recognized the distinction between *objective* and *subjective* parts of opinionated documents. (Bruce & Wiebe 1999, Yu & Hatzivassiloglou 2003). Whilst objective statements focus on the provision of factual information about the topic of the opinionated document (e.g. fact-oriented descriptions of products, restaurants, movies etc. etc.), the subjective parts focus on actually presenting the authors subjective evaluation, which can but need not necessarily be linked to the factual information given. It has been argued that micro-level classification can lessen the influence of features that are not discriminative regarding the class decision such as - in the case of opinionated texts – expressions of neutral polarity or objective statements (McDonald et al. 2007). Furthermore, it allows a deeper insight on different aspects of the domain under investigation – whilst an opinionated document may be negative in general, single domain aspects may be attributed with non-negative polarities. With respect to this, Turney (2002) has argued for a distinction between opinionated evaluations of elements and sub-aspects of the given domain and evaluations that focus on judging the opinionated document as a whole. Hu and Liu (2004) use this idea to generate opinion summaries of product reviews separated into elements of the product.

This distinction in fact leads to sentence-level sentiment evaluation consisting of two separate classification steps, first distinguishing subjective from objective sentences and afterwards classifying their polarities. It is reasonable to assume that objective sentences by definition do not contain any polarities other than neutral. However, this perspective has come under scrutiny, since it is well possible to frame factual statements in a negative or positive fashion, depending on what the author wants to express while conveying the information (Feldman 2013).

For sentence-level evaluation of sentiments, two main approaches can be distinguished: Using pre-coded *sentiment dictionaries* to detect polarized expressions explicitly, and *supervised learning* approaches, similar to classic Text Classification.

Dictionary-based sentence classification is based on seed data that is partially (Hatzivassiloglou & McKeown 1997, Popescu & Etzioni 2010) or fully generated by humans (Bruce & Wiebe 1999). This seed list, consisting of adjectives and adverbs with a known semantic orientation (i.e. positive or negative) can then be expanded using either an example corpus of domain data or online dictionaries such as WordNet (Fellbaum 1998) that contain synonym and antonym information. For each word in the original seed list, synonyms are extracted from these dictionaries and assigned the same polarity as the original word, conversely antonyms can be extracted and be assigned the opposite polarity. The sentiment dictionaries created via this method can then be used to detect occurrences of the contained terms in the sentences one wants to classify. Overall sentence sentiment evaluation is then based on which sentiment polarities occur in the sentence, either using simple counting methods (Pang et al. 2002) or more elaborated weighting functions (e.g. Liu 2012, semantic orientation in Turney & Littman 2002 and Turney 2002).

Other sentence-level classification methods employ a traditional, *supervised learning* strategy, where previously labelled data is used to train a Machine Learning classifier. The data labels are usually either assigned manually by human coders or are generated/mined in an automatic fashion (e.g. Riloff & Wiebe 2003, Pang & Lee 2004), where example sentence data from the target domain is assigned an overall polarity without explicitly adhering to the specific words in the sentences. The idea is to refrain from learning discriminating aspects explicitly, but also to capture implicit expressions of sentiment, simply by using rather large data sets as training basis.

6.2.4. Meso-level Integration of Sentence-Level Information into Document-Level Classification

In a previous research step (Study III, this thesis), we evaluated subjectivity and polarity of expressions on sentence granularity, in order to provide information on the micro level. The next step is concerned with aggregating this information back to negotiation granularity and to employ it in negotiation-level classification in a purposeful fashion, which we describe as a task occurring at the meso-level.

Whilst in itself, sentence-level Sentiment Analysis can mainly be used to evaluate very short texts (e.g. tweets) and to summarize longer opinionated documents, a particularly challenging task is to integrate and aggregate its information (i.e. single evaluations of sentences) into the larger context of classifying the overall document the sentences stem from. The idea is to combine these levels of granularity so that the information from the micro-level can be employed on the macro-level to enrich the original date in order to achieve an improvement in overall classification accuracy (Täckström & McDonald 2011). A multitude of different methods has been devised to fulfil this integration task, ranging from *filtering* over *aggregation through annotation* and *scoring methods* to *fully integrated hierarchical classification* models.

Enrichment by filtering

A manifest approach to include sentence-level information in order to classify documents is to apply filtering techniques based on the sentence-level classifications. Yu & Hatzivassiloglou (2003) recommend reducing documents to opinionated sentences only and base the document classification on the subset only. Pang & Lee (2004) essentially followed a similar approach by dismissing objective sentences during the sentence-level classification phase. Their results showed that reducing the documents by less important parts improves coherence of the document representations and increases classification quality significantly. Similar approaches have been applied by Mao & Lebanon (2006). Toprak et al. (2010) elaborate on the concept of “polar facts”, i.e. statements that implicitly carry opinions although being objective in themselves. They argue that these polar facts can have a detrimental effect on classification accuracy, if they are not considered for further evaluation on an aggregate level.

Enrichment by annotations & scoring functions

The second possibility to enhance the negotiation documents with our sentence-level information is to include information about sentence-level subjectivity and polarity into the documents per annotation. In its plainest form, the whole negotiation transcript is retained and every sentence is annotated with the subjectivity/polarity classification result. These annotations can take on various forms, from simple tags that are appended to the sentence to weighting functions that may include additional information. For example, a weighting function could value sentences with positive or negative valence with respect to their position within the document, giving stronger weights to sentences that occur later on. Zaidan et al. (2007) use human annotations to provide rationales to explain why specific sections of a document are especially indicative of its overall polarity.

A common characteristic of opinionated documents is that generally, the final evaluation expressed is placed in the later sections of the document, most likely at the very end. Hence, approaches were developed that incorporate positional information about attributes into the feature representation, resulting in an improvement in classification accuracy (e.g. Raychev & Nakov 2009). For negotiation data, Twitchell et al. (2013) use a cascaded model where on micro level a trajectory score on an integrative/distributive scale is calculated. Most importantly they emphasize that a weighting function is purposeful for scoring the micro-level statements in order to put stronger emphasis to statements occurring later in the negotiation. The formula they implement is given as

$$\frac{x_i^3}{x_{total}^3}$$

where an utterance *x* at position *i* is weighted by the total number of utterances in the transcript, resulting in a score on a 0-1-scale.

Alternatively, a global score may be calculated from individual sentence values. Possible approaches include simple word counts of sentiment words detected in the document (Pang et al. 2002), summing up polarities over all sentences and determining overall sentiment based on whether the overall score is positive or negative (Liu 2012), combining the sums with a positional weighting factor (Zhang et al. 2009), or the usage of averaged scores of semantic orientations in the single sentences combined with additional weightings (Turney 2002).

Note that these scorings can either be introduced as an additional field in the classification process or to replace the original document-level evaluation altogether, thus leaving the realm of a Machine Learning-centric approach regarding document level classification. However, it is mostly preferred to combine the approaches with classic feature-based methods in order to improve classification accuracy (Turney 2002).

Integrated Models

Lastly, in rare cases researchers have attempted to integrate sentence and document-level information directly, by using combined models. Mao & Lebanon (2006) employ a complex aggregation of sentiment flows relying on conditional random fields and report increases in classification accuracy compared to stepwise vocabulary-based prediction. McDonald et al. (2007) employ a Viterbi-Algorithm in order to jointly classify document and sentence level sentences and report that sentence level-classification can be improved if the document label is known. Vice versa, the results they obtained were inconsistent. In the direction from fine to coarse, they rather recommend cascaded models, where classification results from sentence level are used as annotations in a document level classifier. Due to the scarcity of research on this subject, joint models are still subject to research.

6.2.5. Macro-level Negotiation Classification

The final step in classifying negotiation success consists of document-level classification of the resulting negotiation transcripts.

In general document-level classification projects the alternative approaches are similar to sentence-level classification. Supervised as well as unsupervised approaches may be employed (Liu 2010). Whereas supervised approaches treat the classification task as a Machine Learning problem with the goal classes being positive and negative, unsupervised approaches mostly employ formulas to calculate an aggregated sentiment score for the documents and base the classification on these scores. As on the sentence-level, feature and opinion lexicon generation plays an important role for both of the approaches. Such opinion lexica can be generated using sets of initial generic seed words. However, approaches where lexica are extracted from a corpus of domain data, or being generated by knowledge-engineering methods are more common. Oftentimes, aspects of dictionary-based approaches are combined with corpus learning to arrive at a final opinion lexicon. Turney & Littman (2003) use generic negative and positive seed words in order to learn a lexicon from a corpus of data. Subsequently, they expand their lexicon using association operators from the AltaVista search engine. Similar expansions can be done using online dictionaries such as WordNet or the General Inquirer Lexicon which contain synonym and antonym information.

As a specific type of Text Classification, various different forms of document-level representation have been discussed in Sentiment Analysis literature. Pang et al. (2002) suggest simple unigram and bigram representations in a multinomial bag-of-words-model and report good performances. Paltoglou & Thelwall (2010) experiment with variations of TF-IDF scoring methods to account for relative frequencies of terms occurring in the training corpus. However, a singular best representation on document level has not emerged up to this point and representational details remain strongly dependent on the specific domain and problem where classification means are to be employed upon. A particular issue at this stage is that of dimensionality of the data with respect to the overall data set size. When longer documents are transformed into term-document matrices, these matrices tend to be very sparsely populated, which can distort the classification results – especially if the overall sample set size is comparatively low. There exist several methods to alleviate this, either seeking to filter out irrelevant dimensions (*feature selection*) or to collapse multiple dimensions with shared meanings into one (e.g. *stemming, lemmatization or feature construction*). Feature construction is particularly interesting for domains that tend to be very specific – such as negotiations – since they allow the researcher to employ domain knowledge into the classification process. In its basic form, single terms occurring in the documents are subsumed and replaced by a feature category. These categories can either be manually constructed (e.g. the finance lexicon in Das & Chen 2004), automatically generated through algorithmic extraction processes (Popescu & Etzioni 2010), or Singular

Value Decomposition (Sebastiani 2002), or through any hybrid approaches between these two extremes (e.g. Sokolova et al. 2006, Fortuna et al. 2006)). This is in line with the general classification principle of reducing data set dimensionality via generalization to higher level concepts (Han & Kamber 2006). Whilst for complex topics, automatic extraction means from document collections have been attempted, it is still argued that these processes should ideally performed semi-automatically incorporating expert domain knowledge manually (Cimiano et al. 2005). The primary goal is to arrive at a set of feature categories (represented by lists of terms) that can be targeted and evaluated by sentiment expressions. These representations have been shown to increase performance in Text Classification and Sentiment Analysis tasks over classic word stem-based representations (Mullen & Collier 2004, Bloehdorn & Hotho 2006).

Aside from introducing domain knowledge and opinion lexica, the introduction of low-level contextual information has been shown to provide improvements for classification tasks (Kennedy & Inkpen 2006). The most prevalent construct with respect to contextual modification is that of negation terms, which are capable of inverting the meaning of a sentiment expressed. Furthermore, intensification words such as “really” or “very” can strengthen the evaluation expressed in a sentence (Wilson et al. 2006, Wilson et al. 2009). Incorporating these valence shifters in distinctive features further favours classification performance (Polanyi & Zaenen 2004).

6.3. Approach Taken in the Paper

Based on these three layers we discussed in the previous section, we devised a classification approach for electronic negotiations which this paper follows and which is shown in figure 12.

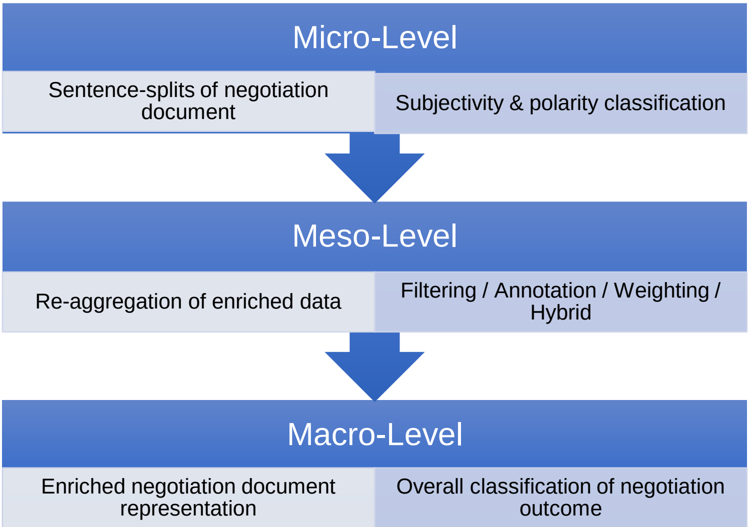

Figure 12: Multi-Layered Classification Approach on E-Negotiations

Negotiation transcripts are first split into single sentences through sentence-based Tokenization – as implemented in RapidMiner (Hofmann & Klinkenberg 2013), the tool we used to model our classification processes and to conduct the model training for this paper.

On the micro-level (i.e. sentence granularity) we first classify the sentences according to their subjectivity and polarity. To this end, a corpus of 28667 sentences from negotiation experiments were manually annotated by human coders in the context of a previous study (study III, this thesis). They were asked to classify the sentences with respect to two dimensions, first whether they think the sentence expresses a subjective assessment or whether the sentence simply conveys factual information. On the second level, each of these sentences was assigned a polarity, either positive, negative or neutral. Based on this manually annotated data, we generated training sets for three classification models, one for subjectivity classification, one for polarity classification of subjective sentences and one for polarity classification of factual sentences, the detailed results of this classification process is contained in said previous study. The strongest performing models were selected for this study to classify the negotiation data.

The next phase consists of the meso-level re-aggregation of the sentence-data in order to enrich the macro-level transcript. Here – as can be seen in figure 12 – four different general approaches were selected and experimented with. A filtering approach which simply omits those parts of the transcript that are less relevant for classification, an annotation-based approach which simply includes the sentence level classification results as additional features into the negotiation transcripts, a weighting approach which uses Twitchell et al's (2013) weighting function to emphasize statements occurring later in the negotiation, and lastly a hybrid approach which is a combination of the filtering and weighting method.

Various alternatives have been evaluated for the filtering approach, since it was not necessarily clear beforehand which approach would yield the best classification results. The first method included retaining only polarized statements, i.e. all opinions and factual utterances that were classified as either positive or negative. While this method seemed to work fairly well for complete negotiations, we found that for partial negotiations this greatly reduced the negotiation content, to the extremes of only retaining a single sentence. It seems that in the data set in general, polarized expressions are concentrated on the last half of the negotiation. Another reason is that most of our sentence-level classifiers suffered from biasing towards the majority class – neutral statements and would rather misclassify polarized statements as being neutral. After thorough evaluation, we retained all polarized statements, whether they had been classified as factual or subjective. This is contrary to classical approaches in Sentiment Analysis (e.g. Yu & Hatzivassiloglou 2003) where usually, factual statements are omitted altogether from subsequent analysis. However, research on emotions in electronic negotiations showed that factual statements can indeed convey emotional information based on content and wording of the statements – a notion which has been discussed in Sentiment Analysis as well (Toprak et al. 2010). The emotional information is of a more implicit nature, but nonetheless existing (Griessmair & Köszegi 2009). Additionally, factual statements conveyed in negotiations usually revolve around the decision-theoretic aspect of the negotiation and it seems manifest that evaluative cues regarding the decision-level conflict are exchanged within these factual statements.

As a second intuitive approach, we used the sentence-level classifications as a means of enriching the original data in that the class decisions for subjectivity and polarity were added to the original transcripts after each sentence as additional features. This allows to include the information while at the same time refusing to explicitly aggregate the sentence-level classes. The rationale for this approach was mainly that we tried to account for the difficulty in looking for an adequate aggregation function. In electronic negotiation transcripts oftentimes, positive sentiments in general outweigh negative sentiments, which means it is difficult to explore a direct role of these sentiment utterances but rather to study them in context of the sentence in which they are uttered. Since it is not directly intuitive to deduce context-oriented weightings for our sentiment scorings, we decided to let the model implicitly deduce the context from the sentence using the classic n-gram representation. Hence for this model, we also increased our granularity of n from bigrams to trigrams.

Regarding the weighting approach, we decided to follow Twitchell et al. (2013) and use the positional adaption function presented in their research, since results reported when using said weighting measure were satisfactory and representative for our classification problem at hand – the negotiation transcripts used by Twitchell et al. resemble the ones used in our research regarding length and domain in a sufficient fashion. The exact weighting procedure is as follows: We use the classification data from sentence level and introduce the label sets (*opinion, fact, positive, neutral, negative,* and the combinations *opinon_positive* etc.) as additional features into the term-document matrix. Occurrences of these features are firstly weighed by the confidence levels of the class decisions given by the sentence classifiers, then by Twitchell et al.'s positional weighting function. For aggregation to negotiation level, we averaged the resulting scorings for the features, resulting in a form of weighed frequency score. These weightings were subsequently used along with the other regular features as model training input.

Lastly, we opted for a combination approach of filtering and weighting, in order to reduce our sentence set to the potentially most distinguishing parts regarding negotiation classification while also providing positional information in order for the classifier to compensate for the reduction of data.

The final phase of the overall classification process includes the actual training and classification of the complete negotiation transcripts on the macro-level.

We applied a generalization step of the document level negotiation data using a previously generated Sentiment and Feature Lexicon (Körner & Schoop 2014), which replaces terms occurring in the text with the category associated to the term in our dictionary. Similar to Shah et al. (2004) the feature lexicon encompasses process-specific categories for the features (e.g. terms related to offer exchange, relationship-building or integrative/distributive behaviour) as well as linguistic constructs that can play an important role in modifying sentiment polarity, such as intensification words or negations (Polanyi & Zaenen 2004, Wilson et al. 2009). The sentiment lexicon has been trained on a corpus of negotiation data and encompasses a list of 762 terms with associated polarities (positive or negative). The result of the generalization process is a simplified representation of the document using our predefined category set only. Since this reductionist approach removes or generalizes rare terms occurring in the corpus, we opted for a simple means of representation of the documents, using a classical *bag-of-words* model, with a term frequency approach[6]. In order to account for interaction effects of the modifier classes introduced such as negations, we chose a bigram representation as the degree of granularity.

6.4. Classification Results

In order to adequately evaluate whether the approaches introduce an improvement compared to previous methods, there are different possible baselines to be selected. The simplest and most standard baseline would be to assume a trivial classifier that labels each data point with the majority class, resulting in a baseline accuracy of 50% for our balanced data set. The second baseline presented stems from a previous classification experiment, where only the feature generalization techniques mentioned in chapter 6.3 were applied (study II, this thesis). Here, we give the strongest performer as an additional stronger baseline which will yield answers to the question whether and how macro-level classification is improved by the micro-level enrichment. Comparisons of performance values reported by Sokolova & Lapalme (2012) or Twitchell et al. (2013) are difficult, since neither the data set nor the trained models are the same, thus any absolute statements about significant differences in performance would most likely be invalid (Japkowicz & Shah 2011).

For the micro-level processing, we use a combination of the best-performing classifiers for each single step according to their kappa statistic. Hence, Support Vector Machines were used for subjectivity classification, and Logistic Regression models were used for both polarity classification for factual statements and subjective utterances.

[6] TF-IDF has also been tried as a means of representation, but no with convincing results were obtained. It is likely that the strong dimensionality reduction to only a few categories has a detrimental effect on the proposed advantage of TF-IDF, i.e. accounting for important but *rare* terms in a corpus vocabulary.

On the macro-level, we used the following classifiers for each approach shown in the overall results table 18: A simple Naïve Bayes classifier, as used by Twitchell et al. (2013), a parameter-tuned Support Vector Machine (SVM), a parameter-tuned Logistic Regression model (LR), a parameter-tuned Decision Tree model (DT) and lastly, a tuned version of the k-Nearest Neighbour classifier (kNN). Parameter tuning was performed as follows: For SVM and LR we followed the guidelines presented in Hsu et al. (2003), using a Radial Basis Function (RBF)-kernel – as implemented in RapidMiner – and modifying parameters C and γ stepwise in a grid-based parameter search (C was increased stepwise from 2^{-5} to 2^{11}, γ from 2^{-15} to 2^{3} as described by Hsu et al.). The Decision Tree model applied is the C4.5 implementation of RapidMiner and was tuned similarly, by modifying the scoring criterion (*information gain* and *gain ratio*), minimum size for split (4, 8, 16, and 32), minimum leaf size (2, 4, 8, and 16), minimum gain (0.001, 0.01, 0.1, and 0.5), and confidence (0.1, 0.25, and 0.5) parameters. For kNN optimization we modified distance measures (*Euclidean, Chebychev distance*, and *Cosine Similarity*) and k (using odd values between 1 and 13). All final classification results were obtained using 10-fold Cross-Validation, where we made use of a sampling seed which ensures that for all our different validation runs, the same folds are used, which supports performance comparability of the different classifiers (Demšar 2006). For reasons of simplicity, we only report the best-performing classification model for each approach, selected by the kappa values obtained. Furthermore, table 18 shows performances for complete negotiation transcripts (indicated by the "_Full"-suffix), three quarters of negotiation transcripts (indicated by the "_TQ"-suffix), and half negotiation transcripts (indicated by the "_Half"-suffix). Performance values reported are classification accuracy, micro-averaged precision and recall, F-Measure – which is the harmonious mean of precision and recall (we used β=1 to weigh precision and recall equally, Manning et al. 2008), and the kappa statistic which measures chance-corrected agreement of the classifier with the actual class assignments (as introduced in Cohen 1960, Brennan & Prediger 1981).

Approach	Best Classifier	Accuracy	Precision	Recall	F	Kappa
Trivial_Base	NA	50.00	50.00	25.00	33.33	0
Previous_Base_Half	SVM	60.54	61.66	60.41	61.02	0.209
Previous_Base_TQ	NB	*65.65*	*66.52*	*65.68*	*66.10*	*0.313*
Previous_Base_Full	SVM	69.91	71.79	70.05	70.91	0.400
Filtered_Half	DT	*63.75*	*67.41*	*63.67*	*65.49*	*0.274*
Filtered_TQ	kNN	57.54	57.77	57.61	57.69	0.153
Filtered_Full	DT	76.41	77.98	76.38	77.17	0.527
Annotation_Half	SVM	60.13	60.49	60.30	60.39	0.206
Annotation_TQ	LR	58.41	59.68	58.78	59.23	0.175
Annotation_Full	DT	75.49	77.20	75.45	76.31	0.510
Weighting_Half	LR	59.24	60.09	59.43	59.76	0.188
Weighting_TQ	LR	61.00	61.58	61.02	61.30	0.221
Weighting_Full	LR	79.42	79.99	79.35	79.67	0.587
Hybrid_Half	SVM	55.47	56.51	55.76	56.13	0.114
Hybrid_TQ	LR	55.76	56.38	55.77	56.07	0.115
Hybrid_Full	SVM	*81.56*	*82.25*	*81.58*	*81.91*	*0.630*

Table 18: Classification Performances on the Data Sets

Classification performances were statistically compared using IBM SPSS Statistics. We compared classifiers in three groups, specifically full negotiations, three quarter negotiations and half negotiations. For performance comparison we follow the indications given by Dietterich (1998), *Dem*šar (2006) and Japkowicz & Shah (2011) respectively, where it is suggested that comparisons of multiple classifiers over single domains of data can be done using the individual performance scores for the single folds of the 10-fold Cross-Validations. This is mainly possible due to our folds being chosen identical for all classifiers as mentioned earlier. Using these folds however implies that basic assumptions necessary for conducting parametric tests (such as repeated-measure ANOVA) can easily be violated. Therefore, significance of differences in performance were assessed using the non-parametric Friedman Test (as in Japkowicz & Shah 2011) with the kappa statistic as main performance value. Significant differences are assumed starting from $p < 0.05$.

For full negotiations, a significant difference in performance was detected, $\chi^2(4)$ = *11.763, p = 0.019.* Therefore, as a post hoc test, Wilcoxon signed-rank tests with Bonferroni correction (to account for Type I error accumulation) were used, comparing each of the enriched classifiers to the baseline. Bonferroni correction modified our initial significance level from *0.05* to *p < 0.0125.* There were no significant differences between the baseline and the filtering approach *(Z = -1.599, p = 0.11),* between the baseline and annotation approach *(Z = -1,260, p = 0.208),* and between the baseline and weighting approach *(Z = -2.090, p = 0.037).* However, between the baseline and the hybrid approach, a significant improvement in performance was detected *(Z = -2.666, p = 0.008).* Lastly, the effect size for the significant difference was calculated (as described in Field 2005), resulting in $\mathrm{r}_{\mathrm{hybrid}}$ = *-0.596,* which represents a large effect.

For the three quarters data set, no significant differences in performance compared with the previous experiment baseline were detected, $\chi^2(4)$ = *3.306, p = 0.508.* Hence, an additional comparison with the trivial baseline was conducted, again resulting in no significant performance differences $\chi^2(4)$ = *8.144, p = 0.086,* suggesting that all of the approaches do not yield a consistent significant improvement over the baseline for three quarter negotiations.

For the halved data set, again, no significant differences in performance compared with the previous experiment baseline were detected, $\chi^2(4)$ = *4.990, p = 0.288.* Again, an additional comparison with the trivial baseline was conducted, detecting a significant performance difference $\chi^2(4)$ = *15.940, p = 0.003.* The post hoc Wilcoxon signed-rank test (as above, Bonferroni-corrected to *p < 0.0125),* revealed significant differences between the baseline and the filtering approach *(Z = -2.805, p = 0.005)* and between the baseline and annotation approach *(Z = -2.701, p = 0.007).* No differences in performance were detected for baseline-weighting *(Z = -2.395, p = 0.017)* and for the baseline-hybrid comparison *(Z = -2.095, p = 0.036).* Again, we calculated effect sizes resulting in $\mathrm{r}_{\mathrm{filtering}}$ = *-0.627* and $\mathrm{r}_{\mathrm{annotation}}$ = *-0.604,* both of which represent large effects. The overall results of these tests can be obtained from table 19.

Classifier	Result
Filtered_Half	Better performance than trivial baseline ($p < 0.0125$)
Filtered_TQ	No performance improvement over either baseline
Filtered_Full	Better performance than trivial baseline ($p < 0.0125$)
Annotation_Half	Better performance than trivial baseline ($p < 0.0125$)
Annotation_TQ	No performance improvement over either baseline
Annotation_Full	Better performance than trivial baseline ($p < 0.0125$)
Weighting_Half	No performance improvement over either baseline
Weighting_TQ	No performance improvement over either baseline
Weighting_Full	Better performance than trivial baseline ($p < 0.0125$)
Hybrid_Half	No performance improvement over either baseline
Hybrid_TQ	No performance improvement over either baseline
Hybrid_Full	Better performance than previous baseline ($p < 0.0125$)

Table 19: Significance testing overview

The rather unconvincing results of these tests are to be taken with a grain of salt – we used a very conservative version of the test, which suffers from the comparatively low amount of negotiations available. This introduces greater performance variations over the single folds, which in turn strongly affect the test's results. For the basic test to reach the significance level we assumed, a better performance in 9 out of the 10 folds was a precondition. Vice versa, the performance improvements actually detected as being significant can be assumed so with great confidence, hence it is safe to state that the hybrid approach resulted in a clear improvement over approaches that did not consider micro-level information explicitly.

For all the approaches tried in this paper, we obtain strong decreases in performances when using partial negotiation transcripts. Regarding the filtering and hybrid approach (containing the filtering step) this is in parts to be expected, since most utterances in the negotiation transcripts on the micro level are classified as neutral and thus omitted from further analysis. This often leads to negotiation transcripts being reduced to a minimum of content – in severe partial negotiation cases only to a single sentence. However, with respect to this it is even more surprising that the filtering-based classifier achieves slightly better performances on half negotiation transcripts than the other approaches. Most likely, sentiment-based classifiers rely more strongly on communication close to the conclusion of the negotiation.

Interestingly, using three quarters instead of half of the negotiation transcripts provides only very minimal improvement in classification quality for the weighting and hybrid approach, and even decreases for the other two models instead of a consistent improvement which would naturally be expected. This decrease is similar to what is reported by Twitchell et al. (2013). Presumably, information from the middle of the negotiation transcripts does not provide much information with regard to the actual outcome and can – at least for the classifiers – be strongly misleading.

All in all it can be stated that sentiment-based classifiers are strongly influenced by utterances occurring towards the end of the negotiation. Whilst for complete negotiation transcripts our research yielded similar results to other seminal papers on negotiation classification - Twitchell et al. (2013) reach 85% accuracy with a baseline of 60%, Sokolova & Lapalme (2012) report 71% accuracy with a baseline of 55% – our results decrease sharply on incomplete negotiation transcripts compared to previous research (Twitchell et al. (2013): 80% accuracy, Sokolova & Lapalme (2012): 71%). We assume that this is due to sentiment information only becoming a distinctive factor towards the end of the negotiation. Up to the point where one party's intent is to abandon the negotiation entirely, sentiment information in itself may not yield entirely accurate information about the negotiation result, whilst the performance on full negotiations reaches similar accuracies than is commonly reached using cascaded approaches in pure Sentiment Analysis tasks (e.g. McDonald et al. 2007).

Conclusively, it is difficult to hold up the postulate of early detection of success and failure being possible when using sentiment scores as the main information source for classification. Most likely, a method that integrates sentiment information with information about integrativeness/distributiveness (Twitchell et al. 2013) or informativeness (Sokolova & Lapalme 2012) criteria may yield better scores. However, sentiment-based assessments do provide additional and valuable information resulting in quite strong performances for complete negotiations (or transcripts that are nearly completed). Furthermore, we showed that micro-level analysis can bring additional value to macro level classification tasks –in this respect it seems that approaches that filter out unimportant statements and provide positional weighting information reach the best performances.

6.5. Conclusion

The superordinate research goal discussed over the course of this paper was to provide and present a classification model for electronic negotiation success that can be employed in a Negotiation Support System in order to enable possibilities of ex-ante proactive communication support, i.e. an entry point for dynamic supportive actions in ongoing negotiation situations. To achieve this, we used document-level classification techniques enriched by sentence-level information that decide upon potential negotiation success or failure. Our results show that for complete negotiation transcripts, micro-level enrichment and filtering of the data can indeed provide a reasonable and significant improvement over classification methods that only work on negotiation-level – which is quite promising given the immense complexity of the classification task at hand. If transcript length is reduced, our classification results deteriorate.

Likewise, the classification results should be handled with care in an NSS: They *do not* provide a confident and accurate prospect of the negotiation result – the prediction is always to be seen in comparison with the negotiation data used as training input to the model. Negotiations that have been predicted as failing may very well end up successful without further intervention. The predictive models presented here rather are meant to serve as a potential “activation mechanism” for proactive negotiation support in the form of giving specific advice to the negotiators, inquiring about the negotiation situation and using the input given by the negotiators as a starting point for actions that are beneficial to negotiation success.

One limitation in cascading models for multi-step classification is that errors made in the early classification stages are passed on as correct classification results to later models, resulting in a propagation of errors, which may lead to an impediment of classification results (McDonald et al. 2007, Yessenalina et al. 2010). Joint learning models for negotiation success may well be a feasible alternative for the approach taken in the paper, but only rare examples of research exist that attempt direct combinations in a joint model and resulting performances are inconclusive up to this point.

Also for the classifiers, it is hard to detect any potential ulterior motives behind the usage of polarized formulations, i.e. a strategic usage of emotions in order to influence the counterpart, either in the form of persuasion and schmoozing, or in the form of applying hard tactics that pressurize the counterpart through using negative emotions. This distinction task is, at the current state of research potentially not doable using existing approaches. There exists work on detection of tactical motives in language (such as deception detection in Zhou et al. 2004) and research on sarcasm and irony detection (e.g. Reyes & Rosso 2014), but on a more complex level it is unlikely, that a Machine Learning model oblivious of the negotiation context is capable to provide an accurate detection of strategic applications of affect.

Nonetheless, Predictive Analytics and classification in negotiations provides a rich field for future research. Early detection of success and failure may provide a valuable source of information for proactive NSSs that not only offer passive structuring, but actively reshape the negotiation process through user interaction if something goes awry. To this end, future research should seek to improve classification quality, maybe through employment of combinations from previous approaches – all of which seem to introduce a distinctive source of information for negotiation classification. Especially for partial negotiation transcripts, improvement of techniques is still needed. Lastly, future research should analyse how to convert the classification result into concrete means of advice to the negotiators – most likely in the form of further conflict diagnostics to obtain context information without which any possible NSS advice will only be generic in nature. We believe that such a proactive stance towards NSSs can introduce value for both researchers in the form of diagnostic information and conflict data as well as practitioners in reducing negotiation risk and thus ultimately unnecessary costs of renegotiations or failure.

6.6. References

Bloehdorn, S. and Hotho, A.(2006). Boosting for Text Classification with Semantic Features, In Mobasher, B., *et al.* (eds.), Advances in Web Mining and Web Usage Analysis: WebKDD 2004, pp. 149–166, Springer, Berlin, Heidelberg.

Braun, P., Brzostowski, J., Kersten, G.E., Kim, J.B., Kowalczyk, R., Strecker, S. and Vahidov, R.(2006). e-Negotiation Systems and Software Agents: Methods, Models, and Applications, In Gupta, J.N.D., Forgionne, G.A., and Mora T., M. (eds.), Intelligent Decision-making Support Systems, pp. 271–300, Springer, London.

Brennan, R.L. and Prediger, D.J.(1981). Coefficient Kappa: Some Uses, Misuses, and Alternatives. Educational and Psychological Measurement 41, 687–699.

Bruce, R.F. and Wiebe, J.(1999). Recognizing Subjectivity: A Case Study of Manual Tagging. Natural Language Engineering 1(1), 1–16.

Carbonneau, R., Kersten, G.E. and Vahidov, R.(2008). Predicting opponent's moves in electronic negotiations using neural networks. Expert Systems with Applications 34, 1266–1273.

Cimiano, P., Hotho, A. and Staab, S.(2005). Learning Concept Hierarchies from Text Corpora using Formal Concept Analysis. Journal of Artificial Intelligence Research 24, 305–339.

Cohen, J.(1960). A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement 20, 37–46.

Crystal 2017, https://www.crystalknows.com

Daft, R.L., Lengel, R.H. and Trevino, L.K.(1987). Message Equivocality, Media Selection, and Manager Performance: Implications for Information Systems. MIS Quarterly 11(3), 355–366.

Demšar, J.(2006). Statistical Comparisons of Classifiers over multiple Data Sets. Journal of Machine Learning Research 7, 1–30.

Dietterich, T.G.(1998). Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms. Neural Computation 10, 1895–1923.

Feldman, R.(2013). Techniques and Applications for Sentiment Analysis. Communications of the ACM 56(4), 82–89.

Fellbaum, C. (ed.)(1998), WordNet: An Electronic Lexical Database, MIT Press, Cambridge, MA.

Field, A.(2005), Discovering Statistics Using SPSS, 2nd Edition, SAGE Publications, London, Thousand Oaks, New Delhi.

Fortuna, B., Mladenic, D. and Grobelnik, M.(2006). Semi-automatic Construction of Topic Ontologies, In Ackermann, M.*, et al.* (eds.), Semantics, Web and Mining: Joint International Workshops, EWMF 2005 and KDO 2005, Revised Selected Papers, 121-131, Springer, Berlin, Heidelberg.

Griessmair, M. and Köszegi, S.T.(2009). Exploring the Cognitive-Emotional Fugue in Electronic Negotiations. Group Decision and Negotiation 18, 213–234.

Han, J. & Kamber, M.(2006), Data mining: Concepts and techniques, second edition, 2nd Edition, Elsevier; Morgan Kaufmann Publishers, Amsterdam, Boston, San Francisco, Calif.

Hatzivassiloglou, V. and McKeown, K.R. Predicting the semantic orientation of adjectives, Association for Computational Linguistics 1997 – Proceedings of the 35th Annual, pp. 174–181.

Hine, M.J., Murphy, S.A., Weber, M. and Kersten, G.E.(2009). The Role of Emotion and Language in Dyadic E-Negotiations. Group Decision and Negotiation 18(3), 193–211.

Hofmann, M. & Klinkenberg, R.(2013), RapidMiner: Data Mining Use Cases and Business Analytics, Chapman & Hall/CRC, London.

Hsu, C.-W., Chang, C.-C. & Lin, C.-J.(2003), A Practical Guide to Support Vector Classification, Department of Computer Science, National Taiwan University, Taiwan.

IBM Watson Tone Analyzer 2017, https://www.ibm.com/watson/developercloud/tone-analyzer.html.

Japkowicz, N. & Shah, M.(2011), Evaluating Learning Algorithms: A Classification Perspective, Cambridge University Press, Cambridge, MA.

Kennedy, A. and Inkpen, D.(2006). Sentiment Classification of Movie Reviews using Contextual Valence Shifters. Computational Intelligence 22(2), 110–125.

Kersten, G.E. and Cray, D.(1996). Perspectives on Representation and Analysis of Negotiation: Towards Cognitive Support Systems. Group Decision and Negotiation 5, 433–467.

Kersten, G.E. and Lai, H.(2007). Negotiation Support and E-negotiation Systems: An Overview. Group Decision and Negotiation 16, 553–586.

Kersten, G.E. and Zhang, G.(2003). Mining Inspire Data for the Determinants of Successful Internet Negotiations. Central European Journal of Operational Research 11(3), 297–316.

Kim, S.-M. and Hovy, E.(2007). Crystal: Analyzing predictive opinions on the web. Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning 2007.

Körner, M. and Schoop, M.(2014). Sentiment-Based Assessment of Electronic Mixed-Motive Communication - A Comparison of Approaches, In Brooks, L., Wainwright, D., and Wastell, D. (eds.), Proceedings of the UK Academy for Information Systems Conference, Oxford, UK, 08.04-09.04.

Lea, M. and Spears, R.(1992). Paralanguage and social perception in computer-mediated communication. Journal of Organizational Computing 2, 321–341.

Lee, C.C. and Ou-Yang, C.(2009). A neural networks approach for forecasting the supplier's bid prices in supplier selection negotiation process. Expert Systems with Applications 36, 2961–2970.

Liu, B.(2010). Sentiment Analysis and Subjectivity, In Indurkhya, N. and Damerau, F.J. (eds.), Handbook of natural language processing,2nd edn., pp. 627–666, Chapman & Hall/CRC, Boca Raton, FL.

Liu, B.(2012), Sentiment Analysis and Opinion Mining, Morgan & Claypool, San Rafael, CA.

Manning, C.D., Raghavan, P. & Schütze, H.(2008), Introduction to information retrieval, Cambridge University Press, Cambridge.

McDonald, R., Hannan, K., Neylon, T., Wells, M. and Reynar, J. (2007), “Structured Models for Fine-to-Coarse Sentiment Analysis”, in Carroll, J.A., van den Bosch, Antal and Zaenen, A. (Eds.), *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, Prague, 23.-30.06.2007*, ACL, pp. 432–439.

Mostafa, M.(2013). More than words: Social networks' text mining for consumer brand sentiments. Expert Systems with Applications 40(10), 4241–4251.

Mullen, T. and Collier, N.(2004). Sentiment analysis using support vector machines with diverse information sources, In Lin, D. and Wu, D. (eds.), Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, Barcelona, Spain, 25.-26.06., pp. 412–418.

Paltoglou, G. and Thelwall, M.(2010). A study of Information Retrieval weighting schemes for sentiment analysis, In Hajic, J., Carberry, S., and Clark, S. (eds.), Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, Uppsala, 11.-16.06.2010, pp. 1386–1395.

Pang, B. and Lee, L.(2004). A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts, In Scott, D., Daelemans, W., and Walker, M. (eds.), Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics, Barcelona, 21.-26.07.

Pang, B. and Lee, L.(2008). Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval 2, 1–135.

Polanyi, L. and Zaenen, A.(2006). Contextual Valence Shifters, In Shanahan, J.G., Qu, Y., and Wiebe, J. (eds.), Computing Attitude and Affect in Text: Theory and Applications: The Information Retrieval Series, pp. 1–10, Springer, Dordrecht.

Popescu, A.-M. and Etzioni, O.(2010). Extracting Product Features and Opinions from Reviews, In Kao, A. and Poteet, S.R. (eds.), Natural language processing and text mining, pp. 9–28, Springer, London.

Raychev, V. and Nakov, P.(2009). Language-Independent Sentiment Analysis Using Subjectivity and Positional Information, In Angelova, G.*, et al.* (eds.), Proceedings of the 2009 International Conference on Recent Advances in Natural Language Processing, Borovets, Bulgaria, 14.-16.09., pp. 360–364.

Reyes, A. and Rosso, P.(2014). On the difficulty of automatically detecting irony: beyond a simple case of negation. Knowledge and Information Systems 40, 595–614.

Riloff, E. and Wiebe, J.(2003). Learning Extraction Patterns for Subjective Expressions, In Collins, M. and Steedman, M. (eds.), Proceedings of the 2003 conference on Empirical methods in natural language processing, Sapporo, Japan, 11-12.07., pp. 105–112.

Schoop, M., Köhne, F. and Ostertag, K.(2010). Communication Quality in Business Negotiations. Group Decision and Negotiation 19, 193–209.

Sebastiani, F.(2002). Machine learning in automated text categorization. ACM Computing Surveys 34(1), 1–47.

Shah, M., Sokolova, M. and Szpakowicz, S.(2004). The Role of Domain-Specific Knowledge in Classifying the Language of E-negotiations, Proceedings of ICON'2004, pp. 99–108.

Shmueli, G. and Koppius, O.R.(2011). Predictive Analytics in Information Systems Research. MIS Quarterly 35(3), 553–572.

Smith, P.L., Olekalns, M. and Weingart, L.R.(2005). Markov Chain Models of Communication Processes in Negotiation. International Negotiation 10, 97–113.

Sokolova, M. and Lapalme, G.(2012). How Much Do We Say? Using Informativeness of Negotiation Text Records for Early Prediction of Negotiation Outcomes. Group Decision and Negotiation 21(3), 363–379.

Sokolova, M., Shah, M. and Szpakowicz, S.(2006). Comparative Analysis of Text Data in Successful Face-to-Face and Electronic Negotiations. Group Decision and Negotiation 15(2), 127–140.

Sokolova, M. and Szpakowicz, S.(2007). Strategies and language trends in learning success and failure of negotiation. Group Decision and Negotiation 16(5), 469–484.

Täckström, O. and McDonald, R.(2011). Semi-supervised latent variable models for sentence-level sentiment analysis, In Lin, D., Matsumoto, Y., and Mihalcea, R. (eds.), Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Portland, USA, 19.-24.06.2011, pp. 569–574.

Te'eni, D.(2001). Review: A Cognitive-Affective Model of Organizational Communication for Designing IT. Management Information Systems Quarterly 25(2), 251–312.

Toprak, C., Jakob, N. and Gurevych, I.(2010). Sentence and Expression Level Annotation of Opinions in User-Generated Discourse, In Hajic, J., Carberry, S., and Clark, S. (eds.), Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, Uppsala, 11.-16.06.2010.

Turney, P.D.(2002). Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews, In Isabelle, P. (ed.), Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, Philadelphia, PA, 07.-12.07., pp. 417–424.

Turney, P.D. & Littman, M.L.(2002), Unsupervised Learning of Semantic Orientation from a Hundred-Billion-Word Corpus, Technical Report, National Research Council Canada.

Tutzauer, F.(1992). The Communication of Offers in Dyadic Bargaining, In Putnam, L.L. and Roloff, M.E. (eds.), Sage Annual Reviews of Communication Research: Communication and Negotiation, pp. 67–82, SAGE Publications, Thousand Oaks, CA.

Twitchell, D.P., Jensen, M.L., Derrick, D.C., Burgoon, J.K. and Nunamaker, J.F.(2013). Negotiation Outcome Classification Using Language Features. Group Decision and Negotiation 22(1), 135–151.

Vetschera, R., Filzmoser, M. and Mitterhofer, R.(2014). An Analytical Approach to Offer Generation in Concession-Based Negotiation Processes. Group Decision and Negotiation 23, 71–99.

Walther, J.B.(1992). Interpersonal effects in computer-mediated interaction: A relational perspective. Communication Research 19(1), 52–90.

Wilson, T., Wiebe, J. and Hwa, R.(2006). Recognizing Strong and Weak Opinion Clauses. Computational Intelligence 22(2), 73–99.

Wilson, T., Wiebe, J. and Hoffmann, P.(2009). Recognizing Contextual Polarity: An exploration of features for phrase-level sentiment analysis. Computational Linguistics 35(3), 399–433.

Yessenalina, A., Yue, Y. and Cardie, C.(2010). Multi-level Structured Models for Document-level Sentiment Classification, In Li, H. and Màrquez, L. (eds.), Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, Cambridge, MA, 09.-11.10., pp. 1046–1056.

Yu, H. and Hatzivassiloglou, V.(2003). Towards Answering Opinion Questions: Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences, In Collins, M. and Steedman, M. (eds.), Proceedings of the 2003 conference on Empirical methods in natural language processing, Sapporo, Japan, 11-12.07., pp. 129–136.

Zaidan, O.F., Eisner, J. and Piatko, C.D.(2007). Using "Annotator Rationales" to Improve Machine Learning for Text Categorization, In Sidner, C., *et al.* (eds.), Proceedings of the 2007 Conference of the North American Chapter of the Association for Computational Linguistics, Rochester, NY, 22.04.-27.04., pp. 260–267.

Zhang, C., Zeng, D., Li, J., Wang, F.-Y. and Zuo, W.(2009). Sentiment Analysis of Chinese Documents: From Sentence to Document Level. Journal of the American Society for Information Science and Technology 60(12), 2474–2487.

Zhou, L., Burgoon, J.K., Nunamaker Jr., J.F. and Twitchell, D.P.(2004). Automating Linguistics-Based Cues for Detecting Deception in Text-based Asynchronous Computer-Mediated Communication. Group Decision and Negotiation 13, 81–106.

7. Overall Discussion and Conclusion

The overarching research question posed and analysed in this thesis sought to explore whether and how methods of Predictive Analytics and Sentiment Analysis can provide a means to detect electronic negotiation outcomes using the communication data exchanged as input.

In order to answer this question, the research applied over the course of the dissertation project first encompassed extensive means of analysis on a large corpus of experimental electronic negotiations. In the first step, we extracted frequent feature groups that make up aspects of the negotiation from our corpus and subsequently adjectives and verbs that modify these features. Using existing sentiment lexica, we allocated polarities to the sentiment expressions, resulting in a basic sentiment and feature lexicon for negotiations. In the following step, this lexicon was applied to generate a simplified representation of negotiation transcripts by replacing terms occurring in the dictionary with their category name. Based on this representation, Machine Learning methods were applied to a training set of negotiations, yielding first classification results for negotiation success or failure. The results showed that whilst the classifiers performed well on complete negotiation transcripts, classification accuracy depleted when only partial transcripts were used. Since the models' objective is to function well in an ongoing negotiation situation, the ability to classify partial transcripts is crucial regarding the research objective. Therefore, we tried to enhance the classification models in that we included information from a lower level of granularity – based on single sentences – into the negotiation data. To this end, human annotators classified each sentence of the initial corpus according to its subjectivity and its polarity. Using this sentence-level data, we trained sentence-level classifiers deciding on a) the subjectivity of sentences in negotiation transcripts, b) the polarity of factual statements and c) the polarity of subjective statements. Results indicated that subjectivity and polarity assessment of sentences can be done by these models, again with reasonable accuracy.

In the final step, we re-integrated the information of the sentence-level classification into our overarching document-level classification model. We experimented with and evaluated different means of fine-to-coarse aggregation. The conclusion drawn from these experiments is that micro-level analysis can improve macro-level classification of negotiations, especially when combined means of scoring and filtering techniques are applied. Unfortunately, this increase in classification quality only affected complete negotiation transcripts, which leads to the question whether Sentiment-based representations of negotiations provide particularly useful information regarding an early detection of success or failure. Thus, answering the research question of the thesis, it *is* possible to detect negotiation success or failure using the approaches discussed and presented in the four studies, but there exist several limitations regarding the point in time at which an accurate prediction can be made. However, even a means that can only provide a reasonable classification on the brink of failure can introduce purposeful support with respect to intervening during the actual escalation of conflict. Hence, the final sections of this thesis will wrap up the research in discussing supportive means available to an NSS, including the information obtained via the classification process.

7.1. Proactive Means of Electronic Negotiation Support

Braun et al. (2006) characterize the offer exchange during an e-negotiation as a phase where strategic behaviour is constantly revised by the actors, including the determination of concessions and revision of aspiration levels so as to adapt to the given negotiation situation. Furthermore, the provision of expertise and organising the negotiators' communication are identified as crucial tasks for Negotiation Support Systems. Given that an NSS is enabled to detect crucial points where the negotiation is about to be ended in impasse, various types of concrete means to support negotiators are possible, which will be discussed in this section according to previous research as well as what the findings in this thesis can provide as additional information. Whilst there exist distinctions and discussions about the types of analytic support an NSS may give (e.g. Spector 1997, Kersten & Noronha 1999), for an explicit communication perspective, only surprisingly few examples exist where concrete advise for negotiators is discussed. Most of the time for negotiations, this type of discussion is referred to as behavioural support and thus leading to the realm of (e-)Mediation. In this thesis, we distinguish three different possible approaches which can (and should) be combined by an NSS and subsequently assess how information obtained from classification processes can fit in.

7.1.1. Diagnostics

The first, and most important task an NSS should engage into when allegedly detecting a critical point is direct interaction with the negotiators, firstly in order to verify its assessment of a critical point, and secondly to obtain further information on the conflict situation. Direct inference of conflict diagnostics from the models created in this thesis will most likely very generic in nature and uninformed of the context.

A role model for conflict diagnostics that could be used in such a conflict situation is the Negotiator Assistant (Druckman et al. 2002, Druckman et al. 2004), an online conflict diagnostics tool. Originally devised for international negotiations, it offers guidelines for automatic conflict diagnosis, on which mediation advice is given to negotiators. Diagnosis in Druckman et al. (2002) is based on five categories: *Parties*, where questions about the nature of the relationship between the negotiators are asked, *Issues*, seeking to estimate negotiators' flexibility regarding the negotiated issues, *Delegation activities* investigating negotiation preparation and possible alternatives (BATNA), *Negotiation Situation*, mainly concerning external influences such as time pressure or third party interests, and lastly, *Process*, where information about the general approach of the negotiators (problem-solving vs. competitive) is gathered. Whilst these categories do not transpose directly to B2B negotiation scenarios such as the ones discussed in this thesis and do not contain diagnosis specific to communication activities, they can be used as a basic diagnosis framework, providing context information to the NSS for further advice.

Diagnosis Category	Content
Relationship Perception	Similar to *Parties* in Druckman et al. (2002) Subjective assessment of relationship quality Trust perception
Substantive Conflict Assessment	Similar to *Issues* in Druckman et al. (2002) Identification of conflicting issues Subjective Evaluation of concession behaviour
Strategic Approach Identification	Similar to *Process* in Druckman et al. (2002) Problem-solving vs. competitive
Communication Perception	Subjective evaluation of - partners' communication - own communication

Table 20: Example Categories for Communication Conflict Diagnosis

Similar to the categories defined there, table 20 gives an example of how communication conflict diagnosis could be applied in a system, consisting of four areas which may provide useful context information to an NSS, namely *Relationship Perception, Substantive Conflict Assessment, Strategic Approach Identification,* and *Communication Perception.* Note that these categories rather focus on subjective evaluations of the negotiators as opposed to seeking to gather an objective, factual state. This is due to the notion that communicative processes being an intersubjective construction that is based on subjective evaluations and social perception. Hence, investigating for subjective evaluations will provide a picture of the communication and negotiation state that is seen through the lens of the individual negotiators, and where it is possible to compare differing perceptions.

Relationship Perception is related to the *Parties* category in Druckman et al. (2002). Here, negotiators are asked to provide their interpretation of the relationship quality, as well as perceived trust towards the other party. Potential measures to be applied here are common trust measurements (as for example in Tzafrir et al. 2012). *Substantive Conflict Assessment* is related to the actual decisional conflict taking place and thus focused on negotiated issues as well as negotiators preferences towards these issues. Additionally, on the substantive layer, concession behaviour of the negotiators may be analysed, i.e. the negotiator is asked to evaluate concession quality of the counterpart and to juxtapose it with his own concession behaviour. Furthermore, issues should be identified that are perceived as being especially conflicting. S*trategic Approach Identification* directly corresponds to Druckman et al.'s (2002) *Process* category. Similarly, the negotiators are questioned about their general strategic approach to the negotiators in order to identify whether they employ a joint problem-solving or a competitive perspective. Lastly, it is of importance to let the negotiators give their own opinion regarding the communication situation itself, in the *Communication Perception* category. Here, subjective assessments of both the partner's and the negotiators' own communication behaviour are investigated.

It is important to note that the actual task of evaluating answers given by the negotiators and to transform them into dedicated advice for the specific negotiation situation is by no means a trivial task and has to be subject to extensive future research. Interestingly, neither communication diagnosis nor advice tools have seen substantial progress over the recent years although the foundations on which such a diagnosis can be based are readily available.

7.1.2. Advice and Recommendations

Dedicated advice to employ regarding communication is rarely discussed by itself, rather in conjunction with negotiation-strategic recommendation that seek to cause conscious alterations of negotiators' perspectives in order to manage the conflict situation more efficiently. These perspective changes in recent research specifically include creating awareness for interests and positive qualities of the negotiation counterpart, in order to facilitate a prosocial orientation, and similarly the promotion of a shared identity to emphasize the connectedness between the parties, which in turn improves agreement quality (de Dreu et al. 2006, Harinck & Druckman 2017). Whilst this type of advice is certainly purposeful, especially in a conflicting situation, it does not focus explicitly on specific communication approaches, i.e. the transformation of the mindset created in the negotiator into negotiation messages is still a task that the negotiator has to manage alone.

From an organisational CMC point of view though, there exist models that emphasize communication in the conduction of complex tasks, such as business negotiations. Especially the Cognitive-Affective Model of organisational communication (Te'eni 2001), discusses several communication strategies that should be facilitated and supported in order to manage communication complexity, namely Contextualization, Affectivity, Control-testing and adjusting, Control-planning (Predetermination in Zaidman et al. 2008), Perspective Taking (Involvement in Zaidman et al. 2008) and Attention Focusing. Each of these strategies can be supported by a system in that it either provides structures or gives concrete recommendations that facilitate their usage in negotiations.

Contextualization means to provide contextual information, i.e. reasoning to support one's claims and to consciously share information that may not be known to the recipient to clarify a communication situation and to ensure a common understanding of the communication task. Here, a communication supporting NSS could motivate negotiators to exchange more reasoning and information about issues that have been identified as critical (e.g. in the diagnosis phase), or perhaps even offer a structured form of argumentation for these issues alone in order for the negotiators to clarify their positions and halt the discussion about other aspects during this phase. This allows negotiators to focus their cognitive capabilities on the most critical points, thus reducing complexity until the issue is resolved or until it becomes clear that the issue can only be resolved in context of other issues. This reductionist approach also would seek to counter the tendency of CMC actors to provide too much context that is not needed at this point (e.g. lengthy explanations for non-critical issues that are given simply because they are easily available, Katz & Te'eni 2007). Likewise, complex issues can be separated into parts which are then discussed in succession, as suggested in Druckman et al. (2004).

The methods applied in this thesis best correspond towards Te'eni's affectivity strategies. These strategies include the display of emotions and moods in the messages, providing statements of subjective evaluation regarding to the topic of communication. Especially in scenarios with high affective complexity, where attitudes, subjective evaluations, and relationship-building plays a crucial role for communication success, affectivity strategies are considered important. However, from a cues-filtered out point of view, asynchronous media (as the domain of investigation in this thesis) is not well-suited for affectivity strategies (Te'eni 2001). However, appropriate medium management can enable negotiators to convey their emotions and feelings – which can be supported using for example the content of the sentiment dictionaries created in this thesis. It would for example be possible to detect terms that are associated with negative affect and motivate the negotiators to think about what exact emotion they want to convey; similarly, alternative terms could be suggested either based on the sentiment dictionary or on online thesauri. From the classifiers, feature-sentiment combinations associated with negotiation failure, such as the examples listed in study II could specifically be proposed to be altered and reformulated. Whilst this certainly will not guarantee negotiation success, on a broader level, negotiators could be stimulated to employ a more conscious emotional style in their messages that corresponds to their actual interpretation of the situation and is less prone to be misinterpreted in a negative fashion – a downside inherent to asynchronous, written media (Byron 2008). More generally, a positive emotional writing style could be suggested including avoidance of negation constructs - which have been identified as a common theme in failing negotiations in studies II and IV - and motivating awareness of intensifier usage, which further amplify sentiment expressions and are more likely to elicit a retaliating response from the counterpart.

Control by testing and adjusting includes employing checks and control mechanisms in order to elicit feedback from the counterpart, e.g. clarifying exchanges about terminologies and interpretations of a message. According to the feedback received, communication is adapted. Because of the context-intensity of this strategy, it is difficult to give directed communication advice to negotiators here. On a general level, an NSS could actively suggest employing the strategy in that it proposes to negotiators to explicitly ask for feedback and clarification, or – on the substantive layer – suggest to exchange preferential information and issue priorities.

Similarly, control by planning entails to prepare one’s own communication style depending on likely future developments of the communication process. Therefore, different strategies should be formulated with the general aim of being consistent in one’s communication. Being directly related to control by testing and adjusting, this strategy is likewise dependent of feedback mechanisms and context-sensitive developments in single negotiation instances, hence again it is difficult for a system to give more than general recommendations. Zaidman et al. (2008) suggest for systems to prompt the user when generally expected courtesy responses are omitted such as greetings and closures of messages. It could also be feasible to include a component into an NSS which allows to trace and monitor negotiators’ planning processes – i.e. prior to the negotiation the negotiators are asked to formulate plans in the system for various developments of the negotiation; during the negotiation the system reminds the negotiators of these communication plans and potentially asks for refinement if a plan is not feasible anymore.

Perspective taking means to consciously think about the counterpart’s values, beliefs and perspectives, and to communicate understanding or agreement/disagreement accordingly.

This is closely related to the promotion of shared identity and other affirmation discussed in Druckman et al. (2004) and Harinck & Druckman (2017). Regarding this strategy, aside from recommending to take the other's perspective into account, an NSS could focus on reminding negotiators to actually communicate considerations and understanding regarding their counterpart's position explicitly. Oftentimes in using asynchronous media, actors will assume that their counterpart is aware of them actively considering the counterpart's perspective and will omit communicating this, in the worst case resulting in perceptions that the counterpart's perspective is not taken into account at all – which in turn leads to more distributive behaviour and ignorance by their counterpart.

Lastly, attention focusing encompasses strategically directing the information processing of the receiver of a message, which is usually be achieved by highlighting and capitalization usage as necessary as well as structuring the components of the message sent. Existing communication support (such as in Schoop 2010) already includes mechanisms of structuration and attention focusing, in highlighting issues in the messages thereby simplifying the identification of paragraphs that discuss issues of specific interest. Likewise, NSSs could suggest structural frameworks for messages, consisting of segments predefined in templates (such as for example Greeting → Evaluation of Counterparts offer → Addressing of Issues → Overall evaluation of the negotiation process → Closure) that provide guidance in message composition and structuring. Issue sequences furthermore could for example be modified according to preference structures of the negotiators or according to difficulty of the issues to be resolved.

Communication Strategy	Supportive Function
Contextualization	Motivate explicit explanations for issues Resolve critical issues in distinguished phase Selective Contextualization – provide explanations only for critical issues
Affectivity	Recommendation of alternative termini General suggestion of positive emotional style Awareness of negation and intensifier usage Motivate a conscious management of affectivity
Control-Testing and adjusting	Suggest asking for feedback Suggest to clarify terminologies and definition
Control-Planning	Reminder for courtesy responses Tracing and monitoring of communication plans
Perspective Taking	Recommend to communicate perspective taking
Attention Focusing	Structural message frameworks

Table 21: Communication Strategies (Te'eni 2001) and NSS support potentials

Table 21 gives an overview of the respective strategies and advice means that could potentially be implemented in Negotiation Support Systems. In general, it can be argued that any design facilitating proactive communication support should be built under the consideration to enhance understanding, facilitate information exchange and promote relationship development between communicators (Te’eni 2006). Furthermore, it should not produce additional cognitive load for the negotiators, but rather reduce and structure existing cognitive complexity, which is high enough in the negotiation task itself. Therefore, the recommendations discussed above should be applied with care and only if desired and needed by the negotiators as a result of continuous monitoring of the negotiation process (Te’eni 2006). Here, the combination with diagnostic information at critical points in the negotiation becomes especially important, which could be used as a means to identify which type of support to be enabled to the negotiators, for example by eliciting the communication area that is most problematic and accordingly supporting only strategies that are directly related to it as opposed to enabling all supportive means all the time. For example, if it becomes clear that a counterpart’s explanations for certain issues are perceived as being insufficient, contextualization and control by testing strategies could be suggested and supported.

7.1.3. Visualization

Visualization methods are integral in simplifying complex matters for users of a system, to communicate the system’s assessment of the situation and to promote deeper understanding and reflection by the negotiators. In the Decision Support component of NSSs, visualization methods have been shown to thus facilitate improvements regarding negotiation outcomes (Gettinger et al. 2012). Regarding communication support, visualization elements used in Negoisst such as the semantic enrichment provide clarity where exactly in the negotiation messages an issue was specified and thus remove ambiguity from the interaction (Schoop 2010). By nature, visualizations employ rather a passive supporting role, where the only proactive aspect a system could employ would be to automatically highlight visual elements it deems as useful in the current situation. Nonetheless, the importance of visualizations in providing additional assistance is undisputed in NSS research as well as in Sentiment Analysis research, where oftentimes information is very fragmented over the message data and can be meaningfully aggregated.

Intuitively, most techniques in sentiment visualization revolve around the usage of sentiment dictionaries as created in Study II in this thesis. From a general perspective, sentiment expressions could be highlighted in the message editor, giving the user visual indications of how positive or negative his message will be perceived and providing options to revise messages in order to alter the impression according to his wishes. This can also be used before a dedicated detection of conflict by the classifiers, so as to contribute to preventing a conflict escalation altogether. Similarly, thematic message structuring could be implemented by highlighting feature categories. From a classification point of view, sentence constructs that are associated with negotiation failure could be highlighted, so as to leak classifier information into the message creation process. Mohammad (2012) discusses summarization of sentiment expressions in pie charts and bar graphs, indicating distributions of sentiment polarities over aspects of the domain under investigation. In an NSS, similar visualizations could be employed for sentiment expressions split up by feature categories, as exemplified in Figure 13.

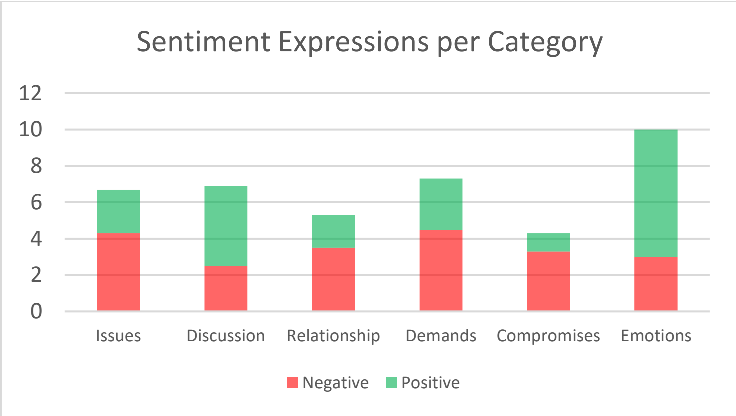

Figure 13: Example Sentiment Visualization per Feature Category

Since in a specific negotiation the names of concrete issues are known, it would also be possible to aggregate sentiment expressions per negotiation issue in order to give a visual indication which issues are potentially problematic. To this end, grammar dependency trees (as in the dictionary generation in study II) could also be employed, since compared to simple co-occurrence detection over sentences they give a more accurate information of sentiments directly modifying feature words and issue terms. Note that the classification models in studies II-IV do not employ dependency trees since the resulting term-document matrices were too sparse and many implicit evaluations were omitted – nonetheless, for visualization purposes dependency trees provide a useful foundation due to their low rate of false positives (i.e. erroneously detecting a feature-sentiment-connection where there is none).

Lastly, the message sequences in NSSs allow for timeline-based representation, allowing to show sentiment developments over the messages. Similar to concession path visualizations used in Decision Support (e.g. Gettinger et al. 2012), these visualizations could allow negotiators to interpolate future conflict development and even asymmetric conflicts where the sentiments of the parties strongly differ from each other.

7.2. Conclusion

7.2.1. Critical Assessment of Sentiment Analysis in Negotiation Classification

As a conclusive critical assessment of the approach in the paper, it can be said that electronic negotiation data provides a valid application scenario for sentiment-based classifiers. Nonetheless, sentiment-based information mainly seems to be an influential factor towards the end of the negotiation process. Whilst this is similar to what has been reported on the influence of positive and negative emotions on the negotiation outcome (Hine et al. 2008), it contradicts the notion that early indicators of negotiation success or failure exist in the communication data (e.g. Simons 1993, Adair & Brett 2005, Curhan & Pentland 2007), or at least argues towards the fact that any of these early indicators in communication are *not* based on negotiator sentiments. This is an interesting result, since Sokolova & Lapalme (2012) as well as Twitchell et al. (2013) discuss classification mechanisms based on polar and emotional statements as possible extension to their approaches. Presumably, and also shown by their research, there exists a plethora of other scales and aspects relevant to negotiation success which need to be integrated in order to provide reliable early prediction methods. However, especially in the later parts of negotiations, Sentiment Analysis can serve as a major pillar in this respect.

Obviously, negotiation results can never exactly be observed from communication data alone, even with powerful enough methods, perfect data representation and sufficiently large data sets. This is an inherent problem of the “context-lean” method of observation. As has been stated multiple times in previous research, communication transcripts alone can never allow for a complete understanding of the social interaction (e.g. Taylor et al. 1996), which always has to be studied considering the manifold contextual factors influencing *which* statements are communicated, *how* they are interpreted and perceived by the other party and how exactly this will influence the progress of the negotiation. Nonetheless, under these rather severe limitations, the methods presented in this thesis offer a valuable proof-of-concept of automatic interpretation of negotiation data.

Regarding the methodological aspect, this thesis contains contributions towards the knowledge base in the Design Science sense in that it presents classification models as design artefacts (i.e. a level 1 contribution according to Gregor & Hevner 2013) and furthermore demonstrated a model paradigmatic approach for incorporating Sentiment Analysis methods in a novel application domain, hence the design principles and approaches to electronic negotiation communication can be considered a level 2 contribution according to Gregor & Hevner (2013).

7.2.2. Limitations

As in any methodological choice of analysis, Data Mining and Predictive Analytics possess unique characteristics that may limit the validity of the results obtained through application of the method. Kersten and Zhang (2003) note the basic difference in sampling to confirmatory statistics is that in Data Mining the data set used for analysis is assumed to be either the whole population or a large fraction of it, which does not hold true for the set of negotiation data used over the course of this dissertation. Our data set stems from negotiation experiments with trained students that were conducted over several years at the University of Hohenheim. These negotiations were embedded in a B2B-Setting, usually in the form that two companies' representatives negotiated a contract between the two companies in a mixed-motive scenario with both distributive and integrative aspects to be negotiated. Hence, we assume that our results therefore may not generalize to entirely different types of negotiation as for example conflict resolution processes or international negotiations on a political level. Nonetheless, for the B2B-case it has been shown that data obtained from experimental setting with trained students is sufficiently similar regarding performance to the data obtained from professional negotiators in a real-world setting (Herbst & Schwarz 2011).

Likewise, the thesis focused only on a reductionist view on negotiation outcomes (success/failure). A more fine-grained distinction of outcomes and the consideration of other outcome aspects such as socio-psychological outcomes or fairness perceptions could as well be studied. In this case, the data set was not fit to employ such a distinction due to size restrictions and unavailability of socio-psychological evaluation data for the respective negotiations. Since communication specifically impacts long-term relationship development, satisfaction and perceptions of fairness (Greenhalgh & Chapman 1998), variables measuring these constructs may provide better performances than those observed in this thesis. Another dimension which has only implicitly been taken into account is the decision making perspective on negotiation development, specifically the influence of offer quality as for example indicated by utility values given by an NSS according to a previously defined preference model. Obviously, showing these utility values to negotiators should influence their decision making behaviour, especially when constructing conceding counteroffers. However, as the possible range of utility values varies with the given negotiation setting, agenda and the definition of the preference model on a case-by-case basis, it is a non-trivial task to unify these values across different negotiations from varying settings in order to achieve comparability and to construct variables that provide distinctive capabilities between negotiation success and failure that can be interpreted by a Machine Learning model. Even if the utility values of a given training/test set were to be unified, it is an almost impossible task for the classifier to put utility values in a previously unseen negotiation into perspective, hence most likely using plain utility values can even distort classification results and impede generalizability and applicability to negotiation problems that the classifier was not trained on. In classic behaviouristic research settings this is not as severe a problem, because the comparability towards an experimental group is always given, and utility ranges often are unified via the specific experimental setting (e.g. by using pre-constructed preference models). However, in classification tasks, absolute utility values should rather be omitted from the sets in favour of relative variables, for example relative concession sizes, differences in relative concession sizes between the negotiators etc. in order to ensure that the classifiers are able to deal with negotiation settings that have not been part of their specific training sets. Conclusively, it is difficult for classification models to capture specific contextual factors, be it utility ranges or more complex factors such as power distribution between the negotiators, initial conflict sizes, psychological negotiation dynamics or changing external influences on the negotiators that all may influence the negotiation outcome. This context-obliviousness is certainly one of the strongest limitations of classification quality for negotiations, which, in its current state is unlikely to outperform actual human judgement from a neutral perspective, for this, representational means of a negotiation situation, Machine Learning models and our understanding of success-influencing factors in negotiations yet has to be improved considerably.

The options of giving advice and recommendations discussed in chapter 6.1.2 also introduce an interesting inherent problem in that if any advice given is followed by negotiators, this will distort the results of the classification methods themselves by effectively removing their own indicators for accurate classification. This may cause any detection of negotiation outcome to fail although the negotiation is actually on the road to impasse. Therefore, advice and recommendation based on the classifiers information should be given with care regarding that it can lead to the classifiers being oblivious of alternate paths to failure that are not contained within the realm of information given to the classification models via sentiment dictionaries.

7.2.3. Future Research Directions

This chapter sums up potential future research directions that can be accessed using the results of this thesis in order to arrive at the research goal of employing proactive communication support in e-negotiation systems.

A wide area of research to be explored consists of fine-tuning the classification techniques explored in the thesis, in order to achieve better classification performances especially on partial negotiation transcripts, which would allow a detection of potential negotiation failure as early into the negotiation as possible. As already suggested in study IV, integrations with other approaches to negotiation classification could provide additional value, since as it seems, each of the approaches explores different content of the negotiation messages and yet arrives at reasonable classification accuracies. Sentiment Analysis and Text Classification as research fields also offer further areas that could be included: Currently, all the models presented take a naïve approach in that they take everything that is written as actual conveyed meaning. This renders the classifiers incapable of detecting complex figures of speech such as irony and sarcasm, which are rhetorical devices known to impact negotiation behaviour (Schroth et al. 2005). Being able to detect such complex constructs, which extend beyond simple connections of words would most likely lead to an improvement of classification performance. However, this task is subject to ongoing research within Sentiment Analysis and would most likely imply to enable the classifier to process complex, paralinguistic cues (such as for example those discussed in Reyes & Rosso 2014).

A related topic, which already has been explored in the context of negotiations is detecting deceptive behaviour of the negotiators (Zhou et al. 2004a, Zhou et al. 2004b, Fuller et al. 2013). Fuller et al. (2013) discuss sets of linguistic and paralinguistic constructs that theoretically allow for automated deception detection in negotiations, employing word quantities, diversity, affective utterances, usage of pronouns, uncertainty indicators, words indicating cognitive information and activation. Although their models suffer from issues due to reliability and domain specificity diminishing the validity of the identified constructs, the results yield promising approaches for future research where the identified indicators could be employed in classification approaches.

Regarding the models explored in study IV, there is a lingering issue of propagating classification errors when using cascaded fine-to-coarse approaches, as also discussed in the limitations section. Since the inclusion of sentence-level information nonetheless improves classification quality, we suggest that future research should seek to improve on existing integrated classification approaches (an example is the employment of a Viterbi algorithm in McDonald et al. 2007) in order to make them a feasible approach for complex classification tasks, reducing the risk of classification errors.

A concept that has only received very brief attention in this thesis is the notion of introducing an opinion holder, i.e. to introduce a separation of the opinions of each of the negotiators.

Being able to identify the distinct opinions of the negotiators could enable the classifier to detect asymmetric conflict situations, where only a single party is close to end the negotiation whilst the other party does not perceive as strong a conflict and would still continue negotiations. Especially in asynchronous CMC scenarios, this situation is not entirely uncommon, since negotiators sometimes fail to perceive the communicated dissatisfaction of the other party, omit to communicate their own dissatisfaction or interpret such statements as a tactical spiel without consequences for the negotiation result. Liu (2012) discusses the applicability of Named Entity Recognition in order to detect opinion holders, in negotiation data personal pronouns could for example be employed, enabling opinion holder identification depending on the authoring negotiator of a message. Still, this task is subject to ongoing research and especially in interaction and discussion sequences open to exploration. Specifically, for negotiations, opinion holder identification vastly increases classification complexity due to expanding the feature space and increasing the necessity of large training data sets to perform accurately. Furthermore, researchers must be careful to not become overly reliant to the training data, in order to produce a general use classifier that uses opinion holders, terms specific to single negotiations (such as names of companies, persons etc.) must be generalized upon – a task which introduces a significant challenge, especially due to potentially identified quality features occurring only scarcely in the data sets.

Regarding proactive communication support based on the classification methods and the specific support approaches given in chapter 6.1., future steps should include empirical validations in the form of implementing support functions in an NSS and employing these functions in ongoing negotiations. Specific topics that should be addressed are the exploration of different user interaction means, how diagnostic information can and should be transformed into advice for the negotiators, the exploration of different visualization alternatives and, lastly and most important an empirical assessment of added value of proactive communication support. Different research questions can and should be addressed over this course:

Firstly, the negotiators' perception of the support should be inquired, including whether negotiators offered support found it to be appropriate and justified, i.e. the proactive activation mechanisms should be put under inquiry. The importance of this is due to users oftentimes assuming a negative stance towards unwanted support, which in extreme cases can result in purposefully disregarding the offered advice and consciously accepting to be worse off than with support (Gettinger et al. 2010). The second question regarding negotiator's stance towards proactive communication support is to evaluate whether they perceive given instructions and advice as being useful – which could for instance be evaluated using models from the theoretical realm of Technology Acceptance (e.g. Davis 1989, Venkatesh et al. 2003). Constructs such as usefulness and ease of use could in this case also serve as indicators of practical relevance of proactive communication support components.

Secondly, the actual influence and impact of proactive communication support on negotiations should be empirically assessed with respect to process as well as outcome variables. During the negotiation process, it would be reasonable to evaluate whether advice given to negotiators by the proactive component actually introduces a change in their behaviour. Therefore, an empirical experiment comparing negotiators that receive proactive support with a control group where support is unavailable could be set up. Changes in behaviour could for example be measured by comparing negotiators communication before and after support was used (e.g. via employment of content analysis) or - to capture interaction effects with the substantive part of the negotiation – concession sizes before and after support. As variables measuring negotiation outcome, the rate of agreements versus non-agreements seem a reasonable evaluation choice, regarding that this is also the goal variable the classification model use. Other measures that should be considered are measures of utility, to investigate for example whether the communication support induced fairer behaviour (contract imbalance) or led to a better exploitation of integrative potential via improved information exchange (joint utility). Lastly, the impact of proactive communication support on socio-psychological outcomes such as satisfaction (Curhan et al. 2006) should be investigated, measuring a change in negotiators perception of the negotiation induced by the supporting component.

These suggestions on future research demonstrate that proactive support of negotiations, be it based on communication, Decision Support or integrated model is a promising and unexplored research field, which in the future may contribute towards holistic negotiation support approaches, ensuring an efficient and effective conduction of electronic negotiations. As electronic negotiations are likely to continue playing an important role in interorganisational communication it is a necessity and should be the task of e-negotiation research to improve upon existing approaches of electronic negotiation support thereby not only providing better methods but also seeking to implement methods and tools into entrepreneurial practice.

8. References

Adair, W.L. and Brett, J.M.(2005). The Negotiation Dance: Time, Culture, and Behavioral Sequences in Negotiation. Organization Science 16(1), 33–51.

Bichler, M., Kersten, G. and Strecker, S.(2003). Towards a structured design of electronic negotiations. Group Decision and Negotiation 12(4), 311–335.

Burke, K., Chidambaram, L. and Aytes, K.(2002). Do Some Things Change Faster than Others? The Dynamics of Behavioral Change in Computer-Supported Groups. Group Decision and Negotiation 11, 293–309.

Byron, K.(2008). Carrying Too Heavy a Load? The Communication and Miscommunication of Emotion by Email. Academy of Management Review 33(2), 309–327.

Carletta, J.(1996). Assessing agreement on classification tasks: the kappa statistic. Computational Linguistics 22(2), 249–254.

Clark, H.H. and Brennan, S.E.(1991). Grounding in communication, In Resnick, L.B., Levine, J.M., and Teasley, S.D. (eds.), Perspectives on socially shared cognition, pp. 127–149, American Psychological Association, Washington, DC.

Cohen, J.(1960). A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement 20, 37–46.

Cortes, C. and Vapnik, V.(1995). Support-Vector Networks. Machine Learning 20(3), 273–297.

Cox, D.R.(1958). The Regression Analysis of Binary Sequences. Journal of the Royal Statistical Society 20(2), 215–242.

Curhan, J.R., Elfenbein, H.A. and Xu, H.(2006). What Do People Value When They Negotiate? Mapping the Domain of Subjective Value in Negotiation. Journal of Personality and Social Psychology 91(3), 493–512.

Curhan, J.R. and Pentland, A.(2007). Thin Slices of Negotiation: Predicting Outcomes From Conversational Dynamics Within the First 5 Minutes. Journal of Applied Psychology 92(3), 802–811.

Daft, R.L., Lengel, R.H. and Trevino, L.K.(1987). Message Equivocality, Media Selection, and Manager Performance: Implications for Information Systems. MIS Quarterly 11(3), 355–366.

Davis, F.D.(1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. Management Information Systems Quarterly 13(3), 319–340.

De Dreu, Carsten K. W., Beersma, B., Stroebe, K. and Euwema, M.C.(2006). Motivated Information Processing, Strategic Choice, and the Quality of Negotiated Agreement. Journal of Personality and Social Psychology 90(6), 927–943.

Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K. and Harshman, R.(1990). Indexing by Latent Semantic Analysis. Journal of the American Society for Information Science 41(6), 391–407.

Donohue, W.A.(1981). Analyzing Negotiation Tactics: Development of a Negotiation Interact System. Human Communication Research 7(3), 273–287.

Druckman, D., Druckman, J.N. and Arai, T.(2004). e-Mediation: Evaluating the Impacts of an Electronic Mediator on Negotiating Behavior. Group Decision and Negotiation 13, 481–511.

Druckman, D., Filzmoser, M., Gettinger, J., Köszegi, S.T., Mitterhofer, R. and Vetschera, R.(2012). 2.0² G eNerationS - Avenues for the Next Generation of Pro-active Negotiation Support, In Almeida, A., Morais, D., and Daher, S. (eds.), Proceedings of the 2012 Conference on Group Decision and Negotiation, Recife, 20.-24.05., pp. 82–84.

Druckman, D., Harris, R. and Ramberg, B.(2002). Computer-Assisted International Negotiation: A Tool for Research and Practice. Group Decision and Negotiation 11, 231–256.

Duckek, K.(2010), Ökonomische Relevanz von Kommunikationsqualität in elektronischen Verhandlungen, 1st Edition, Gabler, Wiesbaden.

Feldman, R.(2013). Techniques and Applications for Sentiment Analysis. Communications of the ACM 56(4), 82–89.

Feldman, R. & Sanger, J.(2007), The text mining handbook: Advanced approaches in analyzing unstructured data, Cambridge University Press, Cambridge, , New York.

Forgas, J.B.(1998). On feeling Good and Getting Your Way: Mood Effects on Negotiator Cognition and Bargaining Strategies. Journal of Personality and Social Psychology 74(3), 565–577.

Forman, G.(2003). An extensive empirical study of feature selection metrics for text classification. Journal of Machine Learning Research 3, 1289–1305.

Forman, G.(2008). Feature Selection for Text Classification, In Liu, H. and Motoda, H. (eds.), Computational methods of feature selection, Chapman & Hall/CRC, Boca Raton.

Friedman, R.A. and Currall, S.C.(2003). Conflict Escalation: Dispute Exacerbating Elements of E-Mail Communication. Human Relations 56, 1325–1347.

Fuller, C.M., Biros, D.P., Burgoon, J.K. and Nunamaker, J.F.(2013). An Examination and Validation of Linguistic Constructs for Studying High-Stakes Deception. Group Decision and Negotiation 22, 117–134.

Gettinger, J., Filzmoser, M., Wachowicz, T. and Wu, S.(2010). Do They Agree Once More? - An Analysis of the Factors That Influence Agreement in the Post-Settlement Phase, In Vreede, G.-J. de (ed.), Proceedings of the 11th Group Decision & Negotiation Conference, Delft, The Netherlands, 21.-23.06., pp. 160–163.

Gettinger, J., Köszegi, S.T. and Schoop, M.(2012). Shall we dance? - The effect of information presentations on negotiation processes and outcomes. Decision Support Systems 53, 161–174.

Greenhalgh, L. and Chapman, D.I.(1998). Negotiator Relationships: Construct Measurement, and Demonstration of Their Impact on the Process and Outcomes of Negotiation 7(6), 465–489.

Gregor, S. and Hevner, A.R.(2013). Positioning and Presenting Design Science Research for Maximum Impact. Management Information Systems Quarterly 37(2), 337–355.

Griessmair, M. and Köszegi, S.T.(2009). Exploring the Cognitive-Emotional Fugue in Electronic Negotiations. Group Decision and Negotiation 18, 213–234.

Habermas, J.(1984), The theory of communicative action, Beacon Press, Boston.

Han, J. & Kamber, M.(2006), Data mining: Concepts and techniques, second edition, 2nd Edition, Elsevier; Morgan Kaufmann Publishers, Amsterdam, Boston, San Francisco, Calif.

Hand, D.J. and Yu, K.(2001). Idiot’s Bayes--Not So Stupid After All? International Statistical Review 69(3), 385–398.

Harinck, F. and Druckman, D.(2017). Do Negotiation Interventions Matter? Resolving Conflicting Interests and Values. Journal of Conflict Resolution 61(1), 29–55.

Hassan, A., Qazvinian, V. and Radev, D.(2010). What's with the Attitude? Identifying Sentences with Attitude in Online Discussions, Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, MIT Stata Center, MA, 09.-11.10.2010, pp. 1245–1255.

Herbst, U. and Schwarz, S.(2011). How Valid is Negotiation Research Based on Student Sample Groups? New Insights into a Long-Standing Controversy. Negotiation Journal 27(2), 147–168.

Jarke, M., Jelassi, M.T. and Shakun, M.F.(1987). Mediator: Towards a negotiation support system. European Journal of Operational Research 31(3), 314–334.

Jarvenpaa, S.L., Knoll, K. and Leidner, D.E.(1998). Is Anybody Out There? Antecedents of Trust in Global Virtual Teams. Journal of Management Information Systems 14(4), 29–64.

Katz, A. and Te'eni, D.(2007). The Contingent Impact of Contextualization on Computer-Mediated Collaboration. Organization Science 18(2), 261–279.

Kersten, G.E. and Lai, H.(2007). Negotiation Support and E-negotiation Systems: An Overview. Group Decision and Negotiation 16, 553–586.

Kersten, G.E. and Zhang, G.(2003). Mining Inspire Data for the Determinants of Successful Internet Negotiations. Central European Journal of Operational Research 11(3), 297–316.

Kiesler, S. and Sproull, L.(1992). Group Decision Making and Communication Technology. Organizational Behavior and Human Decision Processes 52, 96–123.

Komorita, S.S. and Parks, C.D.(1995). Interpersonal Relations: Mixed-Motive Interaction. Annual Review of Psychology 46, 183–207.

Lea, M. and Spears, R.(1992). Paralanguage and social perception in computer-mediated communication. Journal of Organizational Computing 2, 321–341.

Lewicki, R.J., Barry, B. & Saunders, D.M.(2010), Negotiation, 6th Edition, McGraw-Hill, New York, NY.

Lim, L.-H. and Benbasat, I.(1992). A theoretical perspective of negotiation support systems. Journal of Management Information Systems 9, 27–44.

Liu, B.(2012), Sentiment Analysis and Opinion Mining, Morgan & Claypool, San Rafael, CA.

Majchrzak, A., Malhotra, A. and John, R.(2005). Perceived Individual Collaboration Know-How Development Through Information Technology-Enabled Contextualization: Evidence from Distributed Teams. Information Systems Research 16(1), 9–27.

Manning, C.D., Raghavan, P. & Schütze, H.(2008), Introduction to information retrieval, Cambridge University Press, Cambridge.

Markus, M.L.(1994). Finding a Happy Medium: Explaining the Negative Effects of Electronic Communication on Social Life at Work. ACM Transactions on Information Systems 12(2), 119–149.

Mayfield, E. and Penstein-Rosé, C.(2010). Using Feature Construction to Avoid Large Feature Spaces in Text Classification, In Pelikan, M. and Branke, J. (eds.), Proceedings of the 12th annual conference on Genetic and evolutionary computation, Portland, OR, 07.-11.07., pp. 1299–1306.

McDonald, R., Hannan, K., Neylon, T., Wells, M. and Reynar, J.(2007). Structured Models for Fine-to-Coarse Sentiment Analysis, In Carroll, J.A., van den Bosch, Antal, and Zaenen, A. (eds.), Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, Prague, 23.-30.06.2007, pp. 432–439.

Mitchell, T.M.(1997), Machine Learning, McGraw-Hill, New York.

Mohammad, S.(2012). From once upon a time to happily ever after : tracking emotions in mail and books. Decision Support Systems 53(4), 730–741.

Morris, M.W. and Keltner, D.(2000). How Emotions Work: The Social Functions of Emotional Expression in Negotiations. Research in Organizational Behavior 22, 1–50.

Morris, M.W., Nadler, J., Kurtzberg, T. and Thompson, L.L.(2002). Schmooze or Lose: Social Friction and Lubrication in E-Mail Negotiations. Group Dynamics: Theory, Research, and Practice 6(1), 89–100.

Nastase, V., Koeszegi, S. and Szpakowicz, S.(2007). Content Analysis Through the Machine Learning Mill. Group Decision and Negotiation 16(4), 335–346.

Pang, B. and Lee, L.(2008). Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval 2, 1–135.

Pang, B., Lee, L. & Vaithyanathan, S. (eds.)(2002), Thumbs up?: sentiment classification using machine learning techniques, Association for Computational Linguistics, Stroudsburg, PA, USA, from http://dx.doi.org/10.3115/1118693.1118704.

Poole, M.S., Shannon, D.L. and DeSanctis, G.(1992). Communication Media and Negotiation Processes, In Putnam, L.L. and Roloff, M.E. (eds.), Sage Annual Reviews of Communication Research: Communication and Negotiation, pp. 46–66, SAGE Publications, Thousand Oaks, CA.

Postmes, T., Spears, R., Lee, A.T. and Novak, R.J.(2005). Individuality and Social Influence in Groups: Inductive and Deductive Routes to Group Identity. Journal of Personality and Social Psychology 89(5), 747–763.

Purdy, J.M., Nye, P. and Balakrishnan, P.V.(2000). The Impact of Communication Media on Negotiation Outcomes. The International Journal of Conflict Management 11(2), 162–187.

Putnam, L.L.(2005). Discourse Analysis: Mucking Around with Negotiation Data. International Negotiation 10, 17–32.

Putnam, L.L.(2010). Communication as Changing the Negotiation Game. Journal of Applied Communication Research 38(4), 325–335.

Putnam, L.L. and Jones, T.S.(1982). Reciprocity in negotiation: An analysis of bargaining interaction. Communication Monographs 49, 171–191.

Putnam, L.L. and Roloff, M.E.(1992). Communication Perspectives on Negotiation, In Putnam, L.L. and Roloff, M.E. (eds.), Sage Annual Reviews of Communication Research: Communication and Negotiation, pp. 1–17, SAGE Publications, Thousand Oaks, CA.

Quinlan, J.(1979). Discovering rules by induction from large collections of examples, In Michie, D. (ed.), Expert systems in the micro-electronic age, pp. 168–201, Edinburgh University Press, Edinburgh.

Quinlan, J.(1986). Induction of Decision Trees. Machine Learning 1(1), 81–106.

Quinlan, J.(1993), C4.5: Programs for machine learning, Morgan Kaufmann Publishers Inc, San Mateo, CA.

Raiffa, H.(1982), The Art and Science of Negotiation, Harvard University Press, Cambridge, MA.

Razavi, A.H., Inkpen, D., Uritsky, S. and Matwin, S.(2010). Offensive Language Detection Using Multi-level classification, In Farzindar, A. and Keselj, V. (eds.), Proceedings of the 23rd Canadian Conference on Artificial Intelligence, Ottawa, Canada, 31.05.-02.06., pp. 16–27.

Reyes, A. and Rosso, P.(2014). On the difficulty of automatically detecting irony: beyond a simple case of negation. Knowledge and Information Systems 40, 595–614.

Russell, S.J. & Norvig, P.(2009), Artificial intelligence, 3rd Edition, Pearson Education, Upper Saddle River, N.J., Harlow.

Schoop, M.(2001). An introduction to the language-action perspective. ACM SIGGROUP Bulletin 22(2), 3–8.

Schoop, M.(2010). Support of Complex Electronic Negotiations, In Kilgour, D.M. and Eden, C. (eds.), Advances in Group Decision and Negotiation, pp. 409–423, Springer Netherlands, Dordrecht.

Schoop, M.(2015). The Role of Communication Support for Electronic Negotiations, In Kaminski, B., Kersten, G.E., and Szapiro, T. (eds.), Proceedings of the 15th International Conference on Group Decision and Negotiation: Outlooks and Insights on Group Decision and Negotiation, Warsaw, Poland, 22.-26.06., pp. 283–287.

Schoop, M., Jertila, A. and List, T.(2003). Negoisst: a negotiation support system for electronic business-to-business negotiations in e-commerce. Data & Knowledge Engineering 47(3), 371–401.

Schoop, M., Köhne, F., Staskiewicz, D., Voeth, M. and Herbst, U.(2008). The antecedents of renegotiations in practice—an exploratory analysis. Group Decision and Negotiation 17(2), 127–139.

Schoop, M. and Quix, C.(2001). DOC.COM: a framework for effective negotiation support in electronic marketplaces. Computer Networks 37(2), 153–170.

Schoop, M., van Amelsvoort, M., Gettinger, J., Körner, M., Köszegi, S.T. and van der Wijst, P.(2014). The Interplay of Communication and Decisions in Electronic Negotiations:

Communicative Decisions or Decisive Communication? Group Decision and Negotiation 23, 167–192.

Schroth, H.A., Bain-Chekal, J. and Caldwell, D.F.(2005). Sticks and Stones may Break Bones and Words can Hurt me: Words and Phrases that Trigger Emotions in Negotiations and Their Effects. The International Journal of Conflict Management 16(2), 102–127.

Schütze, H., Hull, D.A. and Pedersen, J.O.(1995). A comparison of classifiers and document representations for the routing problem, In Fox, E.A., Ingwersen, P., and Fidel, R. (eds.), Proceedings of the 18th annual international ACM SIGIR conference on research and development in information retrieval, Seattle, WA, 09.-13.07., pp. 229–237.

Searle, J.R.(1969), Speech acts: An essay in the philosophy of language, Cambridge University Press, London.

Sebastiani, F.(2002). Machine learning in automated text categorization. ACM Computing Surveys 34(1), 1–47.

Shmueli, G. and Koppius, O.R.(2011). Predictive Analytics in Information Systems Research. MIS Quarterly 35(3), 553–572.

Short, J., Williams, E. & Christie, B.(1976), The social psychology of telecommunications, Wiley, London.

Simons, T.(1993). Speech Patterns and the Concept of Utility in Cognitive Maps: The Case of Integrative Bargaining. Academy of Management Journal 36(1), 139–156.

Sokolova, M. and Lapalme, G.(2012). How Much Do We Say? Using Informativeness of Negotiation Text Records for Early Prediction of Negotiation Outcomes. Group Decision and Negotiation 21(3), 363–379.

Sokolova, M. and Szpakowicz, S.(2007). Strategies and language trends in learning success and failure of negotiation. Group Decision and Negotiation 16(5), 469–484.

Sproull, L. and Kiesler, S.(1986). Reducing Social Context Cues: Electronic Mail in Organizational Communication. Management Science 32(11), 1492–1512.

Srnka, K.J. and Köszegi, S.T.(2007). From Words to Numbers: How to transform Qualitative Data into Meaningful Quantitative Results. Schmalenbach Business Review 59, 29–57.

Swaab, R., Postmes, T. and Neijens, P.(2004). Negotiation Support Systems: Communication and Information as Antecedents of Negotiation Settlement. International Negotiation 9, 59–78.

Taylor, J.R., Cooren, F., Giroux, N. and Robichaud, D.(1996). The Communicational Basis of Organization: Between the Conversation and the Text. Communication Theory 6(1), 1–39.

Te'eni, D.(2001). Review: A Cognitive-Affective Model of Organizational Communication for Designing IT. Management Information Systems Quarterly 25(2), 251–312.

Te'eni, D.(2006). The language-action perspective as a basis for communication support systems. Communications of the ACM 49(5), 65–70.

Thompson, L.L. and Nadler, J.(2002). Negotiating via Information Technology: Theory and Application. Journal of Social Issues 58(1), 109–124.

Tutzauer, F.(1992). The Communication of Offers in Dyadic Bargaining, In Putnam, L.L. and Roloff, M.E. (eds.), Sage Annual Reviews of Communication Research: Communication and Negotiation, pp. 67–82, SAGE Publications, Thousand Oaks, CA.

Twitchell, D.P., Jensen, M.L., Derrick, D.C., Burgoon, J.K. and Nunamaker, J.F.(2013). Negotiation Outcome Classification Using Language Features. Group Decision and Negotiation 22(1), 135–151.

Tzafrir, S.S., Sanchez, R.J. and Tirosh-Unger, K.(2012). Social Motives and Trust: Implications for Joint Gains in Negotiations. Group Decision and Negotiation 21, 839–862.

van Kleef, Gerben A.(2009). How Emotions Regulate Social Life: The Emotions as Social Information (EASI) Model. Current Directions in Psychological Science 18, 184–188.

Venkatesh, V., Morris, M.G., Davis, G.B. and Davis, F.D.(2003). User Acceptance of Information Technology: Toward a Unified View. Management Information Systems Quarterly 27(3), 425–478.

Vetschera, R.(2016). Concessions Dynamics in Electronic Negotiations: A Cross-Lagged Regression Analysis. Group Decision and Negotiation 25(2), 245–265.

Walther, J.B.(1992). Interpersonal effects in computer-mediated interaction: A relational perspective. Communication Research 19(1), 52–90.

Walther, J.B.(1996). Computer-Mediated Communication: Impersonal, Interpersonal and Hyperpersonal Interaction. Communication Research 23(1), 3–43.

Walther, J.B.(2011). Theories of Computer-Mediated Communication and Interpersonal Relations, In Knapp, M.L. and Daly, J.A. (eds.), The SAGE Handbook of Interpersonal Communication, pp. 443–479, SAGE Publications, Thousand Oaks, CA.

Walther, J.B., Loh, T. and Granka, L.(2005). Let Me Count The Ways - The Interchange of Verbal and Nonverbal Cues in Computer-Mediated and Face-to-Face Affinity. Journal of Language and Social Psychology 24(1), 36–65.

Walther, J.B. and Parks, M.R.(2002). Cues Filtered Out, Cues Filtered In - Computer-Mediated Communication and Relationships, In Knapp, M.L. and Daly, J.A. (eds.), Handbook of interpersonal communication,3rd edn., pp. 529–563, SAGE Publications, Thousand Oaks, CA.

Weigand, H., Schoop, M., Moor, A. de and Dignum, F.(2003). B2B Negotiation Support: The Need for a Communication Perspective. Group Decision and Negotiation 12(1), 3–29.

Winograd, T.(1988). A Language/Action Perspective on the Design of Cooperative Work. ACM SIGCHI Bulletin 20(1), 79.

Witten, I.H. & Frank, E.(2005), Data Mining: Practical Machine Learning Tools and Techniques, Second Edition, 2nd Edition, Morgan Kaufmann, Waltham.

Yang, Y. and Pedersen, J.O.(1997). A Comparative Study on Feature Selection in Text Categorization, Proceedings of the Fourteenth International Conference on Machine Learning, pp. 412–420.

Yuan, Y., Rose, J.B. and Archer, N.(1998). A Web-Based Negotiation Support System. Electronic Markets 8(3), 13–17.

Zaidman, N., Te'eni, D. and Schwartz, D.G.(2008). Discourse-based technology support for intercultural communication in multinationals. Journal of Communication Management 12(3), 263–272.

Zhang, T. and Oles, F.J.(2001). Text Categorization Based on Regularized Linear Classification Methods. Information Retrieval 4, 5.31.

Zhou, L., Burgoon, J.K., Nunamaker Jr., J.F. and Twitchell, D.P.(2004a). Automating Linguistics-Based Cues for Detecting Deception in Text-based Asynchronous Computer-Mediated Communication. Group Decision and Negotiation 13, 81–106.

Zhou, L., Burgoon, J.K., Twitchell, D.P., Qin, T. and Nunamaker Jr., J.F.(2004b). A Comparison of Classification Methods for Predicting Deception in Computer-Mediated Communication. Journal of Management Information Systems 20(4), 139–165.