

# Utterance Embedding for Detecting Argumentative Topics in Assembly Minutes

Hiroto Yano<sup>1,\*</sup>, Soichiro Yasumoto<sup>1,†</sup> and Kazuhiro Takeuchi<sup>1,†</sup>

<sup>1</sup>osaka electro-communication university

## Abstract

Meeting minutes from government and local assemblies are comprehensive documents that meticulously record the deliberations and discussions of each member. These resources provide crucial information about the background of the decisions and retrieve their path through discussions with final approval. Unlike ordinary texts, minutes encapsulate multi-speaker dialogues, making it imperative to identify argumentative topics in which participants exchange different viewpoints on the matter at hand. This paper presents a novel computational model, rooted in machine learning, that uses speaker alternation patterns to transform each utterance into a vector representation. This model lays the foundation for the analysis of complex textual data as graph representation and holds promise for applications in Explainable Artificial Intelligence (XAI) by aiding in the verification of complex textual context summarization. Using these vectorized utterances, we then formulate clusters that capture the argumentative topic and extract discriminative keywords related to the discussion. The effectiveness of our approach is assessed by contrasting the extracted words with a manually written tree-structured summary.

## Keywords

Argument Graph Mining, Argumentative Topics, Utterance Embedding

## 1. Introduction

The government and local council meeting minutes (hereafter referred to simply as assembly minutes), where each member's statements and discussions are recorded and made public, differ from documents created by individuals as they describe the process of resolving differences of opinion among multiple people. From the assembly minutes, one can verify the background of a particular proposal and how it was discussed and approved. In contrast to one-way mass communication, as exemplified by news texts, much information exchange, such as email and social networking applications, is conducted by multiple participants. Dialogue is the most basic form of dynamic information exchange, but it is often characterized as highly individualized, verbose, and repetitive.

In the field of natural language processing (NLP), the AMI Meeting Corpus [1] is an early work that paves the way for sophisticated analyses of multi-party interactions in meeting environments. With the rich set of annotations and transcriptions to the corpus, it has helped

---

*The 2nd International Workshop on Knowledge Graph Reasoning for Explainable Artificial Intelligence, December 9, 2023, Tokyo, Japan*

\*Corresponding author.

†These authors contributed equally.

✉ mi22a012@oecu.jp (H. Yano); mi23a006@oecu.jp (S. Yasumoto); takeuchi@osakac.ac.jp (K. Takeuchi)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



CEUR Workshop Proceedings (CEUR-WS.org)

researchers delve deep into the nuances of human communication. There are some dialogue studies that summarize dialogues. For example, Liu et al. (2019a) [2] collected a dialogue summary dataset from DiDi customer service center logs, and Gliwa et al. (2019) [3] created the SAMSum corpus.

In this paper we focus on assembly minutes that are not mere transcriptions of casual dialogue. Assembly minutes are structured records of the dialogue in a meeting, and as such, they have characteristics similar to formal documents as well as being dialogues. Consequently, deciphering these documents requires a specialized set of skills. This study uses hand-written summaries of the minutes, each of which is presented as a tree structure. These tree representations serve as essential clues in the complicated process of deciphering and analyzing assembly minutes. The current state of AI summarization cannot explain how it analyzes and summarizes an original text.

Our proposed method uses a large-scale language model to identify argumentative topics that can help decipher the structure of these minutes. Unlike general dialogues, where argumentative topics emerge and evolve through the natural alternation of utterances, assembly minutes present a more rigid structure where such spontaneous alternations are absent. To address this, we employ a model that is able to measure the semantic proximity between utterances, thereby facilitating a more nuanced analysis of the assembly minutes.

Argumentative topics are detected by keyword weighting, like tf-idf and LDA (Latent Dirichlet Allocation), which are common methods for textual keyword analysis. Specifically, based on the vector representation of each utterance with the trained model, the utterances are clustered to obtain a set of utterances that most closely match the argumentative topic. Using this set of utterances as a document, the tf-idf score is applied and the top five ranked words are extracted and compared with the words in the summary text.

## **2. Analysis of Assembly Minutes**

In politics, it is important to accurately and fairly inform citizens of the assembly decision-making process to ensure transparency and fairness. One such method is to record and publish ‘assembly minutes’ of the statements of council participants that were deliberated and discussed by the local government council. Citizens can obtain political information, such as whether each participant took a position in favor of or against a particular issue of interest, from publicly available assembly minutes. However, it is difficult to read the vast number of publicly available assembly minutes. In addition, the recent proliferation of digital media has increased the need for fact-checking to detect fake news and verify the truth and accuracy of the information [4, 5]. To enable citizens to handle assembly minutes as primary information, it is important to improve the searchability and visibility of assembly minutes by identifying their discussion structure.

### **2.1. Argument Scheme**

An argument scheme [6] is a general template or structure that represents a general pattern of reasoning or argumentation and is defined to provide a framework for constructing and

analyzing arguments. The theory of the schemes consists of specific propositions, statements, and argumentation patterns.

For example, given the following statements

- Global temperatures are rising.
- Ice in the polar regions is melting.

The argument that follows from these two statements could be: "Global temperatures are rising, therefore ice in the polar regions is melting." This sentence represents a cause (rising global temperature) and effect (melting polar ice). In this example, the argumentation pattern is indeed 'cause-and-effect', and the underlying premise of the argument can be stated as "If global temperatures rise, then ice in the polar regions melts." This premise is generally accepted based on scientific consensus. This relationship is commonly observed in scientific or fact-based arguments.

Using the concept of the Argument Scheme, analyzing the process of argumentation against assembly (or meeting, discussion) minutes is called argument mining. In recent years, research in the field of natural language processing has been conducted with the goal of identifying argument structures in natural language text. The identification of argument structures involves a variety of tasks, such as separating arguments, classifying argument components into claims and premises, and identifying argument relations. Within argument structure identification, the task we focus on is identifying topics in an argument.

Another advantage of applying Argument Schemes to minutes is their ability to visualize and map the structure of an argument. By breaking down an argument into its fundamental components, including premises, inferences, and conclusions, argument schemes can make the underlying logic of an argument more transparent and easier to understand. This can be particularly helpful in complex discussions where multiple arguments and counter-arguments are being made and where it might otherwise be difficult to keep track of the various points being put forward. Figure 1 shows an example of visualization of discussion. In the figure, Speaker A makes two statements in support of one claim. Speaker B, who responds to A, asserts a disagreement with the opinion of A and also forms a structure that is critical of A's supporting opinion. By identifying argumentation patterns for each utterance and tracking the connections between elements, it is possible to analyze the discussion process.

The difference between a discussion and a text lies in the foundational structure. Unlike conventional texts that are composed of sentences, a discussion is composed of alternating utterances from multiple speakers. As shown in Figure 1, speakers A and B take turns to contribute to the discussion. The basis for analyzing such discussions is the adjacency pair of utterances from different speakers [7, 8, 9].

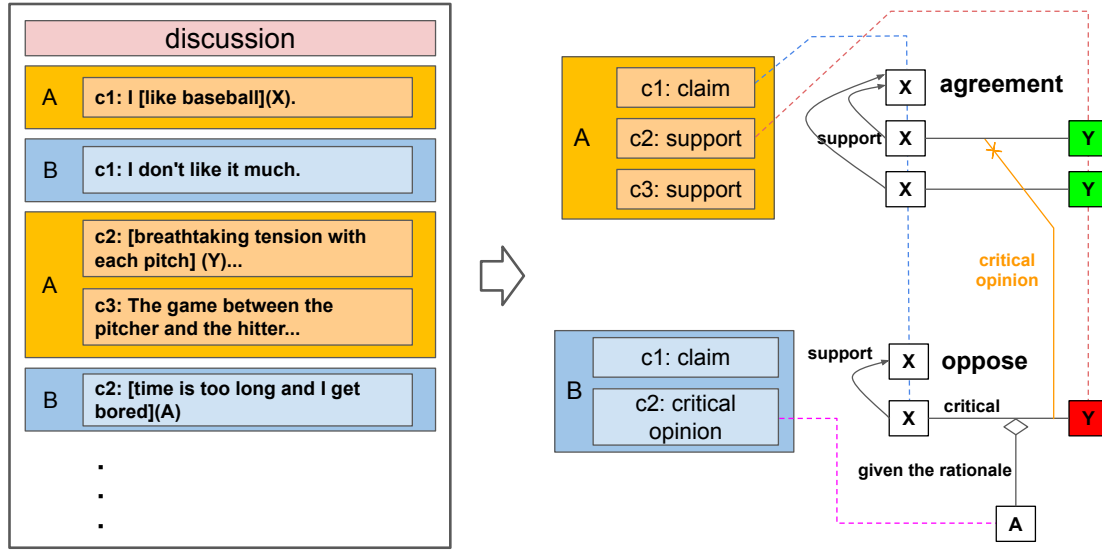
Let target **discussion**  $d_k$ : be a sequence of exchanges.

$$d_k = \{e_1, e_2, e_3, \dots\}$$

It can also be described as:

$$d_k = \{u_1, u_2, u_3, \dots\}$$

Where:



**Figure 1:** Argument Scheme

- $d_k$  is a specific discussion.
- In the first representation, each  $e_i$  denotes an exchange in the discussion.
- In the second representation, each  $u_i$  denotes an utterance in the discussion.
- The subscript  $i$  (whether in  $e_i$  or  $u_i$ ) is a positive integer that represents the position of the exchange or utterance in the sequence, respectively.

**Exchange  $e_i$  (or Adjacency Pair):** A pair of consecutive utterances  $u_i$  and  $u_{i+1}$  where the speaker of  $u_i$  is not the same as the speaker of  $u_{i+1}$ .

$$e_i = (u_i, u_{i+1})$$

Such that:  $s(u_i) \neq s(u_{i+1})$

Where  $s(u_i)$  and  $s(u_{i+1})$  represent the speakers of utterances  $u_i$  and  $u_{i+1}$ , respectively.

## 2.2. Keyword based analysis

In the fields of information retrieval and text processing, quantifying the importance of words within a document relative to a larger corpus is essential for various tasks. One well-known method is the term frequency-inverse document frequency (tf-idf). This numerical statistic quantifies the importance of a term not only based on its frequency in a single document but also offsets its commonality across a larger collection of documents, providing an adjusted measure of term importance. The final tf-idf score for a term in a document is the product of its term frequency (tf) and idf (inverse document frequency) scores. Term Frequency (tf) is derived directly from the BoW representation by counting the number of times a term appears in a document. Inverse Document Frequency (idf) is calculated based on the frequency of a term across all documents in the corpus. In parallel, Latent Dirichlet Allocation (LDA) is another

cornerstone of text analytics. Unlike tf-idf, which focuses primarily on the meaning of terms, LDA is concerned with discovering latent thematic structures present in the corpus. It aims to represent documents as mixtures of topics, where each topic corresponds to a distribution of words.

Notably, both tf-idf and LDA use the basic bag-of-words (BoW) model. Although this model is simple in its representation, focusing on the occurrence of words in a document and ignoring their order, it remains effective in many text-processing scenarios. Rather than relying on documents, this study adopted a discussion-based framework. Although the assembly minutes that are the subject of this paper are typically represented as text, a discussion is not a monologue but consists of utterances from multiple speakers. The unit corresponding to the document in text processing is a discussion characterized by the contributions of multiple interlocutors, where interactions emerge from the interplay of utterances among multiple speakers. The purpose of this paper is to provide such quantifications of word meaning, tailored to assembly minutes.

A text provided as assembly minutes contains multiple discussions that took place on a single day. Since each discussion is considered independent, each discussion can be considered a "document" in tf-idf terms. In other words, when analyzing the minutes, a "document" corresponds to a discussion. To provide clarity and precision in our issue, we now turn to formally represent the problem in this paper.

- $D$  as the target set of discussions. It corresponds to the text of assembly minutes for a specific day.
- $d_i$  as a distinct discussion in  $D$ .

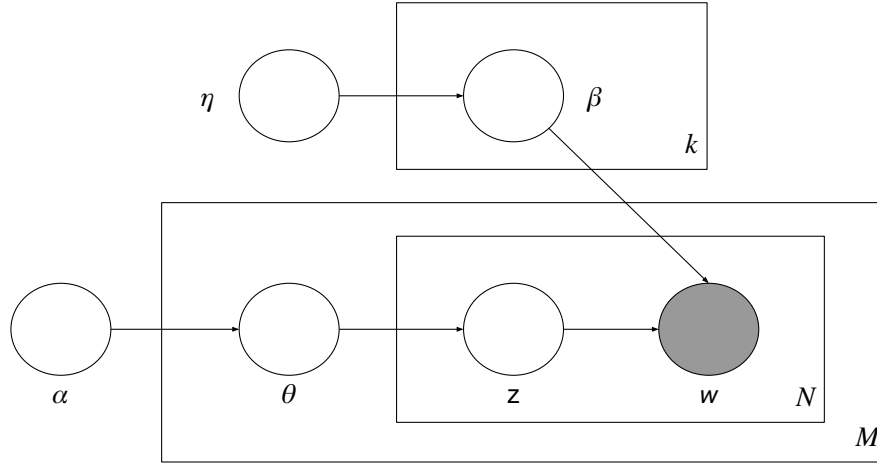
LDA is a probabilistic topic model used to discover topics from a collection of documents. It assumes that each document consists of a mixture of several topics. Each topic is represented as a distribution over words in the vocabulary. While LDA is generally effective in identifying sets of words associated with specific topics, its direct application to the context of real-world discussions can be challenging.

This kind of direct relationship isn't considered in the basic LDA model. Figure 2 shows the plate notation for LDA with Dirichlet-distributed topic-word distributions.

The posterior distribution is intractable to compute directly, so approximation techniques, such as Gibbs sampling or variational inference, are utilized. LDA's strength lies in its ability to reveal underlying topics within text, aiding in document classification, information retrieval, and content summarization.

The BoW model represents text documents as vectors, where each dimension corresponds to a unique word from the entire corpus, and the value represents the frequency or presence of that word in the document. BoW disregards the order of words, focusing only on the occurrence and frequency of words.

LDA is a generative probabilistic model designed to uncover latent topics within a collection of documents. It depicts documents as combinations of topics, with each topic being a probability distribution over words. These outputs, characterized by clusters of co-occurring words, can be interpreted as topics. The problem here is that topics in assembly minutes are not those found in ordinary text or dialogue, but those that form the basis for analyzing structured meeting minutes in a highly empirical way.



**Figure 2:** Plate notation for LDA with Dirichlet-distributed topic-word distributions

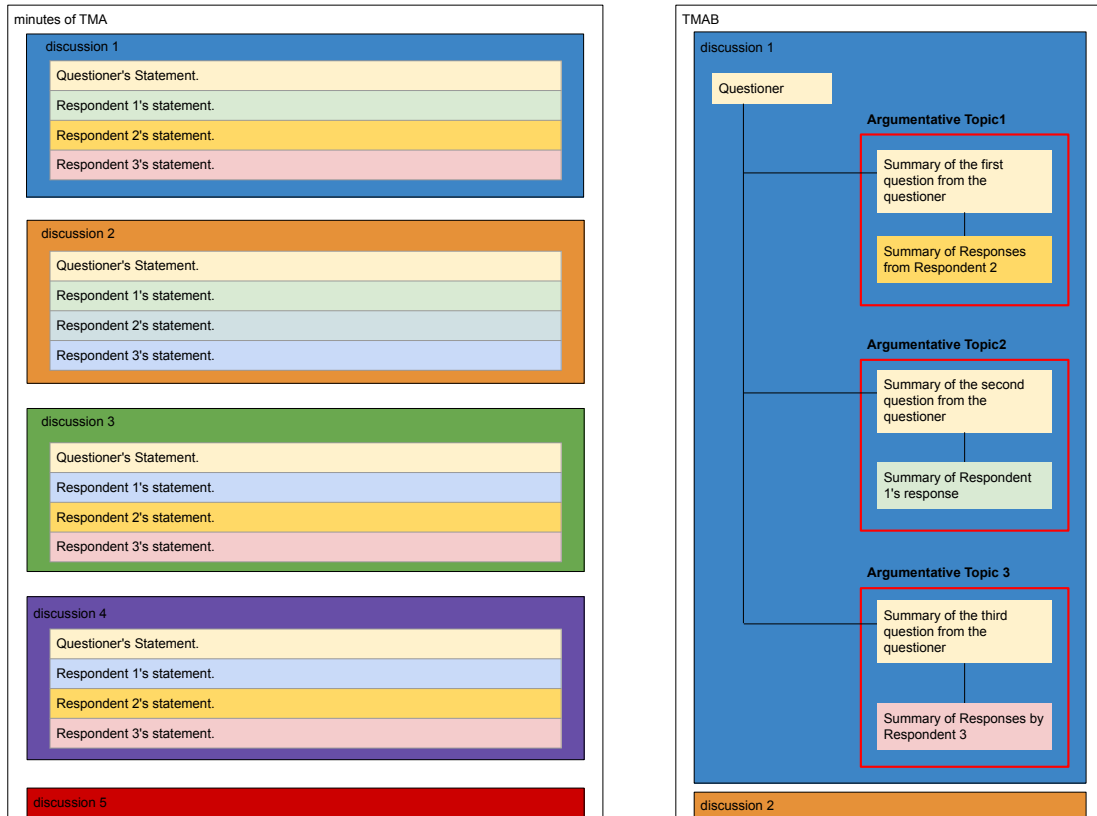
### 3. Data

This paper proposes a method for detecting topics being discussed among multiple speakers from actual assembly minutes, which is a basis to form an argument scheme. For the actual assembly minutes, we use the minutes of the Tokyo Metropolitan Assembly meetings (minutes of TMA), in which the utterances made in actual assemblies are recorded. For the evaluation of the proposed method, on the other hand, we will use the Tokyo Metropolitan Assembly Bulletin (TMAB), in which the utterances by specific speakers and the question-and-answer sessions in response to those utterances are manually summarized.

Figure 3 shows the relationship between the discussions in the meeting minutes and the summaries in TMAB. The summary in the figure is simple, providing a single topic shared between the two speakers, and summarizing each speaker’s argument as a short phrase. Each speaker’s summarized phrase is often shorter than the phrases that we refer to. Each speaker’s assertion about a topic is expressed in as few summary phrases as possible.

There is a difference between general discussions and those in the actual minutes. Of course, the constituent units of text are sentences and discussions are utterances, but the actual minutes do not consist of the general spoken utterances themselves. In a manner of speaking, the minutes can be described as ‘pseudo-written language.’ One characteristic of this pseudo-written language is that it often forms lengthy sentences as shown in Table 1. Such long sentences are difficult to handle. As a preprocessing step for meeting analysis, we divide pseudo-sentences into smaller units, which we refer to as clauses. So, we use preprocessing to analyze utterances by dividing them into ‘clauses,’ which are shorter units than sentences. That is, we have to analyze which topic each ‘clause’ relates to in the utterance.

Another much larger difference exists between utterances in general discussions and those in the actual minutes. That is, in the minutes in Figure 3, each speaker does not take a simple alternation of speakers in a general discussion. Moreover, as the manually provided structural summary of TMAB in the figure shows, there are multiple speakers responding to a single



**Figure 3:** Relationship between a discussion in assembly minutes and its summary in TMAB

speaker's question. This is due to the fact that the assembly is planned and facilitated in advance. We assume that the summary provided by TMAB is an important requirement for the analysis of the assembly minutes. Specifically, one questioner provides multiple argumentative topics, and multiple people answer each argumentative topic. Therefore, without detecting the argumentative topic for each pair of questioner and respondent from the minutes, it is not possible to perform the usual structural analysis of the discussion. In this paper, we propose a method to detect argumentative topics in assembly minutes and evaluate detected argumentative topics compared with TMAB's manual summary.

## 4. Proposed Method

In this section, we present a new computational model rooted in machine learning that uses speaker alternation patterns in meeting minutes as the basis for training data.

Figure 4 shows a schematic of the entire proposed method.

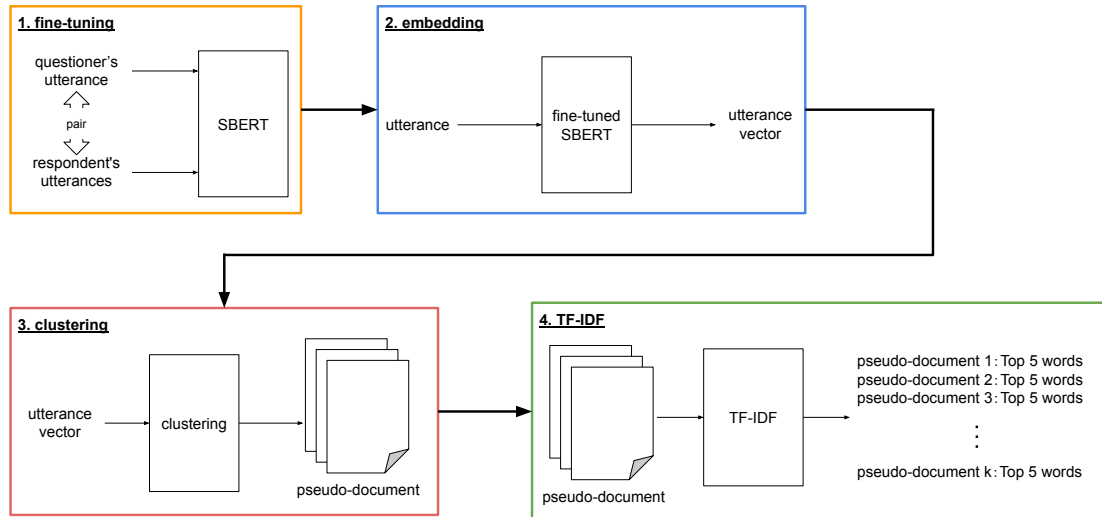
From Figure 4, the proposed method uses the following procedure to extract argumentative topics.

- Fine-tuning SBERT by pairing the utterances of the questioner and the respondent

**Table 1**  
Example utterances in TMA

Questioner	<p>As a member of the Liberal Democratic Party of the Tokyo Metropolitan Assembly, I would like to ask several questions regarding issues of the metropolitan government. I look forward to clear answers from the Governor and other directors.</p> <p>The construction of the Tsukuba Express is being carried out amid great expectations of the people of Tokyo.</p> <p>Ltd. and the three prefectures concerned have announced that the 58-km line between Akihabara and Tsukuba is scheduled to open simultaneously in the fall of 2005. Until now, there was a time when even achieving the opening target of fiscal 2005 was in doubt due to delays in some land acquisitions and other factors, but what is the rationale behind this clarification of the opening date?</p> <p>The Tsukuba Express was originally planned as a national project to alleviate congestion on the Joban Line and to provide access to Tsukuba Science City, as well as to provide a large supply of residential land, and the so-called Housing Railway Law was enacted by the Diet.</p> <p>However, compared to when it was conceived in 1989, the country's socioeconomic situation has changed dramatically today. In fact, the latest announcement revised downward the transportation demand to less than 300,000 persons per day. Since the initial plan was for approximately 600,000 passengers, we have no choice but to estimate the demand at half of that figure, and we expect to face severe business management challenges after the opening.</p> <p>...</p>
Respondent 1	<p>I would like to answer a general question from Councilor Naoki Takashima.</p> <p>Regarding the idea of abolishing the special industrial district building ordinance, in order to strengthen Tokyo's manufacturing and other industries, it is necessary to create an environment that allows factories to be rebuilt or expanded as freely as possible.</p> <p>In response to the Tokyo Metropolitan Government's request, last July the national government abolished the Law on Industrial Restriction, which regulates the location of factories in urban areas.</p> <p>The Tokyo Metropolitan Government is also committed to strengthening its industrial capacity through the early elimination of restrictions on factory locations and various forms of support.</p> <p>We will continue to coordinate with the municipalities to repeal the special industrial district building ordinance.</p> <p>The other questions will be answered by the Director General concerned.</p>
Respondent 2	<p>I would like to answer five questions regarding the new Joban Line, etc.</p> <p>First, regarding the timing of the opening of the new Joban Line, by the end of last year, discussions on the intersection with the Sobu Nagareyama Electric Railway, which had been an issue in the construction process, were completed, and the necessary railroad land was almost completely secured.</p> <p>Based on this, and as a result of coordination among the parties involved, major civil works are expected to be completed by the end of the fiscal year.</p> <p>Furthermore, taking into consideration the progress of the construction of the station building and facilities, as well as the process of running tests of the trains, we have come to the conclusion that the station will open in the fall of 2005.</p> <p>Next, regarding the economic effects of the Joban New Line, according to a study conducted in 1997 by the Joban New Line Project Promotion Council, which was organized mainly by private companies, the direct effects of the Joban New Line project are estimated to be approximately 1 trillion yen for the construction and operation of the railroad and approximately 6 trillion yen for housing development and public investment in areas along the line, for a total of approximately 7 trillion yen over the 30 years from 1996.</p> <p>...</p>





**Figure 4:** Overview of the entire proposed method

- Create utterance vectors using SBERT with fine-tuning
- Clustering of utterance vectors to form pseudo-documents
- Calculate TF-IDF for the formed pseudo-document and extract the top 5 words

The details of each procedure are explained in the following sections.

#### 4.1. Proposed model for utterance embedding

In order to determine the argumentative topic from the meeting minutes, it is necessary to find utterance clauses that are strongly related to the argumentative topic from the questioner's utterance. Also, as mentioned in section 3, the meeting minutes used must detect the argumentative topic for each pair of questioner and respondent, since one questioner provides multiple arguments and multiple people respond to each argument. Therefore, by learning the utterances of the questioner and the respondent as a pair, we form utterance vectors that represent the argumentative topic.

Sentence-BERT [10] (hereafter SBERT) is used to form utterance vectors. SBERT is a modification of the pretrained BERT network that uses siamese and triplet network structures to derive semantically meaningful sentence embeddings that can be compared using cosine-similarity.

In this paper, we perform fine-tuning of SBERT using the training data of the questioner-answer utterance pairs and triples of unrelated utterances in the minutes.

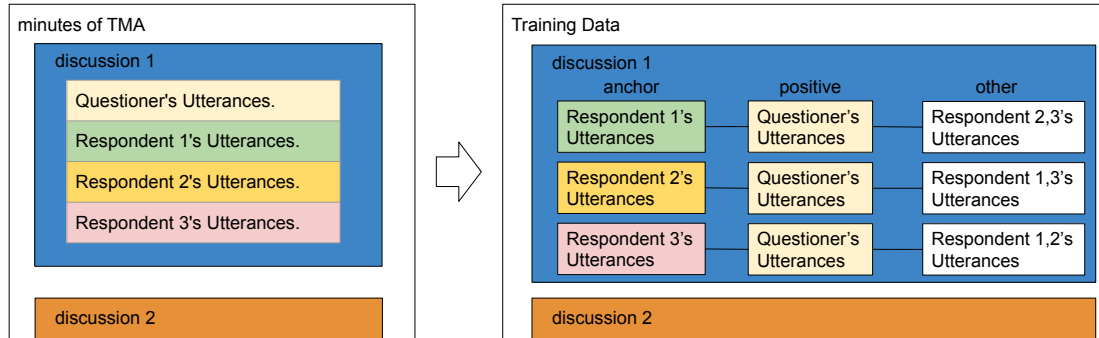
#### 4.2. Training

Create training data for fine-tuning SBERT from meeting minutes.

Figure 5 shows a schematic of creating training data from meeting minutes.

The training data is created with each respondent's utterance for each discussion as the anchor, the questioner's utterance as the positive, and the utterances of respondents other than the respondent who is the anchor as the other.

Using this data, SBERT is trained so that the utterance vectors of a questioner and a respondent within the same discussion in the meeting minutes are close and the utterance vectors of a respondent and another respondent are far apart.



**Figure 5:** Created training data for the pair having the discussion from the meeting minutes

Figure 6 shows the structure of SBERT's learning model. A Dense layer is added after the pooling layer in order to learn the utterance clauses in the meeting minutes with low-dimensional vectors, and in this paper, the training is performed with 10 dimensions.

The vector representation is formed using SBERT after fine-tuning for the utterance clauses of the same meeting minutes used for training.

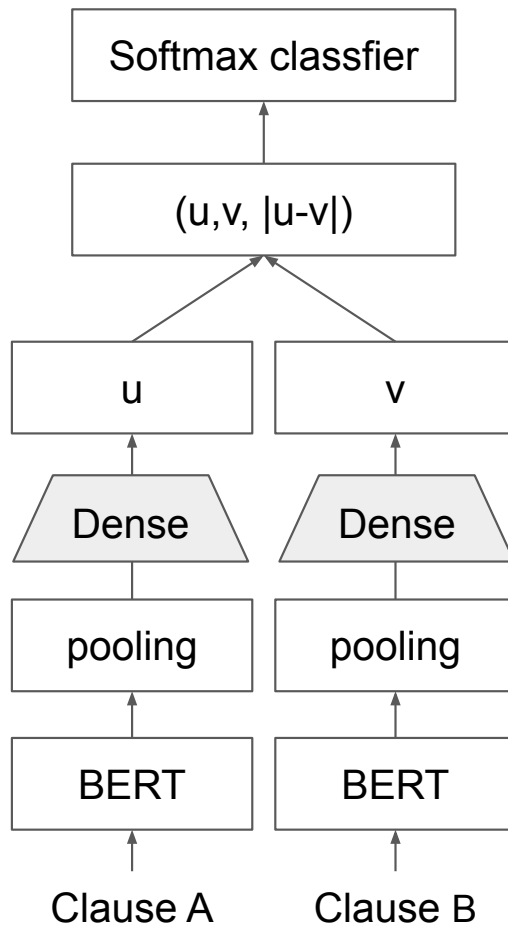
### 4.3. Clustering based on neighbor pairs

Clustering is performed on the formed utterance vectors to form a set of utterances that are strongly related to the argumentative topic. Here, since SBERT is trained on questioner-respondent pair data, each cluster is a set of utterances related to an argumentative topic.

### 4.4. Extracting words for discussion

Extract words corresponding to the argumentative topic from the clusters formed. Inspired by BERTopic[11], this paper investigates the extraction of words corresponding to argumentative topics using tf-idf, where all utterance clauses in each cluster are considered as one pseudo-document.

BERTopic is a topic model that extracts consistent topic representations through class-based tf-idf. First, document embeddings are created using SBERT. Next, the created document embeddings are classified by clustering, and then class-based tf-idf is used to generate topic representations for the clusters that are formed. An important aspect of BERTopic is that it generates topic representations by clustering document embeddings obtained from a large-scale language model and ranking sentences using class-based tf-idf. Other embedding techniques



**Figure 6:** Training of Sentence-BERT model

can also be used if the language model generating the document embedding is fine-tuned for semantic similarity.

Since each cluster is a set of utterance clauses that are strongly related to an argumentative topic, a pseudo-document composed of utterance clauses that discuss a single argumentative topic can be created by concatenating all the utterance clauses that belong to a cluster. Important keywords in the argumentative topic are extracted by weighting words by tf-idf for the created pseudo-document.

Figure 7 shows the method of extraction and evaluation from clusters. First, the utterance clauses belonging to each cluster are concatenated and used as pseudo-document to weight words by tf-idf. Next, the utterance vector is used to filter out unnecessary clauses that do not contribute to the identification of keywords. The center of gravity is calculated from the utterance vectors of each cluster, the similarity between the calculated center of gravity and each utterance vector is measured, and utterance clauses with low similarity are deleted. After

filtering, the top 5 words of tf-idf values are extracted.

Here, the top 5 words of the extracted tf-idf values are compared with the argumentative topic contained in TMAB, and evaluated based on how well the keywords related to the argumentative topic are extracted by the partial match ratio of the words.

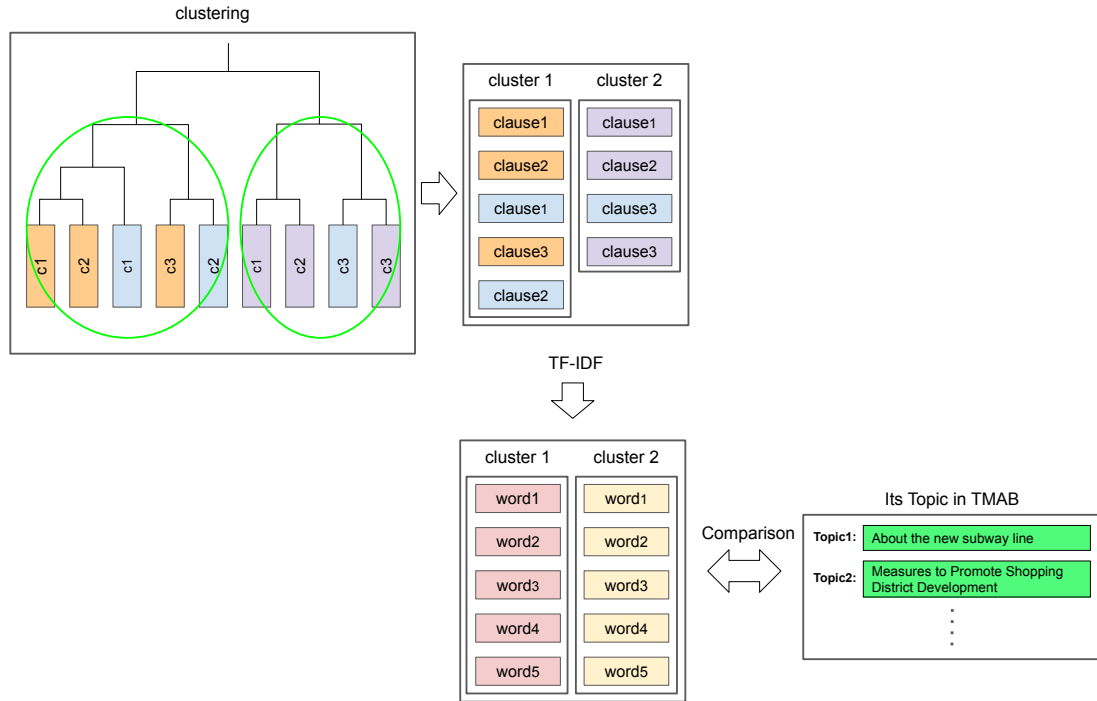


Figure 7: How TMAB compares topics and extracted words

## 5. Experiments and Results

Perform comparative experiments with LDA and tf-idf. LDA is given the combined text of all utterances of the questioner and respondent in a day's worth of conference discussion minutes, and extracts the top five words for each topic. tf-idf extracts Document Frequency (DF) from a day's worth of meeting minutes and Term Frequency (TF) from the combined text of all utterances of the questioner and respondent in the same day's discussion minutes, and extracts the top five words. As with the proposed method, the evaluation is based on the partial agreement rate between the top five extracted words and the argumentative topic contained in the TMAB.

The results of word agreement rates between the top five words extracted by the proposed method, LDA, and tf-idf and the TMAB summary are shown in Table 2. Table 2 confirms that the proposed method is able to extract more keywords that are strongly related to the argumentative topic compared to other methods. This suggests that in the comparison method, since the questioner asked multiple questions in a single text, it was not possible to differentiate

the information from the other questions when weighting words, and thus keywords related to the argumentative topic were not extracted. In the proposed method, clustering using utterance vectors to create clusters of question utterances that are related to the respondent's answer utterances is found to contribute significantly to the extraction of keywords related to the argumentative topic.

**Table 2**

Comparison of proposed method, LDA, and tf-idf

	Target	Proposed Method	LDA	tf-idf
data 1	49	21	9	1
data 2	48	22	9	0
data 3	47	17	7	0
ave.	48	20.0	8.333	0.333

Also,

- Example of agreement for all methods
- Proposed method, example matched by LDA
- Example matched only by the proposed method
- Example of only LDA matched

The table below shows four examples. Table 3 shows an example of the results matched by all methods. Table 4 shows examples of results matched by the proposed method and LDA. Table 5 shows an example of results matched only by the proposed method. Table 6 shows an example of results that matched only the LDA.

**Table 3**

One example where the proposed method, LDA, and tf-idf are all Targeted.

Target	smart interchange
Proposed Method	<b>'interchange'</b> , 'Tachikawa City', 'babysitter', 'Yuriko', 'metropolitan bus'
LDA	<b>'interchange'</b> , 'babysitter', 'shopping district', 'ene', 'electric power'
tf-idf	<b>'interchange'</b> , 'aggregate', 'speech and behavior', '401', 'arrest'

**Table 4**

An example of the proposed method and LDA only when target is applied.

Target	blood donation
Proposed Method	<b>'blood donation'</b> , 'loss', 'foodstuff', 'sightseeing', 'phenomenon'
LDA	<b>'blood donation'</b> , 'foodstuff', 'loss', 'densely wooded area', 'goods'
tf-idf	'2033', 'this year', 'imagination', 'Europe', 'general'

Table 5 shows that LDA extracts keywords related to another question of the questioner, while the proposed method does not extract keywords related to another question, suggesting that the use of questioner and answer pairs in the proposed method is effective for the problem of determining the argumentative topic from the meeting minutes.

**Table 5**

An example where only the proposed method is targeted.

Target	national health insurance
Proposed Method	‘soft’, ‘ <b>insurance</b> ’, ‘some time ago’, ‘beginning’, ‘lighting’
LDA	‘the US armed forces’, ‘self-employed’, ‘Yokota’, ‘accident and sickness benefits’, ‘freelance’
tf-idf	‘script’, ‘Tottori prefecture’, ‘2013’, ‘origin’, ‘fair’

**Table 6**

An example where only the LDA is targeted.

Target	Tokyo Big Sight
Proposed Method	‘analog’, ‘100’, ‘fraud’, ‘pile’, ‘paper-based’
LDA	‘venue costs’, ‘ <b>Tokyo Big Sight</b> ’, ‘return’, ‘job training’, ‘advertisement’
tf-idf	‘INTEX’, ‘ability development’, ‘lengthening’, ‘cancel’, ‘Kawamura’

## 6. Conclusion

Clustering of utterance vectors and extraction of keywords related to argumentative topics using tf-idf was performed. By fine-tuning SBERT with training data using the questioner’s and respondent’s utterances as pairs, an utterance vector representing the argumentative topic was formed. The results suggest that the proposed method, unlike other comparison methods, solves the problem of questioners asking multiple questions in a single text and may be able to identify argumentative topics.

This paper developed a computational model for representing speech vectors by fine-tuning a large-scale language model that incorporates a model of speech alternation within an argument context. This model made it possible to cluster utterances in meeting minutes. Using CLS vectors without fine-tuning BERT seemed to be an appropriate baseline; however, we lost time in the validation experiments while considering fine-tuning methods for better comparisons. In future work, we would like to establish a better baseline and evaluate the validity of this study more precisely. As another future task, we will consider improving the accuracy of the proposed method by conducting a study on summaries in which the proposed method could not agree. In addition, this study uses tf-idf for word extraction, but we would like to consider using a probabilistic model to generate an argumentative topic.

## Acknowledgments

This work was supported by JSPS KAKENHI Grant Number 23H03462.

## References

- [1] J. Carletta, S. Ashby, S. Bourban, M. Flynn, M. Guillemot, T. Hain, J. Kadlec, V. Karaiskos, W. Kraaij, M. Kronenthal, et al., The ami meeting corpus: A pre-announcement, in: International workshop on machine learning for multimodal interaction, Springer, 2005, pp. 28–39.

- [2] C. Liu, P. Wang, J. Xu, Z. Li, J. Ye, Automatic dialogue summary generation for customer service, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019, pp. 1957–1965.
- [3] B. Gliwa, I. Mochol, M. Biesek, A. Wawer, Samsun corpus: A human-annotated dialogue dataset for abstractive summarization, arXiv preprint arXiv:1911.12237 (2019).
- [4] Y. Kimura, H. Shibuki, H. Ototake, Y. Uchida, K. Takamaru, M. Ishioroshi, T. Mitamura, M. Yoshioka, T. Akiba, Y. Ogawa, M. Sasaki, K. Yokote, T. Mori, K. Araki, S. Sekine, N. Kando, Overview of the ntcir-15 qa lab-poliinfo-2 task, in: Proceedings of the 15th NTCIR Conference on Evaluation of Information Access Technologies, volume -, 2020, pp. 101–112.
- [5] Y. Kimura, H. Shibuki, H. Ototake, Y. Uchida, K. Takamaru, M. Ishioroshi, M. Yoshioka, T. Akiba, Y. Ogawa, M. Sasaki, K. Yokote, K. Kadowaki, T. Mori, K. Araki, T. Mitamura, S. Sekine, Overview of the ntcir-16 qa lab-poliinfo-3 task, in: Proceedings of The 16th NTCIR Conference, volume -, 2022, pp. 156–174.
- [6] D. Walton, C. Reed, F. Macagno, Argumentation schemes, Cambridge University Press, 2008.
- [7] M. Higashiyama, K. Inui, Y. Matsumoto, Learning sentiment of nouns from selectional preferences of verbs and adjectives, in: Proceedings of the 14th Annual Meeting of the Association for Natural Language Processing, 2008, pp. 584–587.
- [8] T. D. Midgley, S. Harrison, C. MacNish, Empirical verification of adjacency pairs using dialogue segmentation, in: Proceedings of the 7th SIGdial Workshop on Discourse and Dialogue, 2006, pp. 104–108.
- [9] E. Jamison, I. Gurevych, Adjacency pair recognition in wikipedia discussions using lexical pairs, in: Proceedings of the 28th Pacific Asia Conference on Language, Information and Computing, 2014, pp. 479–488.
- [10] N. Reimers, I. Gurevych, Sentence-bert: Sentence embeddings using siamese bert-networks, in: EMNLP/IJCNLP (1), 2019.
- [11] M. Grootendorst, Bertopic: Neural topic modeling with a class-based tf-idf procedure, arXiv preprint arXiv:2203.05794 (2022).