Contents lists available at ScienceDirect







journal homepage: www.elsevier.com/locate/im

Using discussion logic in analyzing online group discussions: A text mining approach



Shasha Deng^a, Yilu Zhou^b, Pengzhu Zhang^c,*, Ahmed Abbasi^d

^a Shanghai International Studies University, School of Business and Management, 550 Dalian Road, Shanghai 200083, China

^b Fordham University, Fordham Schools of Business, 140 West 62nd Street, Room 413, New York, NY 10023 USA

^c Shanghai JiaoTong University, Antai College of Economics & Management, 535 Fahua Zhen Road, Shanghai 200052 China

^d University of Virginia, McIntire School of Commerce, Charlottesville, VA 22908 USA

ARTICLE INFO

Keywords: Online group discussion Discussion logic, Toulmin's model of argument Conversation disentanglement Coherence analysis Discussion representation

ABSTRACT

We propose a Discussion Logic-based Text Analytics (DiLTA) framework, which combines theories developed in social science and text mining fields. The framework extracts features that uncover discussion logic and uses these features in analyzing online discussions. A series of models are proposed including conversation disentanglement, coherence analysis, and visualization. Validation experiments showed that DiLTA achieved significantly superior performance over existing text analytics methods in reconstructing internal structure. A case study using DiLTA-enabled visualization on a healthcare forum illustrates the great potential of DiLTA in assisting comprehension of the internal linkage, structure, and logic of online group discussions.

1. Introduction

The rapid growth of online communities has dramatically changed the manner in which communication takes place [1]. Online group discussion is a type of computer-mediated communication (CMC) that involves multiple participants and is adopted in online interest sharing, collaborative work, online courses, gaming, technical support, and military command and control [1–4]. There are many situations in which the timely analysis of online group discussion is needed. Organizations can derive considerable benefits from using online group discussion, including increased speed of access to knowledge, enhanced identification of experts, increased number of successful innovations, and reduced costs of communication and operations [5,6]. Teachers can benefit from summarized online discussions for later review [7]. Moreover, military decision makers strive to gather important and urgent information from online discussions in a timely manner [2].

Structured discussion provides many advantages. It leads to quicker cognitive comprehension and deeper understanding and facilitates the decision-making process [8]. It also reveals the structure of social networks, highlighting the connectivity, clustering, and strengths of the relationships among users [9]. In social science, many efforts have been made to create or recover the structure of conversations to facilitate analysis. However, most of the analyses have to be done manually [10]. Vronay, Smith, and Drucker [11] manually coded for intermessage references to recover the structure of a chat log. Nash [12] manually

analyzed 1099 turns from Yahoo! Chat and found that the lag between a message and its response can be as many as 100 turns; she then concluded that nearly half of all turns were "off-topic." Holmer [10] used a combination of manual referencing and automatic content analysis to create a visualization of messages and interaction structures. Guided by Toulmin's model, Savolainen [13] manually identified argumentation patterns from Yahoo! Answers' 100 discussion threads. While these manual approaches tap into both semantic and pragmatic analyses, they cannot deal with large-scale data. Their work provides theoretical foundations for discussion analysis. Among them, Toulmin's model provides a schematic representation of the procedural form of argumentation [13–15]. It is widely verified and adopted in various bodies of discourse analysis research such as e-commerce [16,17], science education [18], and home security [19].

On the other hand, automatic analysis of online group discussions is developing in the areas of summarization, topic detection, and expert identification. Such automatic analysis focuses only on identifying the gist semantics of a discussion [20]. However, social science research requires a deeper understanding of information flow and user interaction behavior. These are difficult to detect automatically, especially when the reply-to structure is explicit and is hidden in the conversation context. Hidden online discourse structures prevent researchers from constructing a holistic and contextual presentation from interrelated messages [21,22]. While social scientists have largely applied the manual approach to reconstruct discussion sequences by adopting

* Corresponding author. E-mail addresses: shasha.deng@shisu.edu.cn (S. Deng), pzzhang@sjtu.edu.cn (P. Zhang), abbasi@comm.virginia.edu (A. Abbasi).

https://doi.org/10.1016/j.im.2018.09.013

Received 12 July 2016; Received in revised form 19 September 2018; Accepted 24 September 2018 Available online 27 September 2018 0378-7206/ © 2018 Published by Elsevier B.V. various schemas and models, to our knowledge, it has not yet been used in automatic discussion analysis.

There are two major challenges in automatic discussion analysis: intertwined and incoherent textual message contents [23]. Because discussion involves multiple users communicating at the same time, their message sequence is often disrupted. This further causes the "information entropy" problem, where incoming messages are not sufficiently organized by topic, and as a result, the content can be difficult to comprehend [24]. Intensive interactions among a large number of participants can further complicate this problem [25]. Thus, advanced analysis of online group discussions relies on the ability to identify subtopics and their corresponding messages as well as the ability to reconstruct message orders. The former is referred to as *conversation disentanglement* [26] and the latter as *coherence analysis* [27].

To address the gap in the existing research, we design and develop a Discussion Logic-based Text Analytics (DiLTA) framework to support the analytics of online discussions. This is the first body of research that links the theories and models related to argumentation and techniques developed in text analytics to automatic discussion analysis [28]. Our approach addresses two fundamental tasks in online discussion analysis: (1) conversation disentanglement and (2) coherence analysis. We evaluated DiLTA framework in a series of experiments that demonstrate the utility of two components in comparison with the existing methods. Furthermore, we show a tree-like representation—a discussion analysis tree (DATree)—generated from automatically reconstructed online discussions, and we discuss its usefulness.

The remainder of the paper is organized as follows. First, we review work related to discussion logic including Toulmin's model and how it has been widely used for manual analysis in social science research. We then review work related to the two major challenges in automatic online discussion analysis: conversation disentanglement and coherence analysis. The next section describes the design of DiLTA framework that attempts to enable better and deeper analysis by identification and incorporation of discussion logic in text analysis. The following section presents a robust evaluation of various facets of the proposed system and illustrates the usefulness of such automatic analysis in various content analytics scenarios. Finally, we conclude the paper with contributions and future directions.

2. Related works

2.1. Discussion logic

Online group discussion resembles argumentation and collaborative decision-making. Such process has been studied in social science where argumentation models are proposed to provide insights into the characteristics of online discussion [15,16,18,29]. For example, Toulmin's model [16] offers a generic representation of all arguments that commonly appear in daily communication, collaborative business decision making, and science education. It reveals the nature of the argument process, especially in tracing the solver's line of discussion, and can be used to guide automatic group discussion analysis. The original model has six argument elements, four of which appear frequently in daily communication: claim, data, backing, and rebuttal [16].

Other studies group the elements in Toulmin's model into *primitive* and *derivative* statements. A *primitive* message is a stand-alone assertion, and a *derivative* message is obtained as a strictly logical or defeasible consequence of others. For example, Raghu, Ramesh, Chang, and Whinston [15] modeled collaborative decision-making as a dynamic process in which individuals assert their positions through both *primitive* and *derivative* statements. Although an individual could make a primitive assertion, a cogent argument requires the assertions to be linked and organized in some logical sense. Therefore, the following logical argument structure proceeds in the form of response exchanges between its proponents and opponents. On the basis of simulation experiments for four types of decision problems, Carbogim, Robertson,

and Lee [30] found that a group discussion can be represented in a twostep process: "First, arguments are generated; then, arguments are evaluated in terms of their acceptability." Group discussion is an intertwined, repetitive process of subtopic and solution generation and its evaluation. We refer to the generation of primitive and derivative messages as *discussion logic*. The identification of discussion logic is the key to comprehending group discussion content.

In group support system (GSS) research, several studies have manually reconstructed discussion message sequences following argumentation and collaborative models [14,31]. Osborne [18] used the Toulmin's argumentation model to assess students' progression in capabilities with argumentation. Kim and Benbasat [16] used the model to assist with investigating the trust-assuring conversation of an Internet store. Raghu, Ramesh, and Whinston [19] used the connectionist mechanism to support collaborative decisions made by home security intelligence agencies. However, such manual processes are time consuming and labor intensive. Few automatic online discussion analyses attempt to uncover the discussion structure that hinders deeper analysis.

2.2. Conversation disentanglement

The identification of discussion logic is challenged by the nature of parallel, intertwined conversations [26]. Entangled conversations, which are highly prevalent in various forms of CMC, including Web 2.0 technologies, occur as a result of multiple simultaneous conversations between two or more users appearing within a single discussion thread [32]. To avoid thread confusion, disentanglement is widely regarded as an essential precursor for more advanced forms of discourse analysis [33]. It is especially important "when there are several streams of conversation and each stream must be associated with its particular feedback" [34]. Consequently, disentanglement is antecedent discussion logic identification.

Prior methods of disentanglement have mostly relied on single-pass clustering methods that compare newer messages against existing conversation clusters [30,33]. While these methods use information regarding content similarity and spatial/temporal proximity between messages, they do not incorporate information pertaining to conversation structure. According to Toulmin's model, a conversation can be decomposed into a beginning act that is succeeded by a series of "reacting" or "continuing moves" [35]. Hence, primitive message identification is of great importance for disentanglement [36]. However, existing disentanglement methods do not attempt to explicitly identify primitive messages. Elsner and Charniak [26] noted that a "detector for utterances which begin conversations could improve disentanglement scores." They empirically demonstrated that a hypothetical method that incorporated enhanced information regarding conversation beginnings could potentially improve performance over existing methods, and they considered this an important future direction. Consequently, given the importance of primitive messages, we believe that the identification of primitive messages in conversation structures is an important step in conversation disentanglement.

2.3. Coherence analysis

Text comprehension involves the construction of a coherent mental representation of situations described by texts. In online group discussions, coherence is represented in terms of reply-to relationships between messages [27]. Automatic coherence analysis attempts to offset the incoherent nature of online discourse by correctly reconstructing coherence relationships between messages using machine-learning algorithms. Two important facets of coherence analysis are the features and techniques that are used.

2.3.1. Coherence analysis features

Past research has used two categories of features to identify

coherence relationships in CMC: system and linguistic features.

System features provide insights into the message context, including header (e.g., date/time, message ID, and subject/title) and quotation information [20]. For instance, Netscan extracted the "contents of Subject, Date, Organization, Lines, Message ID and Reference lines" to generate relationships in Usenet newsgroups, including conversation trees [37]. Similarly, by detecting quoted text, Zest divided each message into contiguous blocks of quoted or unquoted text to support text analysis [38]. However, not all forms of group discussion contain a full range of system features. They are often omitted, especially in synchronous tools, where the only system information that can be extracted is the message ID and time stamp of each individual message.

Linguistic features derived from message content can also provide important cues for coherence analysis. A number of linguistic features have been researched for coherence analysis, including direct address, coreference, lexical relationship, and semantic information [12,27,39]. Direct address occurs when a reply message includes the screen name of the author of a previous message [39]. Lexical relationship is defined as a "cohesive relation where one lexical item refers back to another, to which it is related by having common referents" [12]. It has been shown to be effective for coherence analysis [27]. Coreference also occurs when a lexical item refers to a previously posted lexical item; however, in this case, the relationship is implicit in that it can only be identified by the context [40]. However, coreference identification is a difficult problem in natural language processing (NLP); thus, it is seldom used in coherence analysis [27]. With respect to semantic features, prior work has incorporated overall message similarity by simple textual analysis techniques, such as the bag-of-words approach [27]. Additional relevant forms of semantic information include opinions, emotions, synonymy information, and parts of speech (POS). Such features could be extracted by more advanced NLP techniques; however, they have not been widely adopted [20].

2.3.2. Coherence analysis techniques

Prior methods of coherence analysis include linkage, heuristic, classification, and manual annotation.

Linkage methods construct interaction patterns using predefined rules that are primarily based on system features and assumptions regarding message sequences [38,41,42]. Most linkage methods employ two types of rules: direct linkage and naive linkage [27]. Direct linkage rules assume that users follow system features to post messages and clearly quote the messages to which they respond. Naive linkage rules are then applied to residual messages that are unidentified by direct linkage; these rules assume that all residual messages are responding to either the first message in the thread or the previous message [43]. Linkage methods work fairly well with email-based discussion lists, where "reply-to" relationships are more explicitly defined in message headers, thereby allowing direct linkage rules to identify a large proportion of interactions [37,38,41]. However, group discussions are highly susceptible to disrupted turn adjacency [44]. Consequently, linkage methods have not performed well when applied to web forums or chat discussions [20,27].

Heuristic methods rely on metrics derived from observations of CMC discourse [27]. These metrics are based on a small, fixed assumed set of communication patterns pertaining to system and/or linguistic features [45]. For instance, the hybrid interactional coherence method uses an ordered list of heuristics, where messages unidentified by one heuristic are then evaluated by the next heuristic on the list [27]. Similarly, Khan, Fisher, Shuler, Wu, and Pottenger [36] used finite state automata to identify interaction patterns in multiperson chat rooms. In many of these methods, the choice of heuristics (and their order) was based on prior observations of occurrence [12,27]. Consequently, the effectiveness of heuristic methods is predicated on the validity and generalizability of the set of incorporated heuristics.

Classification methods formulate coherence analysis as a binary classification problem [46]. These techniques couple system and

linguistic features with supervised machine learning methods: predictive analytics algorithms that build models from a set of labeled training data [4]. For example, to handle highly incoherent text from student online forums, Kim and Kang [7] used supervised learning to classify discussion threads. Soon, Ng, and Lim [40] adopted a machine learning approach to identify coreferences of noun phrases both within and across sentences, which have been used for discourse analysis and language understanding. Ackerman [47] suggested that machine learning methods may someday help reduce the gap between computersupported cooperative work systems and user discourse patterns. However, as previously noted, the features used for interactional coherence analysis play a significant role. This is especially true in the case of machine learning classification methods: their performance is highly dependent on the feature sets employed [20]. Moreover, their ability to learn and adapt to different conversation environments makes them better suited for detecting conversation structures than heuristic methods [4].

Manual methods rely on human annotators, who read the entire discussion to reconstruct coherence relationships [48]. For example, Nash [12] manually identified the occurrence of linguistic features, including lexical relationships, direct address, and coreference, for coherence analysis. Linguistic experts are able to accurately identify coherence relationships in online discourse, because they possess a high level of pragmatic competence [49]. Conversely, because of the challenging nature of the task, most (nonexpert) users are less effective [44]. Regardless, manual methods typically outperform automated approaches. Hence, although manual annotation results are obviously not feasible for large-scale interactional coherence analysis, they are often included in experimental studies as an upper bound to shed light on the difficulty of the task [50].

Table 1 provides a summary of some important studies on coherence analysis. As we can see, most studies focus on asynchronous online communities where system features are available. A number of studies have used both linguistic and system features. Although classification is a promising technique in coherence analysis, it has not been widely adopted. None of the existing research has incorporated *discussion logic* as a feature of classification.

3. Discussion logic-based text analytics (DiLTA) framework

Group discussion is a repetitive process of subtopic generation and evaluation. As previously alluded to, this process often results in simultaneous parallel conversations within a single discussion thread [26]. Accordingly, social science studies can shed light on the relationships between messages and conversations within a discussion.

Table 1

Select previous automatic coherence analysis studies on group discussion text.

Previous Studies	Domains		Features		Techniques			
	D1	D2	F1	F2	T1	T2	Т3	T4
Barcellini, Détienne, Burkhardt, and Sack [48]	V		V		V			
Smith [37]			\checkmark					
Newman [42]	\checkmark							
Yee [38]	\checkmark							
Sack [41]	\checkmark							
Fu, Abbasi, and Chen [27]								
Kim and Kang [7]	\checkmark			\checkmark				
Aumayr, Chan, and Hayes [46]	\checkmark			\checkmark				
Nash [12]				\checkmark				
Khan, Fisher, Shuler, Wu, and Pottenger [36]		V		\checkmark			V	
Soon, Ng and Lim [40]		\checkmark		\checkmark				\checkmark

D1: asynchronous; D2: synchronous; F1: system features; F2: linguistic features; T1: manual method; T2: linkage method; T3: heuristic-based method, T4: classification-based method.

Tracing the discussion logic is at the core of these argumentation models. However, they have not been used much in automatic online discussion analysis research. Automatic machine learning methods that use discussion logic features are highly promising with regard to uncovering discussion structures. To our knowledge, previous automatic analyses have ignored such features, and these analyses are generic.

Most current research has studied only asynchronous social media. An important part of today's social media strongly resembles synchronous discussion and is more complicated in nature [12,51]. Thus, simple linguistic and system features alone may not be enough. We believe that more advanced NLP techniques are needed to derive discussion logic features. According to Raghu, Ramesh, Chang, and Whinston [15]'s model, we propose to extract *discussion logic* features to be attributes that can shed light on the relationships between messages and conversations within a discussion. Despite their importance for sense making [32], discussion logic features have not been used much in previous automatic discussion analysis research. Furthermore, to assist people in analyzing discussion structure, we propose the use of social network and tree-based graph representation to visualize the coherence analysis results.

To address these research gaps, we propose a DiLTA framework for discussing representation in online discourse (Fig. 1). The framework takes online discussion messages as input. DiLTA has two major computational components: conversation disentanglement and coherence analysis. The output is a tree-based representation called a DATree. A DATree shows disentangled conversations within a discussion thread, with coherence relationships between messages. Such representation can provide many desired insights for social science research. Moreover, the accuracy of a DATree depends on the performance of the two computational components. The conversation disentanglement component identifies primitive and derivative messages. This is achieved by applying an innovative revised cosine similarity (RCS) measure and a discussion segmentation algorithm (DSA). Coherence analysis further uncovers the discussion structure through linking all the derivative messages by identifying their reply-to relationship. This is done by a machine learning algorithm called transformation-based learning (TBL). The algorithm uses three types of features: system, linguistic, and discussion logic. A residual matching (RM) algorithm is applied to further improve the accuracy. We discuss the detailed design of each

component in this section.

3.1. Conversation disentanglement

The conversation disentanglement component of DiLTA uses linguistic features to compute intermessage similarity, which is then used as input for the proposed DSA.

3.1.1. Revised cosine similarity

Cosine similarity is a popular algorithm to measure the semantic similarity between two documents [52]. Traditional cosine similarity compares all terms that appear in two messages weighted by *tfidf*. Our RCS improves the traditional method with two important refinements: (1) using only the most important POSs to reduce the noise in messages and (2) group semantically similar words to avoid unnecessary entries in similarity vector. Research has shown that noun phrases and verb phrases carry most of the important meaning in a sentence, while conjunctions, adverbs, and adjectives are less important. Thus, we define meaningful terms to be nouns, noun compounds, named entities, verbs, and verb phrases. Instead of considering every term within a document, we focus only on the terms with these POS tags, thereby narrowing the feature space to those terms that are most relevant to the lexical chain. In addition, users often use different words to express a similar meaning [12]. Thus, synonyms were considered by adopting a dictionary-based similarity measure. We are interested in capturing the semantic similarity between terms, and thus a dictionary-based approach is preferred over a context-based or a clustering-based approach. Specifically, we compute a similarity value coef between two terms and incorporate this weight into the *tfidf* calculation, thereby allowing for better representation of the semantic relationships between messages. coef is calculated using existing semantic dictionary such as Xsimilarity. Xsimilarity [53] is an open source project that calculates similarity between any two phrases, sentences, or text documents in Chinese. The algorithm is based on both semantic relationship between words defined in HowNet, a Chinese word ontology, and an algorithm to calculate out-of-vocabulary (OOV) similarity. The project source code is available at https://github.com/iamxiatian/ xsimilarity. This tool has been used by a number of researchers [54,55] in studying semantic similarity in Chinese language. Linguistic

Coherence Analysis



Fig. 1. DiLTA framework.

parsing tools (such as LTP, available at https://github.com/HIT-SCIR/ ltp) were used to perform word segmentation, phrase identification, POS tagging, and dependency tree parsing to construct synonym categories.

Our RCS measure for computing scores for similarity between a pair of messages X, Y is shown in formulas (1)–(3).

$$Rsim(X, Y) = \frac{\sum_{k=1}^{l} \left((S_k^{(x)} * S_k^{(y)}) \right)}{\sqrt{\sum_{k=1}^{l} \left(S_k^{(x')} \right)^2 * \sum_{k=1}^{l} \left(S_k^{(y')} \right)^2}}$$
(1)

$$S_{k}^{(x')} = \sum_{i=1}^{n} (tf_{i}^{(x)} * idf_{i}^{(x)} * coef_{i}^{(x)})$$
⁽²⁾

$$S_{k}^{(y')} = \sum_{j=1}^{m} (tf_{j}^{(y)} * idf_{j}^{(y)} * coef_{j}^{(y)})$$
(3)

Rsim(X, Y) is the score for the similarity between messages *X* and *Y*. We first allocate terms with high similarity to the same category to obtain *k* categories. A term belongs only to the category with the highest similarity value retrieved from the dictionary. Each category *k* here is similar to a term in the original cosine similarity measure. The first term in the category has a *coef* value of 1. All other terms are then represented as tf^*idf^*coef .

After the grouping process, messages X and Y are represented by vectors x' and y', respectively, in our *Rsim* formula (1). The coordinates correspond to the weight of the categories. Formulas (2) and (3) represent the grouping process, where $coef_i^{(x)}$ is a value for the similarity between the i^{th} term in message X and the first term in the same category. Any two terms in the same category can be represented by x' and $x' * coef_i^{(x)}$. $S_k^{(x')}$ is the weight of the k^{th} dimension of vector x'. It is the sum of the term weights for the category, as shown in formula (2). $tf_i^{(x)}$ is the term frequency of the i^{th} term in message X. $idf_i^{(x)}$ is the inverse document frequency of the i^{th} term in message X.

We calculate the *Rsim* score for all message pairs in the discussion thread, and the resulting vector is called **msgSimList**. Our Rsim calculation is similar to the soft similarity score calculation proposed by Sidorov et al. [56].

3.1.2. Discussion segmentation algorithm

To disentangle a discussion thread into its component conversations, we developed a DSA guide by prior studies on discussion logic, specifically the work by Raghu, Ramesh, Chang, and Whinston [15] and Carbogim, Robertson, and Lee [30]. As we mentioned in Section 2, traditional methods of disentanglement have mostly relied on singlepass clustering methods [30,33]. To discover discussion logic, i.e., primitive and derivative messages, DSA adopted the thematization mechanism, which linearizes a conversation to sequentially uncover important themes within a conversation thread [46]. The thematization process uncovers conversation beginnings (i.e., primitive messages) and conversation endings (which we refer to as marginal messages). Note that during this phase, conversation relationships are not reconstructed. Only primitive messages are identified. Derivative messages are connected to their corresponding primitive messages during coherence analysis. Fig. 2 illustrates the DSA segmentation process based on thematization mechanism vs. traditional single-pass clustering methods.

As depicted in Fig. 2, the proposed algorithm has three steps: (1) select candidate primitive message, (2) find marginal message, and (3) validate primitive message. In the first step, the selection of candidate primitive messages is based on similarity between messages, **msgSim-List** vector described in the previous section. If the similarity between the current message and all other messages, the current message is considered as a candidate primitive message. The first message of the entire thread is always a primitive message. The second step identifies marginal message. If the mean similarity between the current

message and a prior given candidate primitive message is the smallest among all remaining messages, this current message is the marginal message. In step 3, we rescan the entire thread to validate the candidate primitive messages. The candidate primitive message is compared with all messages between itself and its immediate ancestor primitive message. If the mean variance of their similarity is greater than the mean variance of the similarity between the candidate primitive message and all messages, the candidate primitive message is a true primitive message. Or, the mean similarity between the candidate primitive message and the marginal message is greater than the mean average of the similarity between the candidate primitive message; at the same time, the mean average of the similarity between the candidate primitive message and all its prior messages is smaller than the maximum similarity between the primitive message and all messages. The detailed algorithm pseudocode is described in Fig. 3.

The rationale for the algorithm is that a message is associated with the primitive message of a new conversation if it has low similarity with most messages in the previous conversation and high similarity with most messages belonging to the new conversation. This intuition is operationalized using the mean and variance of the similarity between messages. The output of the proposed algorithm is the discussion logic feature: a mapping between messages and conversations within a discussion thread, where messages associated with the same conversation have the same discussion logic feature value.

3.2. Coherence analysis

The identification of coherence relationships is modeled as a binary classification problem, where each message pair in the discussion thread either constitutes a reply-to relationship or does not. The attributes used are a feature vector for each message pair; these feature vectors are inputted into a machine learning classifier. Details regarding the coherence analysis features and classification technique are as follows.

3.2.1. Coherence analysis features

DiLTA uniquely uses discussion logic features derived in the conversation disentanglement component in addition to system and linguistic features. In asynchronous communication modes, such as email and web forums, system features that are helpful in coherence analysis include quotations and message headers. However, such features do not exist in all discussions and especially not in synchronous discussions (e.g., chat systems). Asynchronous discussions can become as difficult to deal with as synchronous discussions when users choose not to quote previous messages explicitly. In these cases, only message ID can be captured automatically to determine message sequence and proximity. This information is used to compute the **distRange** feature: the distance between any two messages. While message proximity has been shown to provide some utility in prior coherence analysis studies, its effectiveness is diminished by the socio-technical gap [27] – in this case, through the imposition of a simple, sequential ordering.

As previously alluded to, linguistic features are important for understanding the contextual elements and lexical relationships between messages and, therefore, have important implications not only for conversation disentanglement but also for coherence analysis. We use two important linguistic features: message similarity and message sentiment. The message semantic similarity between two messages (simDegree) is computed using Xsimilarity tool described in the previous section. The message sentiment feature, sentiSeq, indicates whether the message pair contains subjective or objective content. Subjective messages are those that have greater sentiment polarity. Sentiment information is useful because users often express their opinion toward a prior message, and thus, it is likely to be a useful feature. We adopt a straightforward approach to determine whether a message is subjective or objective, where each term in a message is compared against items in the sentiment lexicon to compute a subjectivity score.



Fig. 2. Illustration of DSA process (left) and cluster-based segmentation (right).

Similar to English sentiment lexicons such as SentiWordNet [57], HowNet [58] is a Chinese sentiment lexicon that provides an effective mechanism for inferring sentiment polarity. The lexicon is available at http://www.keenage.com/html/e_index.html. Conversation disentanglement information is essential for reducing the likelihood of creating coherence relationships between messages from different conversations [26]. On the basis of output from the DSA algorithm, two discussion logic features are used for coherence analysis. The **convoSeq** feature is used to denote whether any one message in the pair is a primitive message of a conversation. The **logicPosition** feature provides information regarding the conversation affiliation of the two messages. Thus, a total of five features are used: the system feature **distRange**, the linguistic features **simDegree** and **sentiSeq**, and the logic features **convoSeq** and **logicPosition**.

3.2.2. Coherence analysis technique

Once the features between all message pairs have been extracted, coherence relationships are classified using the corpus-based machine learning approach TBL. A RM mechanism is used to handle the remaining message pairs. We collectively refer to the TBL plus RM method as TBL-RM.

3.2.3. Transformation-based learning

TBL has been successfully applied in spoken dialogue act classification [59]. It starts by learning the best sequence of suitable "transformation rules" from a training corpus (i.e., a set of message pair feature vectors). Consistent with prior work [7], the training corpus comprises all positive and negative (i.e., non-reply-to cases) reply-to cases encompassed in a collection of conversations. For a given message, negative cases are all previous messages with which they have no reply-to relationship. Overall, the number of negative cases considerably exceeds the number of positive cases. One advantage of TBL is that the generated rules are easy to understand and interpret. Each rule derived from TBL is composed of two parts: the combination of feature values used as the input condition for the rule and the associated replyto relationship tag (i.e., classification result for the rule). For example, "1 || 0 || || 0 || 2 || = > 1" is a rule derived by TBL. The five feature values are separated by "||." This rule means that if distRange = 1 and simDegree = 0 and sentiSeq is any value and subtopicSeq = 0 and logicPosition = 2, there is a reply-to relationship between the two messages.

used by TBL. This could be due to discourse ambiguity—a situation in which coherence cues are not explicit in discussion messages and are instead manifested in tacit knowledge that is difficult to express. A residual match algorithm is applied to such messages; it uses a recursive method that considers message pairs' similarity and discussion logic. A residual message is first matched to the most similar primitive message ahead of it. It is then linked to the most similar derivative message from that primitive message region, or the primitive message itself. Fig. 4 provides the pseudocode for the residual match algorithm. Once RM is complete, the reply-to relationships between messages in a discussion thread can be used to construct a conversation tree.

3.3. Discussion analysis tree

The visualization of a discussion thread structure can coherently show the dynamics of communicative interaction and collaboration and depict disentangled conversations [37,39]. It can allow users to better understand the intricacies and nuances of group discussion in a cognitively efficient manner [60]. For example, by the analysis of existing visualization tools such as ArguMed, Convince Me, and Reason!Able, van den Braak, Oostendorp, Prakken, and Vreeswijk [61] showed that these tools contributed to high-quality discussion and more coherent argumentation. Tree-based visualizations are particularly useful for understanding online discourse [44]. There is a need for tree-based representation to illustrate the usefulness in various content analysis scenarios.

The conversation disentanglement and coherence analysis components of DiLTA are combined to create a DATree for each group discussion. Such visualization can enhance people's content analysis capability, provide a holistic view of the discussion, and assist in the discovery of hidden meanings. Important subtopics are floated as top branches, the sequence of discussion is corrected, and disrupted adjacency turns are avoided. In Fig. 5, we present an example of a DATree constructed by DiLTA on the right in comparison with its original unstructured text on the left. In contrast to the unstructured view, a DA-Tree provides a clear visualization. In the tree, each branch represents a conversation, and the nodes under the branches represent messages in the conversations. It is apparent that this particular discussion encompasses multiple conversations, some of which have elaborate interactional coherence patterns.

4. Evaluation

3.2.4. Residual match

During classification, some messages may not fit any of the rules

We evaluate the effectiveness of various components of our DiLTA



Definition of Variables

msgMean: average similarity score between current message and all other messages preAllMean : average similarity score between current message and all previous messages preMean : average similarity score between current message and all previous messages after the previous subtopic residualMean: average similarity score between current message and all remaining messages marginalMean: average similarity score between current message and all remaining messages before the next candidate marginal messages msgVariance : variance of similarity score between current message and all other messages preVariance : variance of similarity score between current message and all previous message after the previous subtopic

1. Initialize the primitive message list of subtopics

If preAllMean = 0, Then set current message = new subtopic;

2. Identify candidate primitive message of new subtopic

- If (distance from the primitive message of previous Subtopic to this current message <2) and (preAllMean < msgMean), Then current message = candidate primitive message of new subtopic;
- If (distance from the primitive message of previous suBtopic to this current message >=2) and (preMean < msgMean), Then current message = candidate primitive message of new subtopic;

3. Find marginal message based on candidate subtopic

We define marginalMeanSet as collection of marginalMean;

Select the maximum similarity score fRom marginalMeanSet;

If its marginalMean is greater than residualMean, Then mark the corresponding message as marginal message;

4.Decide primitive message of new subtopic

If (preVariance > msgVariance) and (marginalMean > msgMean),

Then candidate message = primitive message of new subtopic;

ElseIf marginalMean > msgMean and the maximum value in candidate message's msgSimList is not previous message, Then candidate subtopic = primitive message of new subtopic;

Fig. 3. The proposed discussion segmentation algorithm.

Define X as the residual messages in chronological order. Define x_i as the *i*-th element in X. For each x_i:

- 1) Compute similarity value between x_i and the primitive messages of all conversations. Set the primitive message with the highest similarity value as z.
- 2) Define Y as all the messages that directly reply to z. Define y_i as the *j*-th element in Y.
- 3) If the Y is null, z is x_i 's antecedent message.
- Else, compute the similarity value between x_i and all elements in Y. Set the message in Y with the highest similarity value with x_i as z. And go to 2)

Fig. 4. Residual match algorithm used in TBL-RM.

Output: DATree Representation

Input: Unstructured Online Discussion Texts



Fig. 5. An example of original discussion text and DATree with internal structure.

as well as its overall effectiveness in two experiments. Experiments 1 and 2 assessed the effectiveness of the two computational components of the system in comparison with the existing methods. To assess the discussion representation improvements attributable to the system, we empirically evaluated the accuracy of social network centrality in Experiment 3.

4.1. Dataset

One of the challenges in evaluation is to find the "gold standard" of the internal structure of online discussions. Any post analysis, whether it is manual or automatic, can recover only the original structure as much as possible. To assess the true effectiveness of DiLTA, we recruited 80 participants to join group discussions using a GSS developed by Li, Zhang, and Cao [62]. All participants were graduate students of a business school. They were divided into 20 groups, and each group was composed of four discussants. We chose a general discussion topic: how to address an overproduction problem for a tea bag manufacturer. This thread topic had been used in previous GSS research [62]. The subjects were told to discuss solutions to the overproduction problem and to reach a business decision in 30 min.

The system is designed in a way that a user is required to specify the purpose of his or her message before it is sent to the system. The message can be either a new topic (primitive message) or a response to a previous message (derivative message), where the antecedent message must be specified. Such a reply-to relationship represents the logic behind their discussions. Although such a feature exists with most web forums, users are not forced to specify the purpose of their message, and the gold standard can be missing [4,7,27,63]. All subjects were given a brief training session to ensure that they knew how to use the system. We used these user-generated reply-to relationship tags as the "gold standard." We collected a total of 20 discussion threads. The longest thread contained 71 messages, and the shortest contained 22 messages. The average length was 42. This dataset with the gold standard was

used in the ensuing experiments.

4.2. Experiment 1: conversation disentanglement

In the first group of experiments, we evaluated the effectiveness of the two computational components. Experiment 1 focused on evaluating DSA, the conversation disentanglement component of DiLTA. The outputs of DSA phase are the primitive message and discussion affiliation *variables*, which are used as input for the coherence analysis.

4.2.1. Benchmarks

Most existing methods of conversation disentanglement are described in the computer science literature. They apply topic clustering to compute the similarity between messages. Some use additional linguistic features during clustering. Choi [64] performed segmentation with a bag-of-words and clustering based on the Euclidean distance between messages. Wang and Oard [50] also applied a bag-of-words and single-pass clustering. However, they incorporated information regarding the author, temporal, and conversational contexts (e.g., posting author information, time between messages, and direct address). Shen, Yang, Sun, and Chen [63] used bag-of-words coupled with additional linguistic features and messages weighted by time as input for a single-pass clustering algorithm. Adams and Martell [33] added hypernym information and a message distance penalty as well as direct address information. Elsner and Charniak [26] performed disentanglement using word repetition, discourse-based features, time windows, and direct address as input for a maximum entropy algorithm. For all comparison methods, the parameters were tuned retrospectively to yield the best possible results. Consistent with prior work, microlevel precision, recall, and F-measure were used as our performance measures [63]. Table 2 summarizes the features used in five benchmark systems.

To test the performance of our DSA component, we compared it against the abovementioned disentanglement methods. We used

Table 2

Summary of five benchmark systems with DSA.

Benchmarks	Features
Adams and Martell [33]	TF-IDF, Time-distance penalization, Hypernym augmentation, Nickname augmentation, and Thread extraction
Shen, Yang, Sun, and Chen [56]	Bag-of-terms, TF-IDF, Sentence type, and Personal Pronouns.
Elsner and Charniak [26]	Time, Speaker, Citation, Cue words, Question, Repetition, and Technical jargon
Choi [63]	Similarity
Wang and Oard [50]	Bag-of-terms, TF-IDF, Social and temporal contexts, and constructing expanded messages

Table 3

Results for conversation disentanglement experiment.

Technique	F-Measure	Precision	Recall
DSA	0.617*	0.596	0.671
Adams and Martell [33]	0.423	0.332	0.663
Shen, Yang, Sun and Chen [63]	0.422	0.421	0.452
Elsner and Charniak [26]	0.329	0.252	0.535
Choi [64]	0.284	0.344	0.266
Wang and Oard [50]	0.262	0.699	0.186

 \ast Significantly outperformed comparison methods, with all p values < 0.001.

benchmark source code when available [26]; otherwise, the same method was implemented. Consistent with all these studies, we used recall, precision, and F-measure as our performance metrics.

4.2.2. Results

Table 3 shows the experimental results. DSA outperformed all the five comparison methods in terms of F-measure by a wide margin. It also attained the highest recall rate, suggesting that it was able to identify more of the conversations appearing in the discussion threads than other methods. This was largely attributable to DSA's emphasis on identifying primitive messages (i.e., conversation beginnings). The analysis revealed that DSA correctly identified approximately 70% of the primitive messages, whereas comparison methods typically detected only 50% of the primitives. DSA was also more accurate at identifying marginal messages. Another factor was that the RCS measures included only terms with noun or verb POS to compute the similarity between messages, whereas the comparison methods did not incorporate POS information. While methods such as those suggested by Adams and Martell [33] and Elsner and Charniak [26] also yielded good recall rates, these methods had low precision. Conversely, the method developed by Wang and Oard [50] had good precision but low recall, which means that it was able to accurately associate messages with the conversations identified but failed to detect many of the conversations.

Paired *t*-tests were conducted to evaluate DSA against the comparison methods. The tests were performed on the macrolevel F-measures for 20 discussion threads (i.e., n = 20). DSA significantly outperformed all the five comparison methods (all p values < 0.001). The results underscore the effectiveness of the primitive message detection-oriented DSA method as a viable method for conversation disentanglement.

4.3. Experiment 2: coherence analysis

In Experiment 2, we evaluated the effectiveness of the coherence analysis component of DiLTA: TBL-RM.

4.3.1. Benchmarks

Because automatic machine learning algorithms are affected by two factors—feature and learning algorithms—we conducted three groups of experiments to analyze the utility of TBL-RM.

In the first group, we compared the *feature selection utility* of TBL-RM. Specifically, we compared system, linguistic, and discussion logic features in conjunction against the use of only a subset of the feature categories. We tested whether the use of all three feature categories in unison would outperform the use of a subset of the categories. During this group of comparisons, the TBL classifier with a residual match was used as the learning algorithm. Precision, recall, and F-measure were used as our performance measures. Note that in this classification problem, we were interested only in those message pairs that were classified as having a reply-to relationship. While the number of pairs that were classified as having no reply-to relationships was much higher, including these instances in the performance evaluation would have artificially inflated precision and recall rates for all experimental settings. Thus, our precision and recall metrics were based only on correctly classified reply-to relationships. It is important to note that the gold standard included one message as the antecedent (i.e., primitive) message for each new posting. In other words, the total number of reply-to relationships was fixed and was equal to one less than the total number of messages in the discussion thread. We applied this supposition in the TBL-RM classifier. Hence, the total number of identified reply-to relationships was always equal to the total number of reply-to relationships; this results in equivalent values for precision and recall.

In the second group, we evaluated the *learning algorithm utility* of TBL-RM by comparing with five other machine learning algorithms: J48, LibSVM, Logistic Regression, Naive Bayes, and Random Forest. We fed the algorithms with four settings of features: system + linguistic, system + discussion logic, and a full feature set with system + linguistic + discussion logic. Discussion logic feature is the output from the DSA component.

In the third group, we evaluated the proposed TBL-RM learning method against the existing techniques reviewed in Table 1: manual, linkage, heuristic, and simple classification (without discussion logic feature). All the benchmark methods have been used in previous studies. The heuristic-based benchmark is chosen because it is adopted in [27]. This method relied on three linguistic features derived from the message body: direct address, lexical similarity, and residual match. The direct address match identified coherence relationships on the basis of references to screen names. The Xsimilarity tool was used to compute lexical similarity between every two terms and then obtain the similarity score between messages using VSM. A naive linkage-based residual match rule was applied to the remaining messages [27,43]. The classification-based method used linguistic and system features [7]. It represented the coherence analysis task as a binary classification problem (i.e., to determine whether two messages constitute a reply-to relationship). We extracted four types of features from the message pairs: "time_gap" and "dist" are the intervals of time and distance between message pairs, respectively; "repeatNoun" is the number of repeated nouns between message pairs; and "viewer_timeGap" is the number of message pairs from the same author who had a time interval of less than 5 s.

Coherence analysis is a challenging problem. Group discussion participants often struggle to manually identify coherence relationships [12]. To provide an impression of task difficulty, manually annotated coherence analysis results are often included as an upper bound [50]. Five independent annotators with experience in social media analytics and discourse analysis were asked to annotate the 20 discussion threads with coherence relationships. The macro F-measure, precision, and recall (i.e., average of the five annotators across the 20 discussion threads) were included in the experimental results. As with TBL-RM, the heuristic, linkage, and manual methods also assumed that each message had only a single parent message [27]; hence, their precision and recall values were equivalent. In contrast, the comparison classification-based method did not make this assumption [7].

The classification is a skewed dataset where the number of negative examples (non-reply-to relationships) is much higher than that of positive examples (reply-to relationships). We are interested only in positive precision, or true-positive rate. Thus, our precision and recall metrics are based on correctly classified reply-to relationships and are measured as follows:

$$precision = \frac{Number of Correctly Identified Reply - to Relationships}{Total Number of System - identified reply - to relationships}$$
$$recall = \frac{Number of Correctly Identified Reply - to Relationships}{Total Number of Expert - identified Reply - to Relationships}$$

During the coherence analysis step, each message is assigned to another message to form a pair that comprises a reply-to relationship, except for the first message. Experts also assign each message to another message to form a pair that comprises a reply-to relationship, except for the first message. Thus, the two denominator values, the total number of system-identified reply-to relationships equals the total number of expert-identified reply-to relationships, which equals the total number of message minus one. Thus, precision and recall measures in this experiment have equal values.

4.3.2. Results

The experimental results are shown in Tables 4–6. When comparing feature categories, both linguistic and discussion logic features improved performance over the use of only system features. For both, paired *t*-test p values were significant (< 0.001, n = 20). Using all three feature categories also outperformed the use of system + linguistic and system + discussion logic features (both p values < 0.001). Overall, the results lend credence to the notion that linguistic and discussion logic features are important for bridging the coherence gap resulting from the dichotomy between system features supported by existing communication technologies and discourse practices used in group discussion.

We conducted ablation test to compare five other machine learning techniques using different feature sets, as presented in Table 5. Focusing on machine learning algorithms alone, TBL-RM performed consistently better than all other learning algorithms, achieving 0.481 in F-measure. Among the benchmark algorithms, the best performance was achieved with the Naive Bayes algorithm when all three types of feature are used: system + linguistic + discussion logic, which achieved an F-measure of 0.315. Looking at feature set differences, we found that when paired with system feature, discussion feature alone or linguistic feature alone did not improve the performance much over system feature, it sometimes even decreased the performance when inappropriate machine learning algorithms were used (J48, SVM, and Random Forest).

Table 4

Results for coherence analysis experiment: comparing feature sets.

Comparison	F-Measure	Precision	Recall
TBL-RM Features	0 168	0 168	0 168
System + Linguistic	0.261+	0.261	0.261
System + Discussion Logic System + Linguistic + Discussion Logic	0.243 0.481*	0.243 0.481	0.243 0.481

 * Significantly outperformed all other feature category combinations, with all p values < 0.001.

 $^+$ Significantly outperformed the use of only system features, with both p values < 0.001.

The Naive Bayes and TBM-RM algorithms both achieved reasonable performance when all three feature sets were considered. We observe that the improvement from linguistic feature is even more obvious than discussion logic feature (55.4% vs. 44.6% improvement, respectively, in TBL-RM). We believe that although the proposed discussion logic feature is an important feature in improving restructuring interrupted online discussions, one cannot ignore linguistic features. The best performance is achieved when all three feature sets are considered (186.3% improvement over system feature alone with TBL-RM).

When all three existing techniques were compared, TBL-RM outperformed the comparison heuristic, linkage, and classification methods by a wide margin. Paired *t*-test results were significant (all p values < 0.001). The poor performance of the linkage method was attributable to disrupted turn adjacency; over 80% of the time, the adjacent messages in the discussion thread had no reply-to relationship with one another. Consequently, naive linkage yielded poor results. The comparison classification method attained good precision but poor recall. This was attributable to limitations in the coverage provided by the classifier's rules, which were mostly based on system features related to message proximity and time gaps. The limited use of linguistic features and lack of discussion logic attributes contributed to the classification method's low recall. While the heuristic method performed better, its performance was adversely affected by the use of discourse patternrelated assumptions that did not hold in this context. The method's overreliance on direct address and coreference-based interaction cues was problematic, because these items were seldom used in the test bed group discussions. As expected, manual identification outperformed all the automated methods. However, the performance gain over TBL-RM was not significant (p value = 0.519). Given the infeasibility of manual identification over large volumes of data, the relatively equitable performance of TBL-RM suggests that it may constitute a viable automated alternative. Overall, the results demonstrate the efficacy of the proposed coherence analysis method, which combines system, linguistic, and discussion logic features with a classification method and RM.

4.4. Experiment 3: improving accuracy of social network centrality measures

After showing the effectiveness of DiLTA, our second experiment focused on the usefulness of DiLTA. While we believe that such automatic online discussion analysis can be useful in many ways, one of them is in social network analysis. According to a recent Gartner report, the organizational use of social network analysis is on the rise. From an organizational discourse perspective, important applications of social network analysis include understanding power dynamics and identifying experts and influencers [65]. Social networks derived from conversations can illuminate participant roles using measures such as degree centrality, betweenness, and closeness [27]. However, computation of these measures requires precise values for in-degree: the number of message responses to a participant [45,46]. Otherwise, participant roles can be distorted—either exaggerated for some or understated for others [27].

Inaccurate discussion structure can distort representations of participants' roles in online group discussions. The differences between actual and projected social network centrality measures can shed light on the level of distortion [27,46]. Three commonly used measures are degree centrality, closeness centrality, and betweenness centrality. Degree centrality is the total number of outlinks (sent messages) and inlinks (received/reply-to messages) associated with a discussant; it is a measure of a discussant's level of participation and interaction within a discussion thread [46]. Closeness centrality is a measure of the level of interaction between participants within a group, with greater interaction between discussants indicating greater closeness. Betweenness centrality is an important measure of how critical an individual is for the flow of communication between other discussants in a conversation [27]. For a given discussant, it is computed as the proportion of the

Table 5

Results for coherence analysis experiment: comparing learning algorithms.

Learning Algorithm	Measures	Features			
_		System	System + Linguistic	System + Discussion Logic	System + Linguistic + Discussion Logic
J48	Precision	0.163	0.163	0.163	0.163
	Recall	0.163	0.163	0.163	0.163
	F-measure	0.163	0.163	0.163	0.163
LibSVM	Precision	0.163	0.163	0.163	0.143
	Recall	0.163	0.163	0.163	0.119
	F-measure	0.163	0.163	0.163	0.130
Logistics	Precision	0.163	0.099	0.163	0.231
	Recall	0.163	0.105	0.163	0.214
	F-measure	0.163	0.102	0.163	0.222
Naive Bayes	Precision	0.163	0.271	0.123	0.251
	Recall	0.163	0.350	0.138	0.422
	F-measure	0.163	0.306	0.130	0.315
Random Forest	Precision	0.163	0.069	0.138	0.195
	Recall	0.163	0.071	0.114	0.172
	F-measure	0.163	0.070	0.125	0.183
TBL-RM	Precision	0.168	0.261	0.243	0.481
	Recall	0.168	0.261	0.243	0.481
	F-measure	0.168	0.261	0.243	0.481

Table 8

Discussant/Method

Discussant A

Discussant B

Discussant C

Discussant D

Table 6

Results for coherence analysis experiment: comparison with existing techniques.

Technique	F-Measure	Precision	Recall
Heuristic	0.283	0.283	0.283
Linkage	0.168	0.168	0.168
Simple Classification	0.061	0.276	0.035
Manual Identification	0.521 ⁺⁺	0.521	0.521
TBL-RM	0.481 [*]	0.481	0.481

 $^+$ Significantly outperformed the use of only system features, with both p values < 0.001.

 \ast Significantly outperformed all other feature category combinations, with all p values < 0.001.

⁺⁺ Did not significantly outperform TBL-RM; p value = 0.519.

shortest paths between discussants in the network that includes this discussant. We examined the mean absolute percentage error in degree, closeness, and betweenness centrality for the DiLTA coherence analysis module and the comparison heuristic, linkage, and classification methods. The values were computed across all 80 participants in the 20 discussion test beds.

Table 7 shows the experimental results. The conversation disentanglement method in DiLTA had the smallest mean absolute percentage errors for all three social network centrality measures, with error percentages of only 8.5% and 5.2% for degree and closeness, respectively. The DSA error rates were nearly two to three times better than those of the comparison methods. The differences were statistically significant (with all p values < 0.001). Among the comparison methods, heuristic and linkage methods had error rates of 13%–35%, respectively, while the classification method had error values between 47% and 62%. Table 8 presents the gold standard social network and degree centrality values (i.e., sum of in/out degree for each

Table 7

Mean absolute percentage error for social network centrality measures.

Method	Degree Centrality	Closeness Centrality	Betweenness Centrality
DiLTA	0.085 [°]	0.052 [*]	0.208 [°]
Heuristic	0.187	0.142	0.356
Linkage	0.179	0.136	0.339
Classification	0.619	0.627	0.475

 \ast Significantly outperformed comparison methods, with all p values < 0.001.

linkage methods, for one of the discussions in the test bed. The DSA
degree values for the four discussants were almost identical to the gold
standard. In contrast, the heuristic method exaggerated the degree

Degree centrality values for example group discussion.

48

26

9

27

Gold Standard

standard. In contrast, the heuristic method exaggerated the degree centrality of discussant D. Similarly, because of its bias toward discussants with greater posting frequency within discussions [27], the linkage method inflated the value for discussant A. In the case of the latter method, this was attributable to its lower coherence analysis performance. Though not depicted in the figure, this also caused these two discussants' closeness and betweenness values to be overly pronounced in their respective methods' networks, thereby overemphasizing their centrality and perceived importance (see Fig. 6).

participant), along with those generated by the TBL-RM, heuristic, and

DiLTA-DSA

48

28

8

26

Heuristic

39

20

10

41

Linkage

67

16

10

17

5. Case study on healthcare discussions

Our experimental results showed the utility of each individual component of DiLTA, compared with existing benchmark methods, in online discussion generated by GSS. In this section, we conducted a case study on healthcare discussion to illustrate the usefulness of DiLTA.

5.1. Dataset

The healthcare dataset is crawled from two popular online forums: http://bbs.tnbz.com/ and http://www.sunofus.com/. The first one is an online forum for patients with diabetes and the second one is for patients with depression. We randomly picked nine threads each with around 100–200 messages. Table 9 provides an overview of the datasets, including the number of messages, the number of participants, and the number of quotations. The total dataset included over 1000 messages associated with 476 participants.

Unlike the chat room dataset used in Section 4, Web forums lack a gold standard of reply-to structure. Consistent with prior studies [27], all messages in the test bed were labeled by two independent human annotators with experience in discourse analysis. The annotators



Fig. 6. Social network for example group discussion.

labeled each thread message with respect to primitive status and replyto relationships. The annotators first underwent several rounds of training on messages from web forums that were not part of the test bed. Then, each of them independently annotated the messages in the test bed, thus attaining a Cohen's Kappa value of over 75% for interannotator agreement. All disagreements were solved by annotators' discussion. The annotated dataset was used as the gold standard.

5.2. Performance validation

We respectively validated the performance of the conversation disentanglement component and coherence analysis component of DiLTA on a healthcare domain test bed in comparison with three benchmark methods: Linkage method, Classification method, and Heuristic method (shown in Table 10). The linkage method performed the worst because it used only quotations and assumed that all residual messages (i.e., ones not containing any quotations) were replying to the previous message. Classification method and heuristic method performed similar in terms of final coherence analysis result. However, the classification method is able to identify more corrected conversation appearing in the discussion thread, thereby showing a higher recall rate. However, the poor precision shows that it is not able to associate messages with the correct subtopic, thus losing the logic in discussion. Although these test beds come from different channels and domains, our proposed framework performances are consistent with those of Experiment 1a and Experiment 1b in terms of F-measure, precision, and recall on each thread of the test bed. Because the total number of identified reply-to relationships was always equal to the total number of reply-to relationships, it resulted in equivalent values for precision, recall, and F-measure. We observed that system features such as quotation and timestamp are more effective on web forum datasets than on

а	chat	room	dataset
а	unat	100111	ualasel.

5.3. DATree comparison

To further demonstrate the usefulness of DATree, Fig. 7 compares the gold standard tree representation (top-left chart) with DATree generated by DiLTA (top-middle chart) and three benchmark methods (top-right and bottom charts), using one of the discussions in our case study. From the shape of the tree, it is obvious that DATree generated by DiLTA most closely resembles the gold standard in terms of link structure between nodes. The gold standard and DiLTA grouped the messages in a similar way. It is clear that there are four major issues in discussion (with primitive message numbers 0, 25, 40, and 76). They contain more layers and sketch the discussion process in detail. One can deem these issues as the main discussion topics. The linkage method identifies only one issue. It distorts the relationships among issues and provides an incorrect holistic view of the discussion. Conversely, the classification method generates many single leaf nodes, thus resulting in a wide tree with many branches. However, it fails to associate the message nodes to the correct branch (or primitive message). The tree structure explains the high recall and low precision of the classification method. The heuristic method generates less than desired branches in comparison to the gold standard. It is clear that DiLTA-based DATree best resembles the gold standard tree, showing the superior performance by capturing discussion logic in our framework.

In addition to the overall conversation structure, inaccurate coherence relationships can distort representations of participants' roles in online group discussions. In addition to degree centrality, closeness centrality, and betweenness centrality, another widely used measure is PageRank score [66], which aims to allow the propagation of influence along the network of nodes instead of simply counting the number of

Table	9
-------	---

Overview of neartificare data

	Number of Messages	Number of Participants	Number of Quotations	http
Health1	123	57	66	http://bbs.tnbz.com/thread-346598-1-1.html
Health2	204	99	101	http://bbs.tnbz.com/thread-310949-1-1.html
Health3	56	21	39	http://bbs.tnbz.com/thread-743810-1-1.html
Health4	100	49	63	http://bbs.tnbz.com/thread-349243-1-1.html
Health5	136	77	60	http://bbs.tnbz.com/thread-327255-1-1.html
Health6	129	91	61	http://bbs.tnbz.com/thread-327255-1-1.html
Health7	105	48	52	http://bbs.tnbz.com/thread-599203-1-1.html
Health8	99	17	48	http://www.sunofus.com/bbs/thread-2105937-1-1.html
Health9	94	17	58	http://www.sunofus.com/bbs/forum.php?mod = viewthread&tid = 94139&page = 1
Total	1046	476		

Table 10

DiLTA performance validation on healthcare domain dataset.

Component of DiLTA	DILTA			Linkage Method				
	F-measure	Precision	Recall	F-measure	Precision	Recall		
Conversation Disentanglement Coherence Analysis	0.644 0.792	0.594 0.792	0.728 0.792	0.189 0.620	1 0.620	0.111 0.620		
Component of DiLTA	Classification			Heuristic				
	F-measure	Precision	Recall	F-measure	Precision	Recall		
Conversation Disentanglement Coherence Analysis	0.226 0.590	0.169 0.590	0.429 0.590	0.238 0.617	0.363 0.617	0.279 0.617		



Fig. 7. DATree presentation for example discussion thread from healthcare domain forum.

Table 11

Wilcoxon signed-rank test for social network PageRank score measure.

	DiLTA vs. Gold Standard	Linkage Method vs. Gold Standard	Classification vs. Gold Standard	Heuristic vs. Gold Standard
Z	-0.443 ^a	- 8.000 ^a	-1.328 ^a	-2.118 ^a
Asymp. Sig. (2-tailed)	0.658 ⁺⁺	0.000 [*]	0.184^+	0.034 [°]

^a Based on negative rank.

 * Significantly different from the gold standard; p value < 0.05.

⁺ Significantly different from the gold standard; p value = 0.184.

⁺⁺ Significantly different from the gold standard; p value = 0.658.



Fig. 8. Social network of discussion thread from healthcare domain forum.

other nodes pointing at the current node. We used the Wilcoxon signedrank test on the PageRank score measure to compare the DiLTA and the linkage methods.

Table 11 shows the Wilcoxon signed-rank test results. The differences between DiLTA and the gold standard were not significant (with p value = 0.658). Surprisingly, the differences between the classification method and the gold standard were not significant either (with p value = 0.184). However, the other two benchmark methods, by contrast, were significantly different from the gold standard on the health domain test bed. Consistent with Section 4, DiLTA can improve the accuracy of the PageRank score in measuring influence in social network.

5.4. Social network representation

Fig. 8 shows the gold standard social network in comparison with the results generated by DiLTA and the three benchmark methods. Each node represents a user. The node sizes are proportional to the PageRank score, and the link thickness represents the reply-to degree weight. It is apparent that DiLTA most closely resembles the gold standard in terms of node sizes and links between nodes. Node 1 has the highest PageRank score, followed by nodes 26, 3, and 34. The linkage method, conversely, tends to exaggerate the PageRank values of many nodes, such as nodes 8, 9, 78, and 79. This is consistent with prior studies [20,27,45,46], which have also observed that linkage methods inflate PageRank scores. The classification method incorrectly picked nodes 3, 9, 54, 54, and 83 as more important nodes, while the PageRank scores for nodes 26 and 34 were discounted. Similarly, the heuristic method also amplified the importance of nodes 3, 9, 51, 54, 78, and 83, while diminished the importance of nodes 26 and 34.

The results of the experiments show that DiLTA can reasonably recover the true online discussion structure, especially when direct referencing and quoting features are not available. The performance of DiLTA is almost as good as that of human experts. Thus, it has great potential to facilitate social science analysis. The output of DiLTA can be represented by a DATree to lay out the structure of a discussion text and further support online discourse analysis.

6. Conclusions

This work proposes a DiLTA framework to support the analytics of online discussions. We attempt to connect social science and computer science studies of online discussions by incorporating the spirit of argumentation models into automatic discussion analysis. The goal of our approach is to recover the internal discussion logic to facilitate a more advanced and pragmatic level of analysis. We recruited students to participate in online discussions in which their true conversation intention and coherence could be reported. By using this dataset, we evaluated our approach in a series of experiments and demonstrated the utility of each individual component in comparison to the existing methods. Furthermore, the performance of the framework is validated by using an online healthcare discussion dataset. The two datasets represent different styles of discussions. Social network analysis experiments were performed to show that key discussants could be better identified with DiLTA framework.

Our contributions are twofold. First, this research contributes to the theory of argumentation models. The design and development of DiLTA rely on extensions of Toulmin's model as a strong theoretical foundation. The findings provide evidence that argumentation models can be generalized to various datasets. We further show that argumentation theories such as Toulmin's model indeed provide important cues in discussion and lead to better understanding of conversation structure. This research also provides a new possibility for theory development. The use of automatic analytical system such as DiLTA can potentially help social scientists to develop and validate new argumentation models.

Second, this research contributes to technology development of online discussion analysis. Compared to traditional automatic discussion analysis, which focuses on system and linguistic features only, this new method of textual analysis looks at intermessage semantics and relationships. Several other innovative techniques are also introduced. To detect the discussion logic features, a novel DSA algorithm is introduced. The use of discussion logic features in TBL-RM is promising for unstructured user-generated discussion texts in which system features and linguistic features are less effective. The automatic tree-based representation DATree demonstrates the advantage of turning unstructured discussion text into a structured tree format, which greatly helps decision makers to obtain valuable information from massive user-generated discussion texts in an effective fashion.

The results of this research have further implications. From a user's perspective, the holistic hierarchical structure of discussions can help decision analysis practitioners to quickly understand and grasp the discussion content [67]. The DATree provides an appropriate level of description for online group discussion. The methodology of creating DATree visualization applies to other types of social media, such as online user reviews, Twitter data, email, newsgroups, and forums.

From a group decision-making perspective, the goal of online group discussion is to find a scientific and rational solution for decision problems [68,69]. Viewing group discussions as a DATree is potentially useful for supporting group rational decision-making and helps a discussion supervisor to make appropriate interventions [65]. The interpretation of a DATree offers clues to make interventions. Furthermore, the relationship between these issues can help to sketch the procedure of group reasoning and provide information regarding other decision problems [19]. It has the potential for an in-depth analysis of success or failure factors in an organization.

From an enterprise and organizational perspective, we can apply the DiLTA technique to enterprise and organizational board discussions for strategic and business decision problems. The output—the DATree—conveying issue networking and individual behavior, provides a relationship among these issues in different discussions and human contributions among issues. It can also evaluate the individual's level of expertise and support to identify trustworthy experts. Finally, we can determine the discussant's performance in a discussion according to the individual's behavior.

7. Limitations and future directions

Similar to all other studies on coherence analysis, this research has some limitations because of resource constraints. First, although our framework is generic, the current experiment was conducted on Chinese group discussions. It will be interesting to apply the framework in different languages and compare performance. Second, to generate the gold standard for evaluation purposes, we relied on an experimental dataset, where the true discussion structure can be recovered. Such a gold standard cannot be achieved from random social media data. However, it will be interesting to see the performance of the model on other forms of social media data, such as Twitter and Facebook, as the next step. Third, our discussion logic feature is derived based on two simplified categories of messages: primitive and derivative messages. As the first step in adopting Toulmin's model, we believe that it shows the power of using this feature. In the future, we plan to incorporate a more complex version of Toulmin's model, which will include four argument elements: claim, data, backing, and rebuttal. It is not clear whether using four argument elements will yield significant improvements over the two argument categories. It is also worth noting that our Rsim calculation can be optimized to reduce computational complexity as illustrated by Sidorov et al. [56]. In our experiment, because Rsim was used as one of the steps to perform subtopic segmentation, we did not expand the paper on details of optimization. Furthermore, although this research qualitatively illustrates the usefulness of structured discussion text with tree-based representations, it will be interesting to further study different forms of tree-based representation in the future.

References

- C.-L. Hsu, Y.-C. Liao, Exploring the linkages between perceived information accessibility and microblog stickiness: the moderating role of a sense of community, Inf. Manag. 51 (2014) 833–844.
- [2] S.C. Herring, Computer-mediated communication on the internet, Annu. Rev. Inf. Sci. Technol. 36 (2002) 109–168.
- [3] S. Hrastinski, The potential of synchronous communication to enhance participation in online discussions: a case study of two e-learning courses, Inf. Manag. 45 (2008) 499–506.
- [4] L. Wang, M. Lui, S.N. Kim, J. Nivre, T. Baldwin, Association for Computational LinguisticsPredicting Thread Discourse Structure Over Technical Web Forums, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing2011, Predicting Thread Discourse Structure Over Technical Web Forums, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing (2011) 13–25.
- [5] J. Bughin, M. Chui, The rise of the networked enterprise: web 2.0 finds its payday, McKinsey Q. 4 (2010) 3–8.
- [6] M. Chau, J. Xu, Business intelligence in blogs: understanding consumer interactions and communities, Mis Q. 36 (2012) 1189–1216.
- [7] J. Kim, J.-H. Kang, Towards identifying unresolved discussions in student online forums, Appl. Intell. 40 (2014) 601–612.
- [8] R. Rienks, D. Heylen, vd.E. Weijden, Argument diagramming of meeting conversations, Workshop at the 7th International Conference on Multimodal Interfaces (ICMI) (2005) 85–92.
- [9] P. Mutton, Inferring and visualizing social networks on internet relay chat, Proceedings of the Eighth International Conference on Information Visualisation (2004) 35–43.
- [10] T. Holmer, Discourse structure analysis of chat communication, Language@ Internet 5 (2008) 1–17.
- [11] D. Vronay, M. Smith, S. Drucker, Alternative interfaces for chat, Proceedings of the 12th Annual ACM Symposium on User Interface Software and Technology (1999) 19–26.
- [12] C.M. Nash, Cohesion and reference in English chatroom discourse, Proceedings of the 38th Hawaii International Conference on System Sciences (2005) 1–10.
- [13] R. Savolainen, The structure of argument patterns on a social Q&A site, J. Am. Soc. Inf. Sci. Technol. 63 (2012) 2536–2548.
- [14] C. Chesñevar, S. Modgil, I. Rahwan, C. Reed, G. Simari, M. South, G. Vreeswijk, S. Willmott, Towards an argument interchange format, Knowl. Eng. Rev. 21 (2006) 293–316.
- [15] T. Raghu, R. Ramesh, A.-M. Chang, A.B. Whinston, Collaborative decision making: a connectionist paradigm for dialectical support, Inf. Syst. Res. 12 (2001) 363–383.
- [16] D. Kim, I. Benbasat, The effects of trust-assuring arguments on consumer trust in Internet stores: application of Toulmin's model of argumentation, Inf. Syst. Res. 17 (2006) 286–300.
- [17] L. Zhou, J. Wu, D. Zhang, Discourse cues to deception in the case of multiple receivers, Inf. Manag. 51 (2014) 726–737.
- [18] J. Osborne, Arguing to learn in science: the role of collaborative, critical discourse, Science 328 (2010) 463–466.
- [19] T. Raghu, R. Ramesh, A.B. Whinston, Addressing the homeland security problem: a collaborative decision-making framework, J. Am. Soc. Inf. Sci. Technol. 56 (2005) 310–324.
- [20] A. Abbasi, H. Chen, CyberGate: a design framework and system for text analysis of computer-mediated communication, MIS Q. 32 (2008) 811–837.
- [21] K. Fisher, S. Counts, A. Kittur, Distributed sensemaking: improving sensemaking by leveraging the efforts of previous users, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (2012) 247–256.
- [22] M. Khalifa, V. Liu, Semantic network discussion representation: applicability and some potential benefits, Professional Communication, IEEE Trans. 49 (2006) 69–81.
- [23] K. Zechner, Automatic summarization of open-domain multiparty dialogues in diverse genres, Comput. Linguist. 28 (2002) 447–485.
- [24] S.R. Hiltz, M. Turoff, Structuring computer-mediated communication systems to avoid information overload, Commun. ACM 28 (1985) 680–689.
- [25] S. Farnham, H.R. Chesley, D.E. McGhee, R. Kawal, J. Landau, Structured online interactions: improving the decision-making of small discussion groups, Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work (2000) 299–308.
- [26] M. Elsner, E. Charniak, Disentangling chat, Comput. Linguist. 36 (2010) 389–409.
 [27] T. Fu, A. Abbasi, H. Chen, A hybrid approach to web forum interactional coherence
- analysis, J. Am. Soc. Inf. Sci. Technol. 59 (2008) 1195–1209.
 S. Deng, P. Zhang, Y. Zhou, Turning unstructured and incoherent group discussion
- [28] S. Deng, P. Zhang, Y. Zhou, Turning unstructured and incoherent group discussion into DATree: a TBL coherence analysis approach, the Thirty Second International Conference on Information Systems, (2011) Paper 22.
- [29] X. Yang, Y. Li, C.-H. Tan, H.-H. Teo, Students' participation intention in an online discussion forum: Why is computer-mediated interaction attractive? Inf. Manag. 44

S. Deng et al.

- [30] D.V. Carbogim, D. Robertson, J. Lee, Argument-based applications to knowledge engineering, Knowl. Eng. Rev. 15 (2000) 119–149.
- [31] M. Turoff, S.R. Hiltz, M. Bieber, J. Fjermestad, A. Rana, Collaborative discourse structures in computer mediated group communications, J. Comput. Commun. 4 (1999) available online http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101. 1999.tb00104.x/full.
- [32] S.E. McDaniel, G.M. Olson, J.C. Magee, Identifying and analyzing multiple threads in computer-mediated and face-to-face conversations, Proceedings of the 1996 ACM Conference on Computer Supported Cooperative Work (1996) 39–47.
- [33] P.H. Adams, C.H. Martell, Topic detection and extraction in chat, IEEE International Conference on Semantic Computing (2008) 581–588.
- [34] D. Te'eni, Review: a cognitive-affective model of organizational communication for designing IT, Mis Q. 25 (2001) 251–312.
- [35] E. Auramäki, R. Hirschheim, K. Lyytinen, Modelling offices through discourse analysis: the SAMPO approach, Comput. J. 35 (1992) 342–352.
- [36] F.M. Khan, T.A. Fisher, L. Shuler, T. Wu, W.M. Pottenger, Mining chat-room conversations for social and semantic interactions, in: technical Report LU-CSE-02-011, Computer Science and Engineering, Lehigh University (2002).
- [37] M. Smith, Tools for navigating large social cyberspaces, Commun. ACM 45 (2002) 51–55.
- [38] K.-P. Yee, Zest: discussion mapping for mailing lists, Proceeding on the ACM 2002 Conference on Computer Supported Cooperative Work (CSCW 2002), (2002).
- [39] J. Donath, A semantic approach to visualizing online conversations, Commun. ACM 45 (2002) 45–49.
- [40] W.M. Soon, H.T. Ng, D.C.Y. Lim, A machine learning approach to coreference resolution of noun phrases, Comput. Linguist. 27 (2001) 521–544.
- [41] W. Sack, Conversation map: an interface for very large-scale conversations, J. Manag. Inf. Syst. 17 (2000) 73–92.
- [42] P.S. Newman, Exploring discussion lists: steps and directions, Proceedings of the 2nd ACM/IEEE-CS Joint Conference on Digital Libraries (2002) 126–134.
- [43] D.E. Comer, L.L. Peterson, Conversation-based mail, Acm Trans. Comput. Syst. 4 (1986) 299–319.
- [44] S. Herring, Interactional coherence in CMC, J. Comput. Commun. 4 (1999) available online http://jcmc.indiana.edu/vol4/issue4/herring.html.
- [45] T. Anwar, M. Abulaish, Mining an enriched social graph to model cross-thread community interactions and interests, in: the 3rd international workshop on modeling social media, ACM (2012) 35–38.
- [46] E. Aumayr, J. Chan, C. Hayes, Reconstruction of threaded conversations in online discussion forums, the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM-11), (2011), pp. 26–33.
- [47] M.S. Ackerman, The intellectual challenge of CSCW: the gap between social requirements and technical feasibility, Hum. Interact. 15 (2000) 179–203.
- [48] F. Barcellini, F. Détienne, J.-M. Burkhardt, W. Sack, A study of on- line discussions in an open-source software community: reconstructing thematic coherence and argumentation from quotation practices, Springer, New York, the Communities and Technologies Conference (C&T'05), 2005, pp. 301–320.
- [49] G. Kasper, Can pragmatic competence be taught, NetWork 6 (1997) 105–119.
- [50] L. Wang, D.W. Oard, Context-based message expansion for disentanglement of interleaved text conversations, Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics (2009) 200–208.
- [51] G. Murray, S. Renals, Term-weighting for summarization of multi-party spoken dialogues, Machine Learning for Multimodal Interaction, Springer, 2007, pp. 156–167.
- [52] A. Huang, Similarity measures for text document clustering, Proceedings of the Sixth New Zealand Computer Science Research Student Conference (NZCSRSC2008) (2008) 49–56.
- [53] X. Tian, Study on chinese words semantic similarity computation (in chinese), Comput. Eng. 33 (2007) 191–194.
- [54] Y.L. Zhang Yiwen, Zhang Yifei, Li Qing, Cheng Jiaxing, Friends recommended method based on common users and similar labels (in chinese), J. Comput. Appl. 33 (2013) 2273–2275.
- [55] H.J.-Z. Li Feng, Zhou-J.un Li, Wei-M.ing Yang, Study on domain-corpus driven calculation method of sentence relevance (in chinese), Comput. Sci. 43 (2016) 188–192.
- [56] G. Sidorov, A. Gelbukh, H. Gómez-Adorno, D. Pinto, Soft similarity and soft cosine measure: similarity of features in vector space model, Comput. Y Sist. 18 (2014) 491–504.
- [57] A. Esuli, F. Sebastiani, SENTIWORDNET: a high-coverage lexical resource for opinion mining, Proceedings of the 5th Conference on Language Resources and Evaluation, (2007), pp. 417–422.
- [58] Z. Dong, Q. Dong, C. Hao, Hownet and its computation of meaning, Proceedings of the 23rd International Conference on Computational Linguistics: Demonstrations (2010) 53–56.
- [59] E. Brill, Transformation-based error-driven learning and natural language processing: a case study in part-of-speech tagging, Comput. Linguist. 21 (1995) 543–565.
- [60] R.M. Fuller, U. Murthy, B.A. Schafer, The effects of data model representation method on task performance, Inf. Manag. 47 (2010) 208–218.
- [61] S.W. van den Braak, H.V. Oostendorp, H. Prakken, G.A. Vreeswijk, A critical review of argument visualization tools: Do users become better reasoners? Workshop Notes

of the ECAI-06 Workshop on Computational Models of Natural Argument, (2008), pp. 67–75.

- [62] J. Li, P. Zhang, J. Cao, External concept support for group support systems through Web mining, J. Am. Soc. Inf. Sci. Technol. 60 (2009) 1057–1070.
- [63] D. Shen, Q. Yang, J.-T. Sun, Z. Chen, Thread detection in dynamic text message streams, Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (2006) 35–42.
- [64] F.Y. Choi, Advances in domain independent linear text segmentation, the 1st North American Chapter of the Association for Computational Linguistics Conference (2000) 26–33.
- [65] L. Heracleous, R.J. Marshak, Conceptualizing organizational discourse as situated symbolic action, Hum. Relat. 57 (2004) 1285–1312.
- [66] H. Kwak, C. Lee, H. Park, S. Moon, What is twitter, a social network or a news media? Proceedings of the 19th International Conference on World Wide Web (2010) 591–600.
- [67] M. Galley, Automatic summarization of conversational multi-party speech, Proceedings of the Twenty-First National Conference on Artificial Intelligence (AAAI-06), (2006), pp. 1914–1915.
- [68] J. Jay Nunamaker, Future research in group support systems: needs, some questions and possible directions, Int. J. Hum. Stud. 47 (1997) 357–385.
- [69] A.D. Moor, M. Aakhus, Argumentation support: from technologies to tools, Commun. ACM 49 (2006) 93–98.

Shasha Deng is an assistant professor of Information Management and Information Systems in the School of Business and Management at Shanghai International Studies University. She holds a Ph.D. in Management Information Systems from Shanghai Jiaotong University, an MS and Bsc. in Computer Science from Central South University. Her research aims to improve predictive analytics and pattern discovery with applications in social media, online review, and mental health by designing and developing intelligent technologies. Her work has been published in several academic journals and conference proceedings including *MIS Quarterly, PLos One, Journal of Computational Information Systems*, International Conference on Information Systems (ICIS), and Hawaii International Conference on System Sciences (HICSS). Her research has been supported by grants from the NSFC and other industry project.

Yilu Zhou, Ph.D., is an associate professor of Information Systems in the School of Business at Fordham University. Her research interests include natural language processing, text mining, computational data analytics, multilingual knowledge discovery and human-computer interaction. Dr. Zhou received a Ph.D. in Management Information Systems at the University of Arizona where she was also a research associate at the Artificial Intelligence Lab. She received her BS in Computer Science from Shanghai Jiaotong University. Yilu has published in academic journals including MIS Quarterly, Journal of the American Society for Information Science and Technology, IEEE Intelligent Systems and Decision Support Systems. She is a recipient of an NSF grant to study mobile app maturity rating for children.

Pengzhu Zhang is Chair Professor and Director of the Department of Management Information Systems, Antai College of Management and Economics, Shanghai Jiaotong University. Pengzhu received bachelor's and master's degrees from Shandong University of Science and Technology, and a Ph.D. from Xi'an Jiaotong University. His research focuses on the decision support in innovation and health management. His research has been published in many journals including Decision Support Systems, Information & Management, Computers in Human Behavior, PLos One, Journal of the American Society for Information Science and Technology, Journal of Nanoparticle Research, International Journal of Information Technology & Decision Making, Government Information Quarterly, and Journal of Management Analytics, as well as some important journals in Chinese.

Ahmed Abbasi is Murray Research Professor of Information Technology in the McIntire School of Commerce at the University of Virginia. He is Director of the Center for Business Analytics and coordinator for McIntire's Executives-on-Grounds program. Ahmed received his Ph.D. in Information Systems from the University of Arizona, where he also worked as a project-lead in the Artificial Intelligence Lab. He attained an M.B.A. and B.S. in Information Technology from Virginia Tech. Ahmed's research interests relate to predictive analytics and natural language processing with applications in security, health, social media, and customer analytics. His research has been funded through multiple grants from the National Science Foundation. He has also received the IBM Faculty Award, AWS Research Grant, and Microsoft Research Azure Award for his work on Big Data. Ahmed has published over 70 peer-reviewed articles in journals and conferences, including top outlets such as MIS Quarterly, Journal of Management Information Systems, ACM Transactions on Information Systems, IEEE Transactions on Knowledge and Data Engineering, and IEEE Intelligent Systems. One of his articles was considered a top publication by the Association for Information Systems. He has also won best paper awards from MIS Quarterly and WITS. Ahmed's work has been featured in various media outlets, including the Wall Street Journal, the Associated Press, WIRED, CBS, and Fox News, He serves as senior editor for Information Systems Research, and associate editor for ACM Transactions on MIS and IEEE Intelligent Systems. He also previously served as associate editor at Information Systems Research and Decision Sciences Journal, and is a senior member of the IEEE.