Kaya, Muhammed-Fatih

Article — Published Version

Pattern Labelling of Business Communication Data

Group Decision and Negotiation

Provided in Cooperation with:
Springer Nature

Suggested Citation: Kaya, Muhammed-Fatih (2022) : Pattern Labelling of Business Communication Data, Group Decision and Negotiation, ISSN 1572-9907, Springer Netherlands, Dordrecht, Vol. 31, Iss. 6, pp. 1203-1234, https://doi.org/10.1007/s10726-022-09800-2

This Version is available at:
https://hdl.handle.net/10419/306656

Group Decision and Negotiation (2022) 31:1203–1234
https://doi.org/10.1007/s10726-022-09800-2

Pattern Labelling of Business Communication Data

Muhammed-Fatih Kaya[1]

Accepted: 6 September 2022 / Published online: 28 October 2022

Abstract
Systematic pattern recognition as well as the corresponding description of determined patterns entail numerous challenges in the application context of high-dimensional communication data. These can cause increased effort, especially with regard to machine-based processing concerning the determination of regularities in underlying datasets. Due to the increased expansion of dimensions in multidimensional data spaces, determined patterns are no longer interpretable by humans. Taking these challenges into account, this paper investigates to what extent predefined communication patterns can be interpreted for the application area of high-dimensional business communication data. An analytical perspective is considered by taking into account a holistic research approach and by subsequently applying selected Machine Learning methods from Association Rule Discovery, Topic Modelling and Decision Trees with regard to the overall goal of semi-automated pattern labelling. The results show that meaningful descriptions can be derived for the interpretation of pre-defined patterns.

Keywords Business Communication Data · Machine Learning · Pattern Recognition · Pattern Labelling · Association Rule Discovery · Topic Modelling · Decision Tree

1 Introduction

Pattern recognition has its roots in engineering and represents an important sub-discipline of Machine Learning (ML). It involves the automated determination of systematic rules and patterns in large datasets through the iterative application of specific algorithms (Bishop and Nasrabadi 2007). Powerful database systems exist in today's

✉ Muhammed-Fatih Kaya
Muhammed-Fatih.Kaya@uni-hohenheim.de

[1] University of Hohenheim, 70599 Stuttgart, Germany

world that are able to store and manage large data streams through the systematic use of analytical methods. They enable value-added information to be derived from raw data that is not cognitively detectable by humans (Basit 2003). As a result, the application of pattern recognition techniques and the associated interpretive attempts are nowadays used in a wide variety of application contexts (Biggio and Roli 2018; Garlanda et al. 2018). ML methods have also been actively investigated in negotiation interactions with the aim of supporting negotiation participants in designing effective negotiation processes in the research field of electronic negotiations (Carbonneau et al. 2008; Kersten and Zhang 2003; Körner and Schoop 2017; Sokolova and Szpakowicz 2005; Zeng and Sycara 1998). In view of the increasing technological progress and the spread of more powerful information and communication technologies, the importance of electronic negotiations is growing continuously and represents a topic of high strategic relevance for many companies from a business management and information technology perspective (Agndal 2007). Decision-makers are increasingly tempted to conduct negotiations electronically in an asynchronous and dislocated manner due to the increasing globalisation in the age of digitalisation (Kersten and Noronha 1999). Significant time and cost advantages are gained through the use of electronic channels (Foroughi 1998). In addition to these well-known advantages, a broader view to current world events in relation to the extraordinary pandemic situation underlines the necessity of using electronic communication platforms in a wide variety of areas (Byrnes et al. 2021; Klich 2021). In order to maintain economic relations in the globalised world, international contracting parties are more dependent than ever on using electronic means of communication to be able to protect their own health and the surroundings from contagion.

Communication, as a central element for reaching agreement in negotiations, is especially used in conflicts of interest (Olekalns et al. 2008). A study by Putnam and Jones (1982) shows that successful negotiations differ from failed ones in their communication behaviour so that the analysis of business communication is of particular importance. Communication behaviour can reveal important information about how and what parties negotiate during the negotiation process. It should serve the goal of enabling the understanding of negotiation interactions as a meta-goal (Putnam and Roloff 1992). The analysis of negotiation communication data to derive behavioural patterns has mainly been conducted by using manual coding (Srnka and Koeszegi 2007; Weingart et al. 2004). The goal of this approach is to divide the communication data into pre-defined negotiation units (e.g., negotiation sentences, thought units, etc.). These units are subsequently coded and assigned to pre-defined categories by independent human coders to derive communicational patterns (Donohue et al. 1984). Particularly in light of the rapid growth of data, the processing of manual coding became much more difficult to conduct (Crowston et al. 2012). Especially the processing of large datasets by manual coding leads to an increased effort for the coders and requires a significant amount of work which could be used efficiently elsewhere.

The application of ML methods to detect patterns in communication data is particularly suitable for larger quantities of data. Nevertheless, derived results must presuppose interpretability which can be understood by humans (Karanika et al. 2020). While the human brain is trained to cognitively process natural language, processing

unstructured and natural language from high-dimensional data space as is the case with business communication data may pose certain challenges for machine processing and interpretability (Vázquez et al. 2020). The investigation of communication data in the context of electronic negotiations can deliver important indicators for the modelling of negotiation behaviour. But it is more complex to analyse unstructured communication data in comparison to the structured utility data of negotiation interactions. This is the reason why the communication part of negotiations is going to be focused within the scope of this research paper.

This research paper uses a set of clustered negotiation sentences as input data whose structural patterns were determined in previous studies by considering the underlying data characteristics (Kaya and Schoop 2022). Based on these given set of structured negotiation sentences, this study aims to investigate to which extent descriptive indicators can be derived from each cluster. To overcome the previously mentioned challenges of machine-based processability of large amounts of communication data and to be able to derive interpretable pattern-based descriptions, selected labelling techniques from ML will be investigated. Our research is based on an explorative analytical perspective. It applies and evaluates different ML methods, i.e. two unsupervised methods (namely Topic Modelling and Association Rule Discovery) and a supervised method (namely Decision Tree), to electronic negotiation communication.

Individual strengths of different ML-methods could counteract the data-driven challenges of machine-processability. Hence, a combined application could lead to a maximisation of semantic indicators. This paper does not only aim to present a methodological potential analysis of how well the individual methods perform on pre-structured communication data by considering their individual strengths and weaknesses. Rather, the complementary nature of methods is going to be examined to investigate whether and to what extent supplementary or rather validating indicators can be derived through the diverse use of ML. This would generate considerable added value with respect to the combined application of ML and would also increase the quality of the results. This leads us to the following research question:

Which value-adding pattern descriptions can be derived taking into account a holistic ML-based approach based on high-dimensional business communication data?

After a detailed reflection on the field of electronic negotiation support systems and the presentation of previous studies on the description of negotiation behaviour, a methodological reflection on the possibilities regarding the interpretation of cluster patterns follows. Taking into account the overarching research question and the challenges associated with it, a research procedure is presented in the following chapter with which selected patterns should be examined according to their meaning. On this basis, the results of applied procedures are presented and discussed taking into account a holistic approach. Finally, central findings are summarised and a research outlook is presented.

2 Theoretical Background

2.1 Support of E-Negotiations

Negotiation is needed in situations where differences between at least two parties occur and no clearly defined rules for resolving the conflicts exist (Lewicki et al. 2010). According to Bichler et al. (2003), negotiation is an iterative communication and decision-making process in which the negotiating parties are not able to achieve their respective negotiation goals through unilateral actions. Negotiators pursue the goal of achieving their pre-defined negotiation goals. To this end, the parties exchange information, arguments, offers and counter-offers. To enable negotiations to be conducted as effectively as possible and to accompany negotiating participants in the active process, classic face-to-face negotiations can be handled more efficiently through the use of electronic negotiation systems (Schoop 2021a). Negotiation Support Systems (NSSs) have been developed to support negotiations in terms of structured processing and thus achieve better negotiation results as one support variant of electronic negotiations (Schoop et al. 2003). The NSS called Negoisst, whose business communication data is going to be examined within this work, has been used in university teaching for over 17 years. As a holistic NSS, it combines the support functionalities of communication support with decision support (Schoop 2010, 2021a). Process-oriented assistance is to be ensured by helping the negotiation participants to design the communication process with the help of communication support in an optimal way (Schoop et al. 2003; Weigand et al. 2003; Kersten and Lai 2007). Here, communication takes place in an asynchronous manner by textually exchanging semi-structured negotiation messages in natural language (Schoop 2005). After an exchange of offers and counter-offers, which are documented in negotiation messages, the negotiation process ends with either an agreement or a rejection (Pruitt 2013).

2.2 Negotiation Behaviour

Negotiation researchers have been investigating the existence of certain behavioural patterns in negotiations to optimise communication exchanges for many years (Pruitt 2013; Weingart et al. 2004). Researchers investigate how partners negotiate with each other in order to fulfil the negotiation goals and which special behavioural patterns play an elemental role in negotiations. Negotiation behaviour is measured by expressions in communication (Pesendorfer and Koeszegi 2006; Weingart et al. 1990). Especially in the context of electronic negotiations, communication plays an elemental role and acts as a transmitter for conveying elemental indicators regarding the individual negotiation behaviour to the negotiation partner. Communication in form of messages, which are exchanged from negotiators, represents the only available source of behaviour disclosure due to the absence of nonverbal communication in e-negotiations (Hargie and Dickson 2004). This highlights the elemental interdependency between the negotiation behaviour and communication interactions in the field of electronic negotiations as well as the need for systematic analysis of negotiation communication data. Nevertheless, it should also be noted that even if the com-

munication part of negotiations is very important for the reproduction of negotiation behaviours, there is also a contractual part of negotiations which will not be focussed in this research paper as mentioned in the introduction chapter, where e.g. individual preferences, aspiration and reservation levels are changed on a formal base (Rangaswamy and Shell 1997; Vetschera 2007).

In the negotiation process, negotiators always pursue the goal of asserting as many of their preferences as possible within the framework of a contract agreement through the use of negotiation strategies (Thompson et al. 2010). This goal can be achieved in various ways so that studies deal with the process-oriented implementation of negotiations and analyse the entire negotiation process specifically phase by phase (Adair and Brett 2005; Preuss and van der Wijst 2017; Weingart and Olekalns 2004). In line with previous studies, it can be summarised that negotiation researchers consider the disclosure of information regarding the subject matter of the negotiation to be an important contribution to maintaining the negotiation process and achieving the goals. It can symbolise trust towards the negotiating partner and is frequently observed in integrative negotiation processes, especially at the beginning of negotiations (Adair and Brett 2005; Harinck and Ellemers 2006). In addition, the exchange of information on the negotiating issues can lead to mutual understanding and can open new potentials for working on alternative solutions (Liss 2011). This is often reflected in an increased joint negotiation outcome (Kaya et al. 2017; Rangaswamy and Shell 1997). The elementary role of the use of concessions should not be neglected in this sense. Concessions represent an elementary building block of integrative negotiation strategies. Especially in the context of electronic negotiations, the systematic use of concessions can increase the probability of acceptance and enable the achievement of a pareto-optimal outcome (Filzmoser and Vetschera 2008; Vetschera 2016). This, as well as the transparent sharing of information, also has a positive influence on the relationship between the contracting parties. Relationship building has a positive impact on building mutual trust and can lead to an integrative negotiation behaviour and thus to an increased potential for success in negotiations (Lewicki and Polin 2013). This kind of negotiation behaviours and others convey an important contribution to the successful conduction of negotiations. Nevertheless, in addition to this content-related focus of negotiation theory, a methodological focus is also necessary with regard to the question of how these and additional behavioural patterns can be derived and interpreted from exchanged communication messages.

While the detection of patterns in the underlying data represents a first methodological challenge in unstructured negotiation messages, it is of great importance to conduct research regarding the way whether and how structural patterns could be understood by humans in the high-dimensional data space. Even if the detection of systematic regularities on syntactical base delivers a first important contribution, there is need for knowledge discovery and information retrieval on semantical base. Otherwise, derived patterns could not be integrated into value-adding series of decisions from decision makers since patterns that are not interpretable and thus not usable for humans would be generated.

2.3 Ways of Pattern Labelling in High-Dimensional Business Communication Data

The use of ML methods is indispensable for the fast and targeted processing and analysis of large and unstructured datasets as they are able to find systematic patterns in the datasets that are not detectable to the human eye (Kaur and Agrawal 2018; Tandel et al. 2019). However, the associated preliminary step of processing unstructured business communication data should by no means be neglected as the underlying natural language is often subject to a so-called curse of dimensionality as a result of the processing. This exponential increase can significantly complicate the performance of analyses (Debie and Shafi 2019). This challenge of dimensionality expansion can be counteracted through a series of milestones to reduce the dimensions in dataset. An approximated real-world representation has to be generated with a reduced dimensional density where structural patterns have to be determined to be able to analyse the data structured in a detailed manner (Khalid et al. 2014). These milestones of deriving a database of negotiation sentences as well as the detection of structural patterns by systematically grouping negotiation sentences into pre-evaluated clusters have been conducted in previous studies (Kaya and Schoop 2022).

Based on this fundament of clustered negotiation sentences, this paper targets the question of to what extent machine-based indicators can be derived in these clusters and whether those can be interpreted by humans. Interpretability is defined as the possibility of patterns to be understood by humans. The goal of achieving accurate (high number of correct pattern predictions) as well as valid patterns (patterns that are consistently confirmed by more than one pattern prediction method) is often equated with a high degree of interpretability. A look at various studies shows that this equation cannot be fulfilled in all circumstances. While highly accurate patterns bring great challenge in interpretability, it may be that results reveal better interpretability potential with a solid medium accuracy. Consequently, the user has to make a trade-off between accuracy and interpretability in some situations (Chang et al. 2009). Especially when processing data from high-dimensional data space, users are often confronted with the question of whether characteristic descriptions can be derived and whether a respective label can be assigned to the pattern group.

In principle, different groups of ML methods can be used for the interpretation of pre-defined patterns. It should be noted that the methods to be applied should have a descriptive property in order to be able to reflect the characteristic of patterns to be examined in depth.

Assume that the following sentence:

"*Dear Partner, I think that we need further concessions to reach an acceptable negotiation state.*"

was assigned to a structural cluster (e.g., cluster 1) as one of many intensively preprocessed negotiation sentences in previous studies. Decision Tree, Association Rule Discovery and Topic Modelling that contribute to the ML-based description of structural patterns are presented in the following; they offer different potentials to derive semantic indicators. An in-depth insight into the process-oriented execution of the ML processes and into the algorithms can be taken from the appendix (see appendix).

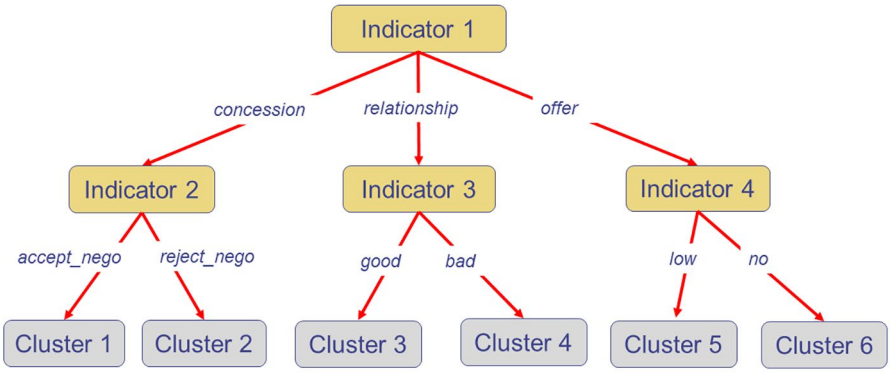

Fig. 1 Decision Tree Example

Decision Trees (DTs) can be used as a first descriptive approach. They belong to the so-called white-box algorithms and to the type of methods that are among the best to interpret compared to the other classification approaches (Molnar 2020). DTs can operate on large datasets and are relatively fast compared to other classification techniques (Cherfi et al. 2018). The overall goal of DTs is to automatically find a tree representation (consisting of a root node, decision nodes, branches and leaf nodes as target classes) by deriving systematic rules. It pursues the goal of performing classifications on pre-defined target classes with as few decision paths as possible (Myles et al. 2004; Priyanka and Kumar 2020) (see Fig. 1). Derived decision paths reveal information about the internal decision-logic and provide insights into the question of how the splitting criteria of data objects look like in detail (Guidotti et al. 2018). These derived decision paths (marked in red in Fig. 1) could be re-used as main DT-results to describe the characteristics of respective target classes. In our application case, the information about already determined allocations to cluster partitions (derived from previous studies) could be assumed as target classes (e.g. target class "cluster 1" in our example). These target classes could be predicted as accurately as possible by deriving decision rules. These rules can be interpreted in the following to get an understanding of which underlying decision paths are needed to describe the respective target class. Generated decision-logic can provide important semantic indicators regarding the characteristic properties of patterns (Amatriain et al. 2011). A decision path for our example cluster 1 (the decision path on the far left) indicates that the word element "concession" in combination with the next indicator "accept_nego" (indicator 2) describes the first cluster best. This kind of information could be re-used as sentence-based indicators.

While some scientists praise the good handling of high-dimensional data as well as the fast computational power of DTs, others are sceptical about the determination of ideal decision paths (Japkowicz and Stephen 2002). This is the reason why three different variants of DTs, namely a Gradient Boosting DT (GBDT), a Random Forest DT (RF) and a classical DT in form of a C4.5 approach are evaluated in this paper. While the GBDT optimises the DT by generating a series of interconnected DTs where the error rates from previous DTs in a sequential way (Ayyadevara 2018),

the RF starts generating individual DTs that are combined to a superordinate DT (Cutler et al. 2012). The renown C4.5 algorithm by Quinlan (1993) generate DTs using optimising internal quality criteria to reach the best possible data split in the tree representation (Quinlan 2014). All detailed functionalities of the corresponding ML-approaches are described in the appendix.

$$\boldsymbol{Cluster}\ 1 : concession \rightarrow accept\_nego$$

Association Rule Discovery (ARD) is one of the unsupervised ML methods which can be also used to analyse clustered negotiation sentences. ARD examines the strength of co-occurrences and their correlations in a set of attribute values in large datasets and generates if (premise) and then (consequence) relationships as so-called association rules (*if* → *then*) (Babi et al. 2017; Chen et al. 2015). An in-depth look at the implication of association rules shows that the consequence part can only occur if the premise part occurs. Those co-existences can be considered for describing the clusters in the form of association rules. The rules have to fulfil certain quality measures of the Apriori algorithm which are described in the appendix (Zheng et al. 2001). The goal of this descriptive technique is thus not to describe the derived internal decision-logic regarding the occurrence of certain events as in the case of the DT through a predictive approach. It aims to determine the joint existence of data entries in the form of if-then relationship patterns (Agrawal and Srikant 1994). Our example shows that the following association rule might occur if there is a significant existence of similar sentence constellations (including the words "concession" and "accept_nego") in cluster 1: *concession* → *accept_nego*. This rule shows that there exists a clear if-then relationship between these two word-constructs which indicates that a negotiation acceptance is only addressed in the negotiation sentences if a reference to concessions are made. ARDs have the important advantage that they can operate very quickly on the underlying datasets because of their exclusively descriptive nature. In addition, ARDs are known for the fact that their results are easier to understand for the interpreter due to the presented association relationship (Ahmed et al. 2014). However, this is countered by the aspect of the danger of overloading. Especially when quality measures are set too low, it can occur that a large number of rules are returned which are no longer interpretable for humans (Zheng et al. 2001). Consequently, a suitable level should be chosen for the considered quality measures.

Another unsupervised way to describe predefined cluster patterns is ***Topic Modelling*** (TM). The application of TM is particularly suitable when applied to textual data and offers the possibility to explore text collections in a thematical way (Kim et al. 2012). As an explorative and probabilistic model, TM aims to group a collection of words from textual data that occur together into topic-based subgroups (Alghamdi and Alfaqhi 2015). A document collection (in our case: negotiation sentences assigned to a cluster) may consist of multiple topics where each topic consists of words of a probability distribution. Hence, topics in a text corpus are declared as calculated probability distributions of words which define weights considering the relative frequency of underlying words and their interrelations in a specific corpus (Gan and Qi 2021). Those words are combined into common topics which show a statistically significant interrelation (Vayansky and Kumar 2020) (see Fig. 2). TM

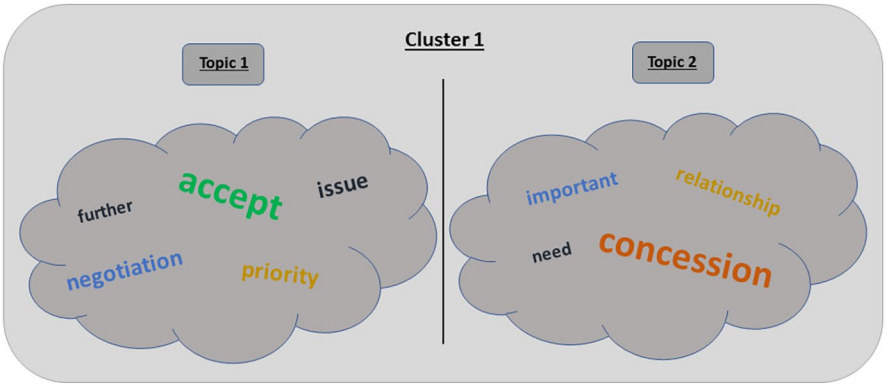

Fig. 2 Topic Model Example

calculates the topics of the text collection, the topic proportions in the individual texts and which words belong to the respective topics (Blei et al. 2003). Details of the applied LDA-based ML-approach are described in the appendix (see appendix).

The top n terms contributing to the derivation of a topic are presented in a visual representation as a word cloud (see Fig. 2). The number of topics is automatically determined by conducting an iterative, data-driven evaluation using a so-called RPC measure which proves how well a derived language model fits to a given word distribution in a text corpus (Boyd-Graber et al. 2014). While words with higher contributing probabilities regarding a specific topic get a higher weight and are therefore displayed larger in the word cloud, words with smaller probability distribution are presented smaller (Li et al. 2018). By taking all negotiation sentences and our sample sentence assigned to cluster 1 into account, two topics could be generated in our example (see Fig. 2). While words such as “accept”, “negotiation” and others occur in a statistically conspicuous way and are therefore summarised in one topic (topic 1), words like “concession”, “relationship”, etc. are summarised in the second topic. A deep look into the two topics shows that the words “accept” and “concession” are emphasised in their topics (see Fig. 2). The results of topic analysis are easy for humans to grasp as long as the visualisation of the top-n words remain within an appropriate scope. Furthermore, TM is easily applicable to different levels of aggregation of textual data and can detect subgroups in the data without knowing the target classes (Correia and Gonçalves 2017). Contrary to this, it should be noted that TM is not able to capture word relations between co-occurring words as it is the case with DT and ARD. In summary, both supervised and unsupervised methods offer the possibility of examining cluster patterns in different ways according to their characteristic properties so that a holistic investigation should be considered in further research milestones.

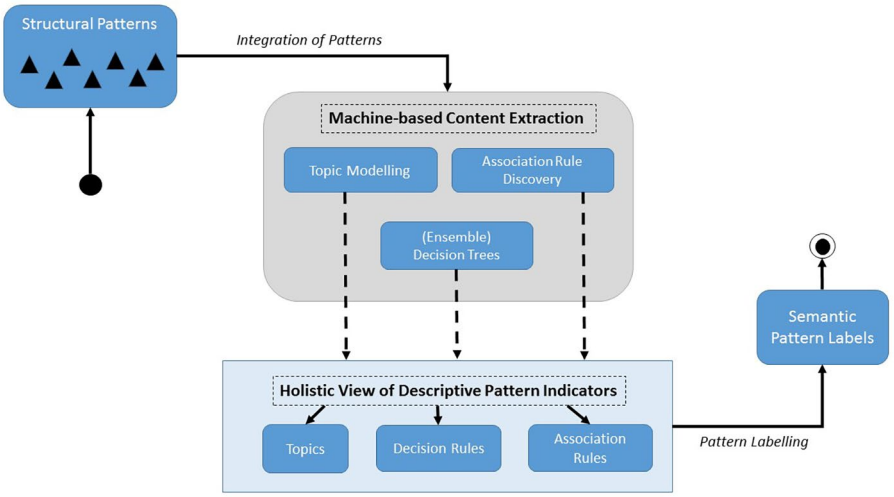

Fig. 3 Overall Research Procedure

Table 1 Extracted Association Rules of OS-dataset with k-Means (k = 5)

Cluster-Partition	Rule	sup	conf	lift
cluster_1	*hear, look, for* → *look_for, for_hear, look_for_hear*	0.023	0.999	41.807
cluster_2	*accept_offer* → *offer, accept*	0.016	1.000	47.278
cluster_3	*term_contract* → *contract, term*	0.024	1.000	37.335
cluster_4	*joint_ventur* → *ventur*	0.020	1.000	45.586
cluster_5	*wil_mak* → *mak_wil*	0.020	1.000	32.142
	hop_agre → *agre, hop*	0.030	1.000	23.003

3 Research Procedure

Considering the described challenges of individual ML-methods from the previous chapter, a research procedure presenting the research approach and overall research goal is needed on a meta-level to understand how pattern labelling could be conducted in a systematic way (see Fig. 3). It has to be noted that the implementation of the holistic approach regarding the labelling of detected patterns also requires some preliminary milestones. In addition to the overall research procedure presented in Fig. 3, a detailed ML-based process pipeline is presented in the appendix to show the processing effort of negotiation communication data and those milestones which have been necessary as a precondition to derive structural cluster patterns (Kaya and Schoop 2020, 2022) (see Fig. 8 in the appendix).

The determination of previously detected structured patterns (in our case: clustered negotiation sentences) represents the starting point for our analysis and the data

basis for answering the research question as mentioned before. For this purpose, all these patterns have to be sequentially integrated into a process of machine-based content extraction in a first step for in-depth investigation by taking the methodological potentials of presented ML-methods into account (see Fig. 3). All in all, one ARD approach investigating the co-occurrences of words in the negotiation sentences, one probabilistic TM approach detecting topic-based interrelations and three complementary DT approaches to derive decision paths describing the structural cluster patterns are applied. While the two descriptive approaches of ARD and TM specifically use the assigned negotiation sentences to cluster patterns as a database, the approaches from type DT follows the goal of conducting its analysis on a predictive level by considering all negotiation sentences and associated target classes (cluster labels). Here, all data must be collectively taken into account contrary to the previous methods.

The considered methods for deriving pattern indicators should not be applied independently from each other to profit from the full range of methodological advantages. Therefore, a holistic approach is recommended which allows to benefit not only from the methodological richness but also from the systematic combination of content-based pattern indicators. Furthermore, the risk of individual approaches not being able to derive pattern descriptions could be minimised by using several complementary approaches. As explained in the previous chapter, each of these ML-methods generate different kinds of descriptive pattern descriptions in the form of topics, decision rule paths and association rules (see Fig. 3). These indicators could then be used in a next step to interpret and to label those structural patterns from clustered sentences according to their content-related characteristics. Semantic labels could be assigned to structural patterns in the final milestone of the research process. This holistic view of combining these differing pattern indicators could provide important insights into the highly relevant question of which semantic meaning structural patterns include. Based on these insights, a pattern labelling can be conducted for each of the structured cluster patterns in a last step.

4 Results

Our research uses business communication data from international negotiations conducted via the NSS Negoisst. This data was collected in ten student experiments between the years 2010 to 2016 in the context of curricular courses. In total, 7026 negotiation messages were exchanged in the context of B2B negotiations which were tokenised into 72,826 negotiation sentences and represent the main database of this work. Some preliminary milestones were needed to be able to derive structural cluster patterns of negotiation sentences as presented in the ML-based process pipeline (see appendix). In order to gain a better understanding of the results from the preparatory steps carried out in previous studies, the key results are shortly summarised in the in the next paragraph.

The negotiation sentences were transformed into machine processable TDMs by evaluating different stemming approaches in the form of Porter and Lovins in the first step (Lovins 1968; Porter 1980; Willett 2006). The following matrices were generated for Porter: [72,826×9661] and Lovins: [72,826×8880] (Kaya and Schoop

2020). Subsequently, several dimensionality reduction approaches were evaluated on both datasets by applying Optimise Selection (OS) and the Principal Component Analysis (PCA) dimensionality reduction techniques (Song et al. 2010; Venkatesh and Anuradha 2019). Hence, one OS-optimised matrix of [72,826×4841] for Porter and another PCA-optimised matrix of [72,826×175] for Lovins were derived as data basis for pattern recognition of previous studies. In the next step, partitioning and density-based clustering methods were directly applied on each of the both OS-optimised as well as PCA-optimised matrices to detect valid cluster separations. This pairwise application and corresponding evaluation of clustering quality have led to the deduction of three partitioned PCA-based datasets and one OS-based dataset due to the reason differing cluster separations were proposed during the evaluation of especially the k-Means clustering method (Kaya and Schoop 2022). In this connection, two partitions with five and seven cluster separations were determined for the PCA-dataset and another five partitions for the OS-dataset by using k-means clustering. Furthermore, another three partitions were determined in the data for the PCA-dataset by using the density-based DBSCAN technique. Hence, 20 cluster partitions were derived in total, as part of a previous study, based on the pair-wise application of clustering methods on each of the both OS- and PCA-datasets (Kaya and Schoop 2022). Nevertheless, these 20 cluster separations should not be understood as totally disjoint separations because differing clustering methods were respectively applied to each of the two datasets to ensure valid pattern separations. Rather, these separations are to be understood as patterns proposed by considering different data preparation approaches, reduction steps and pattern recognition methods by evaluating differing clustering techniques. These steps were necessary to obtain structured negotiation sentences in a clustered form.

The ARD is applied for the derivation of cluster description in the first step of this study (see Fig. 3). Here, the Apriori algorithm of the ARD is applied step by step to each of the 20 cluster partitions by calculating a maximum of 3-grams as n-gram. An n-gram represents a sequence of n consecutive units (here: words) in a text (here: negotiation sentences) and calculates a set of co-occurring words in a textual dataset (Schonlau et al. 2017). Potential co-occurrences for 3-grams used in the ARD process are represented as *word1*, *word1_word2* or *word1_word_2_word3* constellations in premises and consequences of association rules. In terms of quality criteria, a sup_{min} of 0.05 and a $conf_{min}$ of 0.95 are defined for high-dimensional business communication data after several optimisation loops. Hence, only association rules with high validity are presented. In addition, an extract of the rules with the highest positive effect is presented for each cluster partition due to the fact of the larger the lift value the more significant the association rule.

In parallel with the application of the ARD, numerous optimisation iterations of the LDA-based TM are carried out for the pre-determined clusters. This internal evaluation ensures the determination of the optimal number of topics for each of the 20 cluster patterns (see Fig. 3). The RPC presented in the appendix chapter is used to determine the optimum, since an asymptotic course could be observed for almost all cluster partitions with the perplexity measure. The RPC results of the iterative evaluation show that divisions into two subtopics are recommended for almost each of the 20 cluster partitions, apart from the first cluster of the DBSCAN approach.

Four topics could be determined for the first cluster of DBSCAN according to the RPC evaluation. After determining the respective optimal topic number of k = 2 for the majority of partitions, the LDA can be applied using this information to perform the LDA-based TM. The top 10 words of the topic are visualised for each cluster partition in the form of a word cloud to ensure the clarity of existing topics based on an explorative pre-evaluation of optimal word numbers. The pre-evaluation of word numbers shows that there is no significant difference between the generation of word clouds with a word constellation of n > 10 words, so that only low-weighted and thus less significant words would be derived in these cases. One further aspect is the deteriorating visualisation with an increasing number of TM words. As described in the scope of the research procedure, terminologies with greater relevance in the respective topic are strongly weighted and thus mapped larger in the word cloud visualisation. In the following, the results of the ARD and TM are presented step by step and in a complementary parallel manner for each cluster separation starting with OS-dataset based on k-Means which detected five clustering partitions, followed by the PCA-based k-Means-dataset with five partitions, the PCA-based k-Means-dataset with seven partitions and finally the PCA-based DBSCAN approach which detected three cluster partitions.

Starting with the ARD results for the OS-dataset with k-Means (k = 5), association-based indicators could be derived for each of the five clusters. Focussing the first cluster partition, future-oriented indicators referring to the sentence "I look forward to hearing from you" could be detected. While the items of the second cluster refer to the acceptance of offers, the third cluster determines the pattern {*contract*} and clearly refers to contract-based objects. These contract-specific items are also supported by the rules of the fourth cluster with detecting patterns concerning the contract-specific information of joint ventures (see Table 1). The last cluster partition contains the most significant rule with the indicators for {*make*} and {*will*} as the desire to act. This desire is linked to the hope of an agreement as the second significant rule of the partition.

A look at the following TM results for the same database of OS-dataset with k-Means (k = 5) shows that topic patterns could be derived for each of the initially determined five cluster partitions as in the case of the ARD (see Table 1 and see Fig. 4).

For the first cluster of the OS dataset, terms with positive connotations of thanking and looking to the future such as {*thank, good, look, forward, hear*} are recognisable (topic 1) in addition to negotiation-specific words {*negotiation, offer*}, while the second topic contains clear contract-specific elements (see Fig. 4). The first topic of the second cluster is about {*offer*}, while the second topic emphasises the concept of reflection {*think*}. Furthermore, the first topic of the third cluster lists issues referring to the contract-specific content, whereas no clear picture can be discerned for the second topic. Moreover, the first topic of the fourth cluster focuses on the core concept of action with the central term {*want*}. Negotiation items are again identified for the second topic in the same cluster. In the fifth and last cluster, the terms {*make, offer, concession*} and the term for acceptance {*agree*} are highlighted based on the first topic.

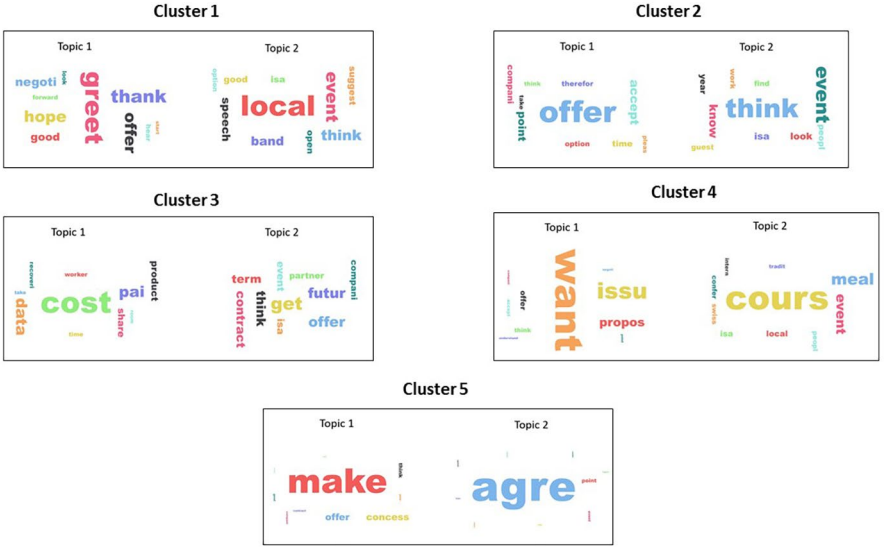

Fig. 4 Topics of OS-dataset with k-Means (k = 5)

Table 2 Extracted Association Rules of PCA-dataset with k-Means (k = 5)

Cluster-Partition	Rule	sup	conf	lift
cluster_1	*joint, ventur* → *joint_ventur*	0.016	0.998	61.804
cluster_2	*thank_quick* → *quick*	0.020	1.000	39.085
	fin_offer → *fin*	0.022	1.000	23.778
cluster_4	*therefor_think* → *therefor*	0.027	1.000	25.024
cluster_5	*look_for, hear* → *for_hear, look_for_hear*	0.030	0.970	32.640

The application on the PCA-dataset with k-Means (k = 5) again generates rules with contract-specific implications for the first cluster this time, while the second cluster includes a thanking indicator which is supported by the rule of the final offer in the further step. While no rules could be derived for the third cluster, the rule of the fourth cluster focuses on the rule "therefore think". The last cluster again points to the sentence "I am looking for(ward)" as detected in the previous dataset. All in all, rules can be derived for four out of five cluster partitions in the case of the ARD this time. The corresponding TM results show a similar behaviour. Individual topics can be detected for all clusters except for the clusters 1 and 5 so that the complementary function of this holistic analysis is emphasised when looking to the fact that no association rules could be derived for cluster 3 (see Table 2; Fig. 5).

Hence, topics can only be derived for three of five cluster partitions using the LDA within the scope of the PCA-dataset. The term {*offer*} is placed in the foreground for the first topic of the second cluster which is supported in the second topic by addition-

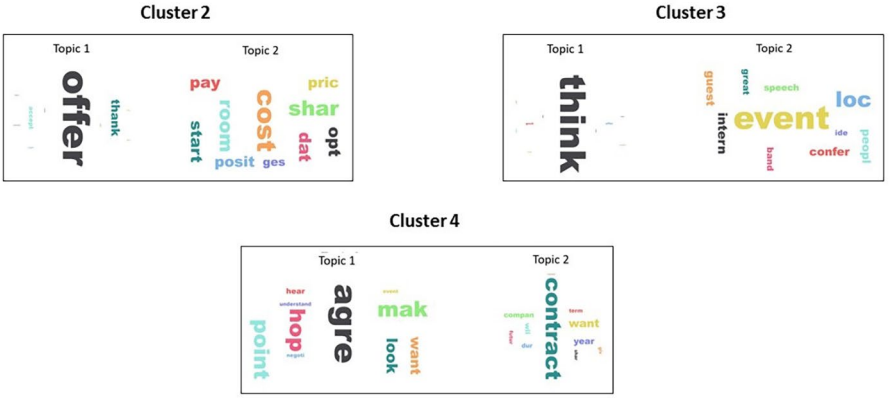

Fig. 5 Topics of PCA-dataset with k-Means (k = 5)

Table 3 Extracted Association Rules of PCA-dataset with k-Means (k = 7)

Cluster-Partition	Rule	sup	conf	lift
cluster_1	*joint, ventur* → *joint_ventur*	0.014	1.000	70.815
cluster_3	*accept_offer* → *accept*	0.100	1.000	6.287
cluster_4	*think_good* → *good*	0.040	1.000	10.501
cluster_6	*hop_hear* → *hop, hear*	0.018	1.000	51.668
	point_view → *view*	0.028	1.000	32.666
	mak, mak_conces → *mak, conces*	0.027	1.000	29.770
cluster_7	*look, for, hear* → *look_for*	0.092	1.000	344.685

ally enumerating contract-specific issues. The third cluster clearly places the concept {*think*} in the centre and again generates relevant contract objects for the second topic. For the last partition (cluster 4), the terms {*agree, hope, make, look, want*} are used, whereas in the second topic the central term of contract implementation is paraphrased with contract-specific issues objects.

The deduction of association rules based on the PCA-dataset with k-Means (k = 7) shows that the ARD was unable to derive rules for the clusters 2 and 5 (see Table 3). However, again contract-specific content could be derived for the first cluster. While the third cluster points to the acceptance of offer, the rules of the fourth cluster refer to the aspect of reflection. In addition, this aspect is accompanied by a positively connoted term {*good*}. Three rules stand out as highly significant for the sixth cluster and reference to the point of view, the aspect regarding "hope to hear" and reveals a clear indicator with regard to the concession making aspect.

The corresponding analysis of TM represents two topics for six of seven originally generated patterns. While association rules could not be derived for cluster 2 as presented in Table 3, topics can be generated for the second cluster partition of the PCA-

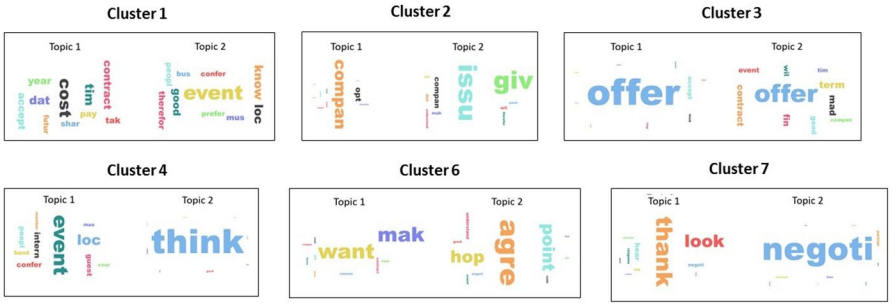

Fig. 6 Topics of PCA-dataset with k-Means (k = 7)

Table 4 Extracted Association Rules of PCA-dataset with DBSCAN

Cluster-Partition	Rule	sup	conf	lift
cluster_1	*relationship, term* → *term_relationship*	0.014	1.000	73.333
	ventur → *joint, joint_ventur*	0.014	1.000	73.333
cluster_3	*joint, ventur* → *joint_ventur*	0.016	0.997	63.912
	look, for → *look_for*	0.022	0.997	44.484

dataset with k-Means (k = 7) (see Fig. 6). Both topics of the first cluster deal with contractual subjects. The same phenomenon can be observed for two topics of the second cluster. Here, the terms like {*company*} for topic 1 and {*give, issue*} for topic 2 are especially highlighted. The third cluster contains topics related to offer. While in the first topic {*offer*} plays a central role beside the less weighted words {*accept, think, hope*}, {*offer*} surprisingly reappears in the second topic in connection with a list of contract items. The fourth cluster divides the topics into contract objects (topic 1) and into a concept of reflection {*think*} in topic 2. For the first topic of the sixth cluster, the concepts of the desire to act {*want, make*} are decisively emphasised. The corresponding topic 2 touches on aspects such as {*agree, hope, point*}. The last partition focuses on the terms {*thank, look, hear*}, whereas the second topic focuses on negotiation.

The last separation of the PCA-dataset whose partitions were detected by the DBSCAN clustering approach generates a majority of association rules with respect to contractual objects. The method detects rules focussing on relational aspects in the further course as presented in cluster 1. Both rules of the first cluster show significant results with an increased lift value and represent ideal confidence values of 1 (see Table 4). In the context of the third cluster, rules are derived with respect to contractual objects of the joint venture. Furthermore, indicators can be derived regarding a future-oriented phrase of “I look forward to hearing from you”. Nevertheless, the ARD is not able to derive rules for the second cluster of the dataset.

This finding is confirmed for the TM-results of the PCA-dataset with DBSCAN in Fig. 6. No topic results can be derived for the second cluster again. This gap is

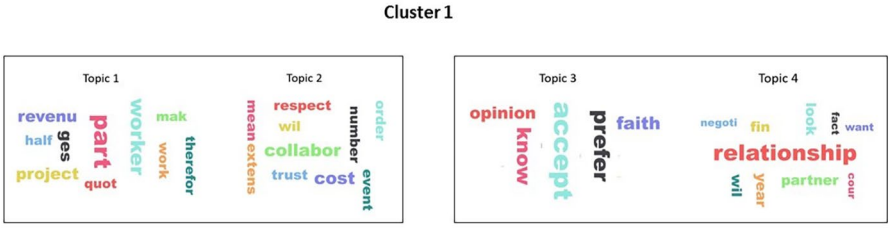

Fig. 7 Topics of PCA-dataset with DBSCAN

complemented by an additional finding that also no topics can be derived for the third cluster this time, so that topics can be derived only for the first cluster of the dataset. However, the DBSCAN-based PCA-dataset is to be treated in a special way compared to the other datasets in our study. While two topics are recommended for almost every determined cluster pattern in the internal evaluation of the datasets, four topics are recommended for the first cluster of the PCA-dataset with DBSCAN (see Fig. 7). A closer look at the first two topics reveals that both topics focus on contractual objects, while the third topic focuses on terms such as {*accept, prefer, faith, know, opinion*}. Finally, the fourth topic emphasises aspects of the relationship {*relationship, partner*} in addition to a mix of contractual objects.

Table 5 summarises the core indicators derived by the applications of ARD and TM and presents an overview of all label indicators that were derived cluster by cluster. It is noticeable that two cluster partitions exist for which descriptive indicators could not be derived neither by the use of ARD nor by the use of TM. There exist some results whose descriptions could only be derived by either ARD or by using TMs. This occurs in 5 out of 20 cases (see Table 5). Hence, indicators could be generated for 75% of the cluster partitions.

In addition to the results of the two previous applications, predictive DTs in the form of RF, GBDT and C4.5 DT are applied as presented in the research procedure (see Fig.3). It can be observed for the RF method that the tree cannot be reduced to an interpretable dimension despite iterative optimisations of the maximum tree depth, minimum gain and despite applying a pre-pruning. Furthermore, a closer look at the confusion matrix reveals that the RF does not work on high-dimensional data. For example, precision and recall values of 0% are obtained for six out of seven cluster partitions of the PCA dataset (k = 7). Similar results could be observed for the GBDT approach. The corresponding tree is certainly capable of learning on the high-dimensional data with an average accuracy over all four datasets of 84.33%. Nevertheless, the interpretation of the tree is not guaranteed by the intensive breadth and depth. Consequently, no descriptions of the cluster patterns can be derived for both the RF and the GBDT on high-dimensional business communication sentences.

A somewhat more promising result is obtained by applying the C4.5 algorithm on the high-dimensional data. To counteract the problem of tree expansion, a tree with pre-pruning (minimal gain) and pruning (pessimistic error calculation) are applied within the optimisations. The application of pruning methods results in trees being reduced to such an extent that only one root node is finally present. An exception can be observed for the DBSCAN-dataset. Corresponding patterns provide results

Table 5 Overview of derived Label Indicators

Cluster Separation Approach	Cluster Partitions	Label Indicators
OS-dataset with k-Means (k=5)	Cluster 1	*thanking, looking to the future, contract-specific issues*
	Cluster 2	*offer-related issues, reflection (think), acceptance of offers*
	Cluster 3	*contract-specific issues*
	Cluster 4	*contract-specific issues, requests*
	Cluster 5	*concession making, agreement, hope*
PCA-dataset with k-Means (k=5)	Cluster 1*	*contract-specific issues*
	Cluster 2	*thanking, offer-related issues (final offer), contract-specific issues*
	Cluster 3**	*reflection (think), contract-specific issues*
	Cluster 4	*reflection (think), agreement, hope*
	Cluster 5*	*looking to the future*
PCA-dataset with k-Means (k=7)	Cluster 1	*contract-specific issues*
	Cluster 2**	*contract-specific issues*
	Cluster 3	*offer-related issues, acceptance of offers*
	Cluster 4	*reflection (think good)*
	Cluster 5	-
	Cluster 6	*want to make, agreement, hope*
	Cluster 7	*thanking, looking to the future*
PCA-dataset with DBSCAN	Cluster 1	*contract-specific issue, preference, agreement, relationship*,
	Cluster 2	-
	Cluster 3*	*contract-specific issue, looking to the future*

* only ARD Results could be derived

** only TM Results could be derived

for the description of the classification decision of the first and third cluster. The model comes to the conclusion that the majority use of the following words in the negotiation sentences leads to classification into the first cluster: {*prefer, faith, opinion, prefer_faith, prefer_faith_opinion*}, so that these words represent a significant component of the first cluster according to the developed DT model. In turn, the model concludes that the avoidance of the listed terms in the negotiation sentences by at least more than half indicates a classification into the third cluster group of the DBSCAN-based PCA dataset.

5 Discussion

Taking into account the presented research procedure (in Fig.3) and ML-based process pipeline (in Fig.8 of the appendix), a wide range of results describing the patterns could be derived. While some methods are better able to contribute to the description of patterns than others, certain methods turned out to be less suitable for our application context of negotiation sentences. An analytical comparison of intra-method results shows that important indicators for describing the cluster partitions could be derived.

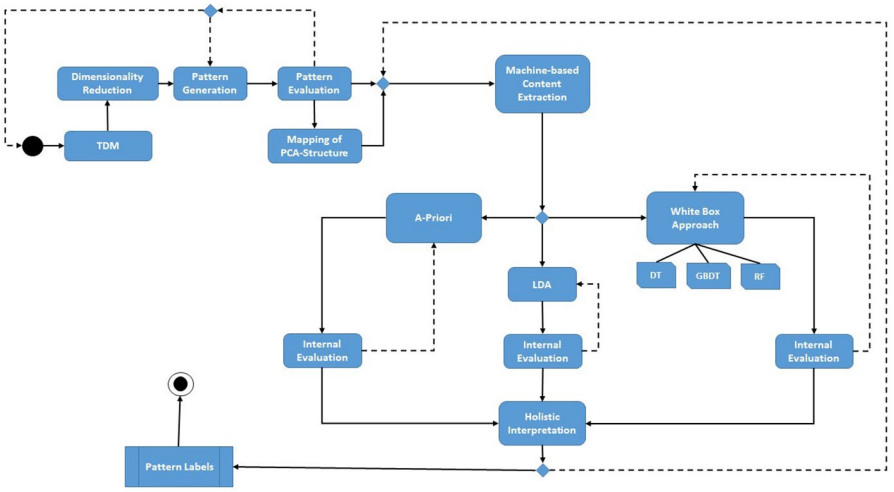

Fig. 8 ML-based Process Pipeline

A closer look at the examined partitions of the ARD results reveals that rule descriptions can be derived for all partitions of the OS dataset (k = 5), so that 16 partitions out of a total of 20 cluster partitions are described by association rules. All derived rules are confident according to the considered quality criteria. It is noticeable that each dataset contains at least one cluster that refers to contract-specific content across all datasets (see Table 5). This observation is also supported by further highly significant rules {*term, contract* → *contract, term*} which indicate the special importance of contractual topics. It should be also emphasised that the Apriori algorithm of ARD comes to the same conclusion despite the use of differently prepared datasets and the consideration of different clustering methods. Furthermore, the rule constellation from the itemset {*look, for, hear*} clearly occurs in each dataset in at least one cluster that points to the future-oriented statement to continue the negotiation interaction with "I am looking forward to hear". Indicators regarding the acceptance of offers {*accept_offer* → *offer, accept*} are addressed in the second cluster of the OS-dataset (k = 5) and in the third cluster of the PCA-dataset (k = 7) as one of the main goals of negotiation processes (see Tables 1 and 3). In this context, two more rules are derived from cluster 2 based on the PCA-dataset (k = 5) linking the final offer to a rule for the acknowledgement in the same cluster partition. The reference to the final offer and the acceptance of offers are considered separately according to the sample partition. An inter-data comparison also reveals that the rule {*wil_mak* → *mak, wil*} from cluster 5 (Table 1) indicates a striving for action, although the type of action is not yet specified at the current stage. In further steps, another rule is determined on the basis of the PCA dataset in Table 3, which points to the term "make". However, this time the action is connected with the elementary element of negotiating concessions: {*mak, mak_conces* ◊ *mak, conces*}. The linking of both rules could thus specify the action to make concessions. A sole universal position is pursued by the rule from

Table 4, which emphasises the importance of relationship building within the framework of a rule constellation: {*relationship, term → term_relationship*}.

A total of 32 topics could be derived for 15 cluster partitions with regard to TM which was applied to the individual cluster partitions. Comparable to the results of the ARD, at least 10 topics point to contractual subjects of the negotiations that represent about one-third of all topics. When looking at the TM results as a whole, it is noticeable that the topic with the strongly weighted term *{offer}* appears for almost every dataset. It can be observed in this context that both topics even consider the term as a central element in the internal comparison based on the two topics from cluster 3 of the PCA dataset (k=7) (see Fig.6). While the term {*think*} is part of two significant association rules, this term also plays an elementary role here in the context of three cluster partitions. This underlines the importance of the use of this term immensely. The topic of the agreement {*agre*} appears in a general view in two topics (see Figs.4 and 5), whereas the terms {*accept, prefer*} represent another topic in the third cluster of the PCA-dataset (see Fig.7). Apart from that, {*relationship*} represents a universal position for the fourth topic of the PCA-dataset of DBSCAN, similar to the ARD.

Linking the results within the framework of a holistic interpretive approach opens up new potentials and either enables the derivation of new information or enriches existing information with rich complementary indicators. Consequently, it is noticeable that common elements can be observed despite the use of partially divergent preparation methods of Text Mining and different clustering procedures when a meta-perspective is taken into account. A summarising overview of core label indicators is provided in Table 5. When looking at the results holistically, it can be seen for the results of the Apriori as well as for the LDA approach that the statement “look forward to hearing” repeatedly appears in rules and in topics which is often used as a transition for following negotiation interactions. Looking at the summary overview as a whole, it is noticeable that this “looking to the future” observations appears in each cluster separation approach in exactly one cluster (see Table 5). This indicates that this word construct was determined to be a significant sentence pattern despite the evaluation of various data preparation and transformation steps, pattern recognition procedures and application of ML procedures to derive descriptive indicators. While the corresponding association rules repeatedly make use of the itemset {*look, for, hear*}, complementary terms such as {*greet, thank, hope*} (cluster 1 in Fig.4) and {*thank, look*} (cluster 7 in Fig.6) are highlighted. Furthermore, a comparison of both types of results shows that the described constellation always results in a mixture of contractual items. Contract-specific issues are addressed in the same context in addition to the presented statement. This can be particularly shown by this holistic approach, since a clear differentiation to contract issues takes place within the framework of the topics for the same cluster partition. A look at the distribution of this indicator across all clusters shows that contract-specific issues always appear in exactly two different cluster partition of different cluster separation approaches (see Table 5). A communication-based transition is elementary in order to be able to achieve the personal negotiation goals and to maintain the negotiation process. Negotiators argue for their demands and express their willingness to hear from their negotiating partner. These statements can be underlined with words with positive connotations such as

hope and thanks in accordance with the derived results (Sokolova and Szpakowicz 2005). The following cluster partitions provide indicators for the continuation of the negotiation interaction: cluster 1 of OS-dataset (k=5), cluster 4 of PCA-dataset (k=5), cluster 7 of PCA-dataset (k=7) and cluster 3 of DBSCAN-dataset. A holistic view reveals that the association rules and the topics particularly emphasise the use of contract items. For the TM in particular, this is made even clearer by the enumeration of the individual negotiation issues, whereas the term {*contract*} also exists occasionally alongside significant issues for the ARD. In addition to the enumeration of issues, the contract items are put into context by the rules of the ARD. E.g. enumerated contract issues are brought into the context of the offer for the first cluster of the PCA-dataset (k=5), whereas for the fourth cluster of the OS-dataset (k=5), issues are connected into a request {*want*}. The explicit integration of negotiation issues into business communication is thus by no means negligible, so that 25% of the clusters focus on negotiation issues (see Table 5). Contractual items can be used in many ways. Besides formulating explicit demands to achieve negotiation goals and using issues as so-called positional information, they can also be used as a means of information disclosure (Olekalns and Smith 2000). Especially in situations when negotiators realise that the process is not moving forward, the transparent sharing of information with regard to contract preferences can positively influence the negotiation (Thompson et al. 2010). Taking into account the results, following cluster partitions can consequently be labelled with indicators of negotiation issues: cluster 3 and cluster 4 from the OS dataset (k=5), cluster 1 and cluster 2 from the PCA dataset (k=7) and cluster 1 from the PCA dataset (k=5).

Since the achievement of a negotiation goal in connection with the achievement of an accepted negotiation reflects one of the elementary goals of a successful negotiation process, it is to be expected that this behaviour of striving for agreement is also represented in the business communication data. This assumption is confirmed in various pattern descriptions. The most significant association of the fifth cluster of the OS-dataset (k=5) explicitly determines the elements from the itemset {*hope, agree*}. This observation is supported by the word cloud of the corresponding TM, which considers the terms {*concession*} and {*make*} as relevant. Making concessions represents one of the central strategic elements in negotiations. As an integrative element, they can contribute significantly to reach negotiation agreements (Filzmoser and Vetschera 2008). Results from previous studies show that the use of concessions is increasingly used in the final phase of negotiations (Lim and Murnighan 1994; Moore 2004). This finding is represented by the descriptions from cluster 6 of the PCA dataset (k=7), where the rule {*mak_concession → mak, concession*} is associated with the heavily weighted terms {*want, make, agre*} from TM. Taking into account this information and the explicit mentioning of concessions, following partitions provide strong indicators regarding the expression of striving for an agreement: cluster 5 of OS-dataset (k=5), cluster 4 of PCA-dataset (k=5) and cluster 6 of PCA-dataset (k=7). However, pattern recognition could determine further partitions in terms of striving for agreement with an offer focus. For example, an additional pattern is determined for the OS-dataset in addition to cluster 5 (cluster 2), which differentiates with the rule {*accept_offer → offer, accept*} and the information coming from the word cloud. This cluster description can be observed even more clearly

for the third cluster of the PCA dataset (k = 7). All in all, the treatment of the contract agreement is explicitly addressed at various levels of the communication in the negotiation process. A reference is also made to the final offer in the further step which is recognisable in a separate cluster partition (cluster 2 of PCA-dataset (k = 5)). In order to be able to convince the negotiating partner with one's own points of view, ideas and arguments, a certain credibility and a certain level of trust must exist within negotiation processes. This is only possible through mutual relationship building (Koeszegi 2004). Previous studies show that trustworthy relationship building between partners leads to an increase in concessions and contributes to an increased likelihood of contract acceptance with higher joint utilities (Kaya et al. 2017; Lewicki and Polin 2013). The ARD results and the third- and fourth word clouds of the first cluster of the DBSCAN-dataset point precisely to this need by clearly demonstrating the relevance of the relationship. While the association rule determines the term {*relationship*} for the first time as part of a significant rule, the aforementioned word clouds make the subject area of the clusters clear through the terms {*accept, prefer, faith*} and {*relationship*}. All in all, the results of our study clearly show that especially the results of TM and ARD have a complementary character besides a validating character. While unilateral indicators allow conditional interpretability, further indicators of the other method contribute with additional descriptive indicators. In addition to this phenomenon, the trained C4.5 model of the DT provides added value regarding the interpretation of the first cluster. The results of the DT underline the relevance of the results of the TM and emphasise in this context that the increased use of the terms {*prefer, faith*} contribute to the classification in the first cluster.

6 Summary and Outlook

Pre-defined cluster patterns from high-dimensional business communication data could be examined in the context of this paper. From a holistic perspective, important behavioural clusters such as the significant use of contractual items, transitions to maintain the negotiation process, indicators of striving for an agreement as well as relationship building elements could be determined at the sentence level of negotiation communication. This clearly shows that the first derivation of label indicators for semi-automated pattern labelling is possible despite the broad dimensional density for the application area of high-dimensional negotiation communication data. Nevertheless, it should be noted that even if some of the methods were able to derive an elementary contribution for describing the cluster patterns, the predictive DTs had their difficulties. A certain part of the predictive white-box algorithms was not able to describe the characteristic with respect to the cluster patterns. Moreover, some clusters were not interpretable despite the derivation of descriptions. This confirms the finding of Chang et al. (2009) that the goal of achieving highly valid performance values of investigated models is not always accompanied by an increased potential for interpretation.

To relativise the generalisability of the derived statements, it should also be noted that the communication data were currently solely examined at the aggregation level of the negotiation sentences to have sufficient data available for the systematic pro-

cessing of the ML procedures. It is important to recall that the analysis of the extent to which indicators can be derived to describe cluster patterns reflects the overall goal of the present work. Consequently, final labelling cannot be concluded on the basis of these first results yet due to the fact additional analyses are needed. This opens potentials for the pattern recognition on additional levels of negotiation communication (e.g. negotiation messages, negotiation phases, etc.) in future research (Adair and Brett 2005; Schoop 2021b). The aspect that the dynamic character of negotiations has not been taken into account so far represents another limitation of this work. An exclusively static approach was considered on the level of negotiation sentences. Exemplary questions concerning the degree to which the structure of negotiation messages change over time (e.g. within the bilateral process of requests and answers) and how this variable course of negotiations can affect the negotiation outcome (accepted and rejected negotiations) are not considered until now. The consideration of additional aggregation levels focussing the dynamic character of negotiations can lead to further complementary communication patterns which could yield exciting results. A total of 20 non-disjunctive clusters could be determined in the current approach through data-driven evaluations of pattern recognition procedures based on varying business communication data on negotiation sentence level (Kaya and Schoop 2020, 2022). However, a review of previous category schemes from manual content analysis shows that negotiation researchers have included further disjunctive negotiation categories in their analyses. Those categories and subcategories are usually at a more fine-grained level (Adair et al. 2013; Koeszegi et al. 2006; Weingart et al. 2004). This could suggest that additional patterns might be hidden in the negotiation data. Nevertheless, it should also be noted that only one dimension of negotiation behaviour is targeted within this research paper. Even if the communication part of negotiations is very important for the reproduction of negotiation behaviour, there is also a contractual part of negotiations whose utility patterns were not considered in this research paper due to the complexity of the communication dimension. A holistic combination with this additional dimension would open new potentials to investigate patterns based on negotiation communication including contract-specific preference and utility information.

This research shows that pattern recognition and the associated interpretation of derived patterns are of fundamental importance. They offer great research potentials for the field of electronic negotiations. Therefore, a further in-depth analysis of determined behavioural communication patterns should be used for the optimisation of negotiation processes in future research milestones. For instance, it could be investigated to what extent determined and labelled communication patterns (e.g. concessions) exist in accepted and rejected negotiations. Furthermore, the question of analysing whether communication patterns exist that emerge in a significant way in specific situations – e.g. in certain negotiation phases as for example in the first or final phase of negotiations. This could provide new phase-specific knowledge. As described in the previous section, linking the pattern results to the dynamic nature of negotiations bear important potentials and can generate extra added value. Predictive models of recommender systems could reuse this kind of labelled communication patterns to derive predictive recommendations for the optimisation of negotiation processes by considering the dynamic character of negotiation interactions (Lu et al.

2015). This could enable a pro-active communication support for negotiators interacting in intertwined electronic negotiations. The results of this paper show that ML is able to infer important characteristic indicators of previously determined communication patterns, despite the challenges of complex natural language and dimensional data expansions. From a methodological point of view and by considering the future-oriented goal of generating recommendations regarding the optimisation of negotiation processes, one valuable next step could also be to analyse whether described communication patterns can be predicted using predictive ML methods. Detected cluster patterns, whose descriptive indicators were characteristically described within this work, could be mapped to labelled communication patterns. ML models might be used with the goal of predicting the right communication patterns in right negotiation situations. Such a research step could serve as a solid basis for the pattern-based generation of predictive recommendations. All in all, the semi-automated description of patterns represents an exciting milestone in ML and reveals great potentials. The world of pattern recognition is and remains at the heart of ML and is indispensable in the age of digital transformation where more and more data are being tracked in today's world.

7 Appendix

7.1 Functionality of used ML-methods

This chapter provides additional insights into the detailed functionalities of the applied ML methods, their procedural execution flows as well as their quality criteria which were taken into account during the internal evaluations. In detail, the following approaches were followed: ARD with the Apriori approach, TM with the Latent Dirichlet Allocation approach (LDA) and the three predictive DT approaches named by C4.5 DT, GBDT and the RF technique.

7.1.1 Association Rule Discovery: Apriori

The Apriori algorithm determines potential candidates from so-called frequent item sets in the data as a first step. In our application case, item sets are defined as negotiation words that are frequently occurring in a set of clustered negotiation sentences. Only those set of words are selected as potential candidates whose relative frequencies exceed a pre-set global support value (sup_{min}) which represents a first quality criterion (Zheng et al. 2001). The support measure is calculated as the percentage of all possible word-interrelations in a given dataset and gives insight how frequently a collection of words occurs together (Dongre et al. 2014). In the following step, Apriori generates a set of association rules in form of if-then relationships based on all determined frequent candidates. A final check is executed to see whether each rule decomposition generated by the algorithm satisfies the minimum confidence

($conf_{min}$) for each potential association rule. The confidence measure represents the frequency of relationships where the then-part of the rule is found true (Dongre et al. 2014). Finally, a lift measure is additionally taken into account. It provides valuable information about to which extent the rule items correlate with each other and reveals new information about the degree of independence between the candidates (Azevedo and Jorge 2007). Consequently, only those rules are made available as a final result which fulfil all quality criteria in the context of the internal evaluation.

7.1.2 Topic Modelling: Latent Dirichlet Allocation

Latent Dirichlet Allocation (LDA) by Blei et al. (2003) represents a renowned sub-method of TM. The basic assumption of the method is that the negotiation sentences presented in the clusters consist of a mixture of topics that are derived from a fixed vocabulary of used negotiation terms (Blei 2012; Boyd-Graber et al. 2014). The general idea of the LDA approach is based on an iterative selection of random text segments. In each of these segments, the statistical cumulation of word groups is recorded. The algorithm calculates the topics of the text collection, the topic proportions in the individual texts and the interrelationships between words and corresponding topics. An iterative learning loop starts after having assigned the words to one of the topics by using the so-called Gibbs Sampling. The Gibbs algorithm pursues the goal of iteratively approximating the probable topic distribution. For this purpose, a value is assigned from e.g. one word of the negotiation set which is based on a conditional probability by also considering the other words (Griffiths 2002; Griffiths and Steyvers 2004).

But the optimisation procedure of LDA needs a parameter input regarding the (optimal) number of desired topics. This step is conducted via a data-driven internal evaluation where the rate of perplexity change (RPC) heuristic is used. RPC considers the perplexity measure which calculates how well a derived language model fits to a given word distribution in a text corpus (Zhao et al. 2015). Perplexity aims to predict the word sequences from a vocabulary which are associated with a topic by taking the probability distribution of word sequences into account (Kastrati and Imran 2013; Wang et al. 2012). This enables a data-driven division into a certain number of sub-topics (Correia and Gonçalves 2017). Based on this information, a series of optimisations are conducted: LDA calculates the probability of topics belonging to a document and the probability of words belonging to a topic. All calculated probabilities are assigned to the words and topics. This assignment of optimised probabilities is executed until a pre-defined upper-bound regarding the number of iterations is reached (Griffiths and Steyvers 2004).

7.1.3 C4.5 Decision Tree

The C4.5 algorithm by Quinlan (1993) represents one of the renowned possibilities to generate DTs as described in Chap.2.3. It uses an information gain ratio (a metric to measure the suitability of attributes to a pre-defined class label) to determine splitting

points for each of the considered attributes (Batra and Agrawal 2018). Based on this first step, a sorted list of normalised information gain ratios is calculated for each of the attributes and the attribute with the highest normalised information gain in the list is initially considered for the determination of a decision node which splits the data in a most effective way. Subsequently, a recursion is executed on respective sub-lists by splitting the attribute so that additional nodes are added as children of the nodes (Quinlan 2014). The overall goal of C4.5 is to derive a DT that is as compact as possible (regarding the decision paths) and exhibits a high predictive power with respect to the pre-set target classes (in our case: clusters of negotiation sentences).

7.1.4 Gradient Boosting Decision Tree

The GBDT follows the overall goal of optimising DTs in a sequential way to generate strong learners from initially weak learners. While so-called weak learners are defined as trees that generate errors and still need to be optimised, strong learners represent trees that have been improved by optimising loss functions to a pre-set threshold (Freund and Schapire 1999). So-called residuals (also known as error rate) are calculated and considered during the optimisations. In the first step, an initial DT with predefined target classes is applied to the entire data set where a weak learning DT is derived. Based on this, a calculation of the residuals is executed and used as target variable in the next step to predict the following DT (Ayyadevara 2018). Consequently, the information (including error experience in form of residuals) from the previous tree is taken into account when predicting the next DT. In each step, all the predicted DTs are connected with each other in a sequential process and are improved with respect to their prediction performances until a final DT is achieved whose tree representation fulfils the defined quality criteria (Friedman 2001). This is done by a so-called boosting algorithm where each derived DT follows the goal of minimising the error of the previous DT in the process pipeline. The use of GBDTs is recommended for the application on high-dimensional data even if it requires expensive computation time under certain circumstances.

7.1.5 Random Forest Decision Tree

The RF method represents an ensemble approach in which a set of DTs is predicted. In this prediction process, individual DTs are generated in a randomised and uncorrelated manner such that each of the predicted DTs derives individual classification decisions by declaring internal decision-rules. The underlying RF algorithm specifies how many DTs are to be generated and combines these trees by a systematic ensemble approach to a superordinate DT by taking into account the average of DT predictions (Cutler et al. 2012). While doing this, RF uses the results of derived decision-rules of a large number of DTs. All in all, a so-called forest of differentiating DTs is generated as a result of the ensemble process where the goal is to reach a maximisation of prediction probability through the variance of individual DTs (Khan et al. 2020). RF has been able to show good classification performance in previous

experiments and can counter the problem of overfitting (Banfield et al. 2007; Pappu and Pardalos 2014).

7.2 ML-based Process Pipeline of the Research Procedure

The ML-based process milestone in Fig. 8 presents some of the preliminary milestones which were needed to derive structural cluster patterns. All milestones starting with "TDM" up to "Machine-based Content Extraction" were conducted in previous studies and serve for the data basis of this research paper (Kaya and Schoop 2020, 2022). In the first step of the overall process pipeline, negotiation sentences were subjected to some Text Mining processing steps by using frequency measures and were mapped into a Term Document Matrix (TDM) in a first step. TDMs stand for a mathematical matrix representation and map documents (in our case: a series of negotiation sentences) to term dimensions (in our case: words derived from a set of negotiation sentences) (Antonellis and Gallopoulos 2006)

Subsequently, expanding matrices were processed to elaborate dimension reduction procedures in the Dimensionality Reduction milestone. Here, Feature Extraction (FE) and Feature Selection (FS) in form of Principal Component Analysis (PCA) and Optimise Selection (OS) were applied to be able to counteract the so called "curse of dimensionality" (Song et al. 2010; Venkatesh and Anuradha 2019). This phenomenon occurs if too many word dimensions are generated during the transformation process of qualitative negotiation sentences into quantitative matrices (Köppen 2000). While PCA aims at extracting reduced, approximated dimension combinations from a set of dimensions by calculating meaningful linear combinations (Abdi and Williams 2010), the OS-approach focuses on the effective selection of data dimensions instead of extracting dimensions by using intelligent ML-based greedy heuristics that approximate local optima to the global optimum (Venkatesh and Anuradha 2019)

Selected partitioning and density-based clustering methods were applied and evaluated on the basis of the prepared database in the milestone of "Pattern Generation". The goal was to extract the most valid cluster patterns from high-dimensional matrices. Detected patterns had to ensure a certain interpretability in this sense. This was the reason why some of the TDMs that were reduced with FE have been transformed into interpretable features of FS after the determination of existing structures to guarantee the interpretation potential while retaining the structural properties

After necessary preparations have been made and significant patterns could be extracted from the high-dimensional communication data, the presented research procedure in Fig. 3 starts concerning the derivation of semantic indicators for the labelling of cluster patterns. The ML-based process pipeline in Fig. 8 reveals a detailed view at the algorithms used and their internal evaluation loops as respectively described in the previous subsection of the appendix chapter. Considering this procedural flow, it is noticeable that each applied approach independently aims at deriving valid results on the methodological level in the form of either topics, associations or decision paths after fulfilling their interval evaluation criteria. All these results are interpreted from a holistic point of view in the last step to be able to assign a suitable label description to the integrated cluster patterns.

Authors' Contributions All authors/the author contributed (equally) to the study conception, design, data preparation, model application as well as optimisation and subsequent methodological analysis. All authors/the author read and approved the final manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data Availability The data is available on request due to data size reasons. As described in the abstract, this paper operates on high-dimensional communication datasets so that the underlying databases to some extent cover several gigabytes.

Code Availability The code is available on request and was used for the analysis as well as model training which was partly executed via a process-oriented ML-tool named by RapidMiner.

Declarations

Funding The author(s) did not receive any funding from organisations or institutions for the submitted research paper. The work is completely self-financed.

Conflicts of interest/Competing Interests *The author(s) has/have no relevant financial or non-financial interests to disclose.*

References

Abdi H, Williams LJ (2010) Principal component analysis. Wiley interdisciplinary reviews: computational statistics 2:433–459

Adair WL, Brett JM (2005) The negotiation dance: Time, culture, and behavioral sequences in negotiation. Organ Sci 16:33–51. https://doi.org/10.1287/orsc.1040.0102

Agndal H (2007) Current trends in business negotiation research. Stockholm School of Economics Research Paper 3:1–55

Agrawal R, Srikant R (1994) Fast Algorithms for Mining Association Rules in Large Databases. Proceedings of the 20th International Conference on Very Large Data Bases, San Francisco, CA, pp487–499

Ahmed S, Paul R, Hoque ASML (2014) Knowledge discovery from academic data using Association Rule Mining. IEEE 17th Int. conference on computer and information technology, pp314–319. https://doi.org/10.1109/ICCITechn.2014.7073107

Alghamdi R, Alfalqi K (2015) A survey of topic modeling in text mining. Int J Adv Comput Sci Appl 6:147–153. https://doi.org/10.14569/IJACSA.2015.060121

Amatriain X, Jaimes A, Oliver N, Pujol JM (2011) Data mining methods for recommender systems. Recommender Systems Handbook. Springer, Boston, pp 39–71

Antonellis I, Gallopoulos E (2006) Exploring term-document matrices from matrix models in text mining. arXiv preprint cs/0602076

Ayyadevara VK (2018) Gradient boosting machine. Pro machine learning algorithms. Apress, Berkeley, CA, pp 117–134

Azevedo PJ, Jorge AM (2007) Comparing rule measures for predictive association rules. European Conference on Machine Learning. Springer, Berlin, Heidelberg, pp510–517. https://doi.org/10.1007/978-3-540-74958-5_47

Babi C, Babi D, Rao MV, Rao VV (2017) Study of Association Rule Mining for Discovery of Frequent Item Sets on Big Data Sets. Int J Appl Eng Res 12:12169–12175

Banfield RE, Hall LO, Bowyer KW, Kegelmeyer WP (2007) A Comparison of Decision Tree Ensemble Creation Techniques. IEEE Trans Pattern Anal Mach Intell 29:173–180. https://doi.org/10.1109/TPAMI.2007.250609

Basit T (2003) Manual or electronic?: The role of coding in qualitative data analysis. Educational Res 45:143–154. https://doi.org/10.1080/0013188032000133548

Batra M, Agrawal R (2018) Comparative analysis of decision tree algorithms. Nature inspired computing. Springer, Singapore, pp 31–36

Bichler M, Kersten G, Strecker S (2003) Towards a Structured Design of Electronic Negotiations. Group Decis Negot 12:311–335. https://doi.org/10.1023/A:1024867820235

Biggio B, Roli F (2018) Wild Patterns: Ten Years After the Rise of Adversarial Machine Learning. Pattern Recogn 84:317–331. https://doi.org/10.1016/j.patcog.2018.07.023

Bishop CM, Nasrabadi NM (2007) Pattern Recognition and Machine Learning.Journal of Electronic Imaging16

Blei DM (2012) Probabilistic topic models. Commun ACM 55:77–84. https://doi.org/10.1145/2133806.2133826

Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet allocation. J Mach Learn Res 3:993–1022

Boyd-Graber J, Mimno D, Newman D (2014) Care and feeding of topic models: Problems, diagnostics, and improvements. Handbook of mixed membership models and their applications, pp3–41

Byrnes KG, Kiely PA, Dunne CP, McDermott KW, Coffey JC (2021) Communication, collaboration and contagion: "Virtualisation" of anatomy during COVID-19. Clin Anat 34:82–89. https://doi.org/10.1002/ca.23649

Carbonneau RA, Kersten G, Vahidov RM (2008) Predicting opponent's moves in electronic negotiations using neural networks. Expert Syst Appl 34:1266–1273. https://doi.org/10.1016/j.eswa.2006.12.027

Chang J, Boyd-Graber J, Gerrish S, Wang C, Blei D (2009) Reading tea leaves: How humans interpret topic models. Proceedings of the Advances in Neural Information Processing Systems Conference, pp1–9

Chen Y, Zhao X, Lin X, Wang Y (2015) Towards frequent subgraph mining on single large uncertain graphs. IEEE International Conference on Data Mining, pp41–50. https://doi.org/10.1109/ICDM.2015.88

Cherfi A, Nouira K, Ferchichi A (2018) Very fast C4. 5 decision tree algorithm. Appl Artif Intell 32:119–137. https://doi.org/10.1080/08839514.2018.1447479

Correia A, Gonçalves A (2017) Topics discovery in text mining. World Conference on Information Systems and Technologies. Springer, Heidelberg, pp251–256

Crowston K, Allen EE, Heckman R (2012) Using natural language processing technology for qualitative data analysis. Int J Soc Res Methodol 15:523–543

Cutler A, Cutler DR, Stevens JR (2012) Random forests. Ensemble machine learning. Springer, Boston, MA, pp 157–175

Debie E, Shafi K (2019) Implications of the curse of dimensionality for supervised learning classifier systems: theoretical and empirical analyses. Pattern Anal Appl 22:519–536. https://doi.org/10.1007/s10044-017-0649-0

Donohue WA, Diez ME, Hamilton M (1984) Coding naturalistic negotiation interaction. Hum Commun Res 10:403–425

Filzmoser M, Vetschera R (2008) A classification of bargaining steps and their impact on negotiation outcomes. Group Decis Negot 17:421–443. https://doi.org/10.1007/s10726-008-9106-1

Freund Y, Schapire R (1999) A short introduction to boosting. Journal-Japanese Soc Artif Intell 14:771–780

Friedman JH (2001) Greedy function approximation: a gradient boosting machine. Annals of statistics, pp1189–1232

Foroughi A (1998) Minimizing negotiation process losses with computerized negotiation support systems. J Appl Bus Res 14:15–26. https://doi.org/10.19030/jabr.v14i4.5648

Gan J, Qi Y (2021) Selection of the Optimal Number of Topics for LDA Topic Model—Taking Patent Policy Analysis as an Example. Entropy 23:1301

Garlanda C, Bottazzi B, Magrini E, Inforzato A, Mantovani A (2018) PTX3, a Humoral Pattern Recognition Molecule, in Innate Immunity, Tissue Repair, and Cancer. Physiol Rev 98:623–639. https://doi.org/10.1152/physrev.00016.2017

Griffiths T (2002) Gibbs sampling in the generative model of latent dirichllocation
Griffiths TL, Steyvers M (2004) Finding scientific topics. Proceedings of the National academy of Sciences 101: 5228–5235
Guidotti R, Monreale A, Ruggieri S, Turini F, Giannotti F, Pedreschi D (2018) A survey of methods for explaining black box models. ACM-CSUR 51:1–42. https://doi.org/10.1145/3236009
Hargie O, Dickson D (2004) Skilled Interpersonal Communication: Research, Theory and Practice, 4th edn. Routledge, London
Harinck F, Ellemers N (2006) Hide and seek: The effects of revealing one's personal interests in intra-and intergroup negotiations. Eur J Social Psychol 36:791–813. https://doi.org/10.1002/ejsp.321
Japkowicz N, Stephen S (2002) The class imbalance problem: A systematic study. Intell data Anal 6:429–449
Karanika A, Oikonomou P, Kolomvatsos K, Anagnostopoulos C (2020) On the Use of Interpretable Machine Learning for the Management of Data Quality. arXiv preprint arXiv. 2007.14677
Kastrati Z, Imran A (2013) Adaptive concept vector space representation using markov chain model. In Proceedings of the 19th International Conference on Knowledge Engineering and Knowledge Management. Linkoping, Sweden: Springer International Publishing, pp203–208
Kaur S, Agrawal R (2018) A detailed analysis of core NLP for information extraction. Int J Mach Learn Networked Collaborative Eng 1:33–47
Kaya MF, Körner M, Lenz A, Melzer P, Sepin S, Schmid A, Schoop M, van der Wijst P (2017) Does Trusting Behaviour Pay in Electronic Negotiations?. Proceedings of the 17th International Conference on Group Decision and Negotiation, pp 289–297
Kaya MF, Schoop M (2020) Maintenance of Data Richness in Business Communication Data. Proceedings of the 28th European Conference on Information Systems (ECIS 2020), An Online AIS Conference.
Kaya MF, Schoop M (2022) Analytical Comparison of Clustering Techniques for the Recognition of Communication Patterns. Group Decision Negotiation 31: 555–589. https://doi.org/10.1007/s10726-021-09758-7
Kersten GE, Lai H (2007) Negotiation Support and E-negotiation Systems: An Overview. Group Decis Negot 16:553–586. https://doi.org/10.1007/978-3-540-48713-5_23
Kersten GE, Noronha SJ (1999) WWW-based negotiation support: Design, implementation, and use. Decis Support Syst 25:135–154. https://doi.org/10.1016/S0167-9236(99)00012-3
Kersten G, Zhang G (2003) Mining Inspire Data for the Determinants of Successful Internet Negotiations. Cent Eur J Oper Res 11:297–316
Khalid S, Khalil T, Nasreen S (2014) A survey of feature selection and feature extraction techniques in machine learning. IEEE Science and Information conference, pp372–378. https://doi.org/10.1109/SAI.2014.6918213
Khan Z, Gul A, Perperoglou A, Miftahuddin M, Mahmoud O, Adler W, Lausen B (2020) Ensemble of optimal trees, random forest and random projection ensemble classification. Adv Data Anal Classif 14:97–116
Kim HD, Park DH, Lu Y, Zhai C (2012) Enriching text representation with frequent pattern mining for probabilistic topic modeling. Proceedings of the American Society for Information Science and Technology 49:1–10. https://doi.org/10.1002/meet.14504901209
Klich A (2021) Electronic Communication with Public Administration in the Time of COVID-19—Poland's Experience. Int J Environ Res Public Health 18:685. https://doi.org/10.3390/ijerph18020685
Koeszegi ST (2004) Trust-building strategies in inter-organizational negotiations. J Managerial Psychol 19:640–660
Koeszegi ST, Srnka KJ, Pesendorfer EM (2006) Electronic negotiations—a comparison of different support systems. Die Betriebswirtschaft 66:441–463
Köppen M (2000) The curse of dimensionality. In 5th online world conference on soft computing in industrial applications (WSC5) 1:4–8
Körner M, Schoop M (2017) Classifying Electronic Negotiations Based on Their Communicational Content. In: Schoop M, Kilgour DM (eds) Proceedings of the 17th International Conference on Group Decision and Negotiation, Stuttgart-Hohenheim, pp247–254
Lewicki RJ, Polin B (2013) The role of trust in negotiation processes. Handbook of advances in trust research. Edward Elgar Publishing, pp 29–54
Lewicki RJ, Saunders DM, Barry B (2010) Negotiation: Readings, exercises and cases. McGraw-Hill, New York
Li X, Zhang A, Li C, Ouyang J, Cai Y (2018) Exploring coherent topics by topic modeling with term weighting. Inf Process Manag 54:1345–1358

Lim SG, Murnighan JK (1994) Phases, deadlines, and the bargaining process. Organ Behav Hum Decis Process 58:153–171. https://doi.org/10.1006/obhd.1994.1032
Liss J (2011) Negotiating the Marcellus: the role of information in building trust in extractive deals. Negot J 27:419–446. https://doi.org/10.1111/j.1571-9979.2011.00317.x
Lovins JB (1968) Development of a stemming algorithm. Mech Transl Comput Linguistics 11:22–31
Lu J, Wu D, Mao M, Wang W, Zhang G (2015) Recommender system application developments: a survey. Decis Support Syst 74:12–32. https://doi.org/10.1016/j.dss.2015.03.008
Molnar C (2020) Interpretable machine learning. Lulu.com
Moore DA (2004) Myopic prediction, self-destructive secrecy, and the unexpected benefits of revealing final deadlines in negotiation. Organ Behav Hum Decis Process 94:125–139. https://doi.org/10.1016/j.obhdp.2004.04.001
Myles AJ, Feudale RN, Liu Y, Woody NA, Brown SD (2004) An introduction to decision tree modeling. J Chemometrics: J Chemometrics Soc 18:275–285
Olekalns M, Smith PL (2000) Understanding optimal outcomes. The role of strategy sequences in competitive negotiations. Hum Commun Res 26:527–557. https://doi.org/10.1093/hcr/26.4.527
Olekalns M, Putnam LL, Weingart LR, Metcalf L (2008) Communication processes and conflict management. The psychology of conflict and conflict management in organizations, pp81–114
Pappu V, Pardalos PM (2014) High-Dimensional Data Classification. In: Aleskerov F, Goldengorin B, Pardalos P (eds) Clusters, Orders, and Trees: Methods and Applications. Springer, New York. https://doi.org/10.1007/978-1-4939-0742-7_8
Pesendorfer EM, Koeszegi ST (2006) Hot versus cool behavioural styles in electronic negotiations: The impact of communication mode. Group Decis Negot 15:141–155. https://doi.org/10.1007/s10726-006-9025-y
Preuss M, van der Wijst P (2017) A phase-specific analysis of negotiation styles. J Bus Industrial Mark 32:505–518
Priyanka SK, Kumar D (2020) Decision tree classifier: A detailed survey. Int J Inform Decis Sci 12:246–269
Pruitt DG (2013) Negotiation behavior. Academic Press
Putnam LL, Roloff ME (1992) Communication and negotiation. Sage
Putnam LL, Jones TS (1982) The role of communication in bargaining. Hum Commun Res 8:262–280. https://doi.org/10.1111/j.1468-2958.1982.tb00668.x
Quinlan JR (1993) C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, San Mateo
Quinlan JR (2014) C4. 5: Programs for Machine Learning. Elsevier
Rangaswamy A, Shell GR (1997) Using computers to realize joint gains in negotiations: toward an "electronic bargaining table. Manage Sci 43:1147–1163. https://doi.org/10.1287/mnsc.43.8.1147
Schonlau M, Guenther N, Sucholutsky I (2017) Text mining with n-gram variables. Stata J 17:866–881
Schoop M (2005) A language-action approach to electronic negotiations. Syst Signs Action 1:62–79
Schoop M (2010) Support of Complex Electronic Negotiations. In: Kilgour DM, Eden C (eds) Handbook of Group Decision and Negotiation. Springer, pp 409–423
Schoop M (2021a) Negoisst: Complex Digital Negotiation Support. In: Kilgour DM, Eden C (eds) Handbook of Group Decision and Negotiation. Springer, Berlin, Heidelberg, pp 1149–1167. https://doi.org/10.1007/978-3-030-49629-6_1
Schoop M (2021b) Negotiation communication revisited. CEJOR 29:163–176. https://doi.org/10.1007/s10100-020-00730-5
Schoop M, Jertila A, List T (2003) Negoisst: A negotiation support system for electronic business-to-business negotiations in e-commerce. Data Knowl Eng 47:371–401. https://doi.org/10.1016/S0169-023X(03)00065-X
Sokolova M, Szpakowicz S (2005) Analysis and Classification of Strategies in Electronic Negotiations. Proceedings of the 18th Canadian AI. Springer, Heidelberg, pp145–157. https://doi.org/10.1007/11424918_16
Song F, Guo Z, Mei D (2010) Feature selection using principal component analysis. IEEE International conference on system science, engineering design and manufacturing informatization, pp27–30. https://doi.org/10.1109/ICSEM.2010.14
Srnka KJ, Koeszegi ST (2007) From words to numbers: how to transform qualitative data into meaningful quantitative results. Schmalenbach Bus Rev 59:29–57. https://doi.org/10.1007/BF03396741
Tandel SS, Jamadar A, Dudugu S (2019) A survey on text mining techniques. IEEE 5th International Conference on Advanced Computing & Communication Systems, pp1022–1026. https://doi.org/10.1109/ICACCS.2019.8728547
Thompson LL, Wang J, Gunia BC (2010) Negot Annual Rev Psychol 61:491–515

Vayansky I, Kumar SA (2020) A review of topic modeling methods. Inform Syst 94:101582
Vázquez FI, Zseby T, Zimek A, Analytics A (2020) (DSAA). IEEE, pp21–29
Venkatesh B, Anuradha J (2019) A review of feature selection and its methods. Cybernetics and Information Technologies 19:3–26. https://doi.org/10.2478/cait-2019-0001
Vetschera R (2007) Preference structures and negotiator behavior in electronic negotiations. Decis Support Syst 44:135–146. https://doi.org/10.1016/j.dss.2007.03.007
Vetschera R (2016) Concessions dynamics in electronic negotiations: a cross-lagged regression analysis. Group Decis Negot 25:245–265. https://doi.org/10.1007/s10726-015-9441-y
Wang Y, Agichtein E, Benzi M (2012) TM-LDA: efficient online modeling of latent topic transitions in social media. In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, pp123–131
Weigand H, Schoop M, Moor A, Dignum F (2003) B2B Negotiation Support: The Need for a Communication Perspective. Group Decis Negot 12:3–29. https://doi.org/10.1023/A:1022294708789
Weingart LR, Thompson LL, Bazerman MH, Carroll JS (1990) Tactical behavior and negotiation outcomes. Int J Confl Manage 1:7–31. https://doi.org/10.1108/eb022670
Weingart LR, Smith P, Olekalns M (2004) Quantitative coding of negotiation behavior. Int Negot 9:441–456
Weingart LR, Olekalns M (2004) Communication processes in negotiation: Frequencies, sequences, and phases. The handbook of negotiation and culture, pp143–157
Willett P (2006) The Porter stemming algorithm: then and now. Program
Zeng D, Sycara K (1998) Bayesian Learning in Negotiation. Int J Hum Comput Stud. https://doi.org/10.1006/ijhc.1997.0164. 48:125-141
Zhao W, Chen JJ, Perkins R, Liu Z, Ge W, Ding Y, Zou W (2015) A heuristic approach to determine an appropriate number of topics in topic modeling. BMC Bioinformatics 16:1–10. https://doi.org/10.1186/1471-2105-16-S13-S8
Zheng Z, Kohavi R, Mason L (2001) Real world performance of association rule algorithms. Proceedings of the seventh Int. conference on Knowledge discovery and data mining, pp401–406. https://doi.org/10.1145/502512.502572