Takelab at the NTCIR-16 QA Lab-PoliInfo-3 Task

Yo Amano
Osaka Electro-Communication University
Japan
mi21a001@oecu.jp

Masayuki Matsumoto
Osaka Electro-Communication University
Japan
gp18a177@oecu.jp

Kousuke Sasaki
Osaka Electro-Communication University
Japan
mi20a003@oecu.jp

Kazuhiro Takeuchi
Osaka Electro-Communication University
Japan
takeuchi@osakac.ac.jp

ABSTRACT

This paper proposes a method for budget argument mining using topic extraction based on utterance classification. We employ a domain-specific word embedding, which is calculated only from the given data, to link budget descriptions with corresponding arguments.

KEYWORDS

argument structure analysis, topic extraction, utterance classification

TEAM NAME

takelab

SUBTASKS

Budget Argument Mining

1 INTRODUCTION

This paper describes an implementation for the budget argument mining task based on our hypothesis that a graph between topics and speaker-specific opinions can visually represent an argument. We analyze the given discussions with topic extraction based on utterance classification and employ a domain-specific word embedding to link budget descriptions with corresponding arguments.

2 OUR APPROACH

In the QA Lab-PoliInfo-3[5] budget argument mining task, we expand one we employed in the QA Lab-PoliInfo-2 topic detection task to a more detailed argument analysis [3]. Specifically, the followings are the extending points:

- Introduce argument analysis for utterance clause type classification
- Represent topics as vectors using word embedding
- Calculate the links from budget description to candidate discussions based on the topic representation

The Figure 1 shows our model that we employed last year. In the figure, there is a premise that arguments can be analyzed using a graph consisting of links that connect words (nodes) with who uttered them. In other words, we regard co-occurred words among participants as words representing the topics to discuss, while words that were used unevenly by only either participant are considered opinions about the topic.

Based on the premise, we extract co-occurrence words as the topics of discussion shared by the participants. Since many words appear in the discussion minutes, we employ the Latent Dirichlet Allocation to word weighting and select the words with their weight.

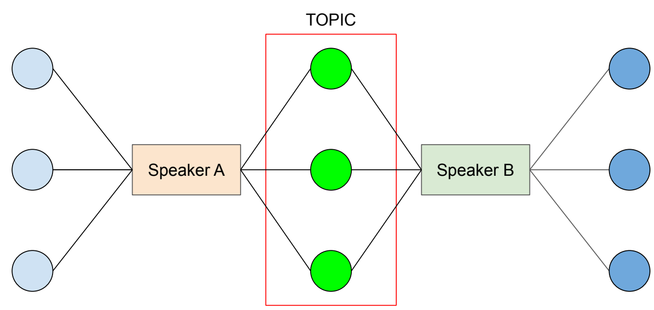

Figure 1: Our argument representation model at QA Lab-PoliInfo-2

We extend the topic extraction model adopted last year to better reflect the structure of this year's discussions. A discussion consists of one or more utterances from more than one speaker. For example, Figure 2 shows an example of a discussion structure between two speakers. In this figure, speaker A interacts with B alternately. Each speaker change is the boundary of the basic dialogue unit that consists of one or more utterance clauses, and we adopted a model that regards a discussion as consisting of a series of adjacency pairs [6] [4].

We regard these serieses of the adjacency pairs as discourse segments and divide a given discussion into discourse segments that correspond to the discussed topics.

There are words in the utterance that strongly reflect the speaker's opinion, and such words are assigned to the nodes that indicate the corresponding speaker' s viewpoint or stance to the topic node. To convert a discussion to a graph representation of the argument, we identify utterance clauses in which each speaker asserted his/her opinion. As training examples for learning this identification model, we use clause type labels that the task organizer for budget argument mining assigns. While the training examples are limited to only clauses containing monetary expressions, we classify all utterance clauses. Figure 5 shows our idea of making the graph representation of the argument from the classification results of the utterance clauses described in Figure 2.

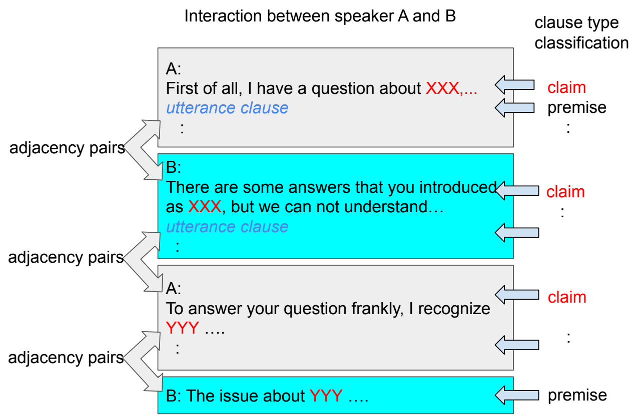

Figure 2: Structure of discussion

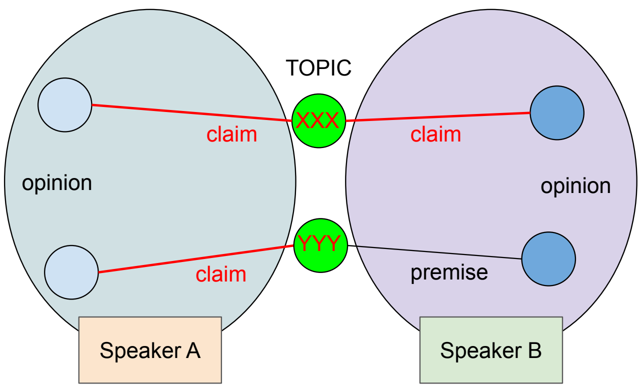

Figure 3: Our argument representation model at QA Lab-PoliInfo-3

Before we assign the topic nodes using clauses classification, we firstly use the Latent Dirichlet Allocation (LDA)[1] to extract topic node candidates from given discussion. LDA requires that the number of topics comprising the discussion be assumed in advance. For each word that appears in the topics, LDA outputs an index value indicating which topic it belongs to.

Figure 4 shows an example of a discussion log for a given day (in this example, February 19, 2019 (Fukuoka City)) divided into 1-39 topics. The vertical axis of the figure shows the average number of adjacency pairs to which the words in each topic can correspond. For example, if the number of topic divisions is one, then all words belong to one topic. Therefore, the general words that the one topic should contain correspond to many adjacency pairs that appear in any given part of the discussion. On the other hand, as the number of topic divisions increases, the fewer words in each topic will correspond to only a few adjacency pairs. Based on this decreasing curve, we use the words extracted with the smallest topic division where the average number of corresponding adjacency pairs closes to 1.0 as topic node candidates.

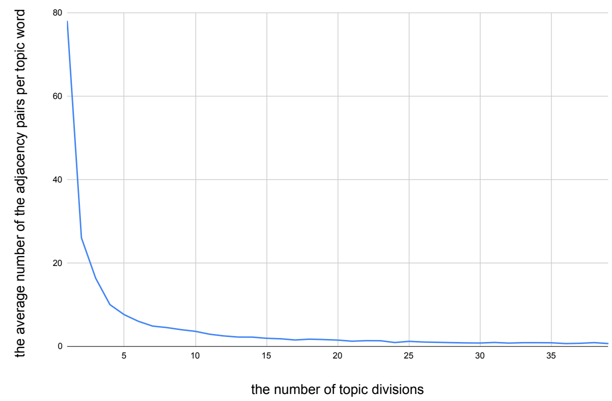

Figure 4: An example of topic separation to adjacecy pairs according to the number of topic devisions

3 UTTERANCE CLASSIFICATION

In our argument analysis, we focus on all the utterances of the speaker, not only the utterance clauses containing monetary expressions. This is because only the utterance clauses containing monetary expressions provide little information to identify the argument class. In other words, we use not only those clauses but also the wider context to extract the topic of the argument.

Discussions are generally structured as shown in Figure 2, with speakers taking turns to speak. For example, when a topic item is proposed, the next speaker states his or her approval or disapproval of the proposal. We regard utterances containing positive words as having a positive opinion, and utterances containing negative words as having a negative opinion. That is the basis of our clause type classification algorithm.

In order to identify argument (clause) classes, we use the sentiment words contained in each utterance using one of the Japanese sentiment dictionaries [2], that has been created by acquiring the evaluation polarity of nouns that frequently occur in the Web corpus.

The dictionary provides information on whether the registered word belongs to p (positive), n (negative), or e (neutral). Table 1 illustrates some of the entries in the dictionary. Using the information of positive words for positive opinions and negative words for negative opinions contained in the sentences, we classify the utterance clauses into the seven argument classes defined in this task.

We use a model constructed with a neural network for the actual calculation of the classification. In our model, we prepare an input layer that corresponds to the occurrence of sentiment words in the target utterance. The hidden layer consists of 100 units and the output layer corresponds to the seven labels defined as the argument classes for this task. Figure 5 shows the learning curve. This figure

Table 1: some examples of the entries in the dictionary

p (positive)	e (neutral)	n (negative)
サービス	10 カ月	ダメ出し
意気込み	3 か月	環境汚染
穏便	遠慮がち	間違い
救出	基本合意	逆境
現実味	減速	厳しさ
賛同	作用	思い過ごし
子育て支援	市民活動	指摘
持続	事業活動	時代遅れ
収益	就職	修正箇所
常識	常人	常識外れ

shows that the identification accuracy of utterances labeled with ‘claim’ types is about 0.75.

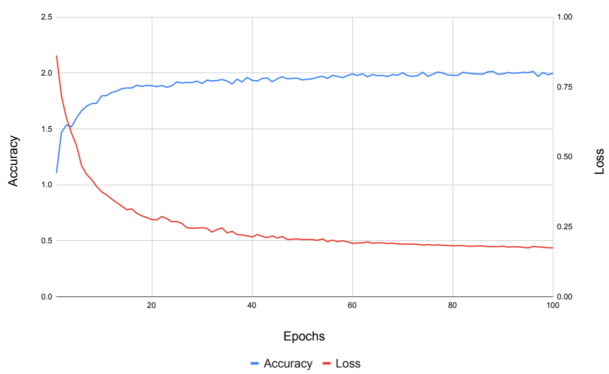

Figure 5: Accuracy (and loss) vs. epochs for training utterance classification

On the NTCIR-16 QA Lab-PoliInfo-3 leaderboard, the best performance of our implementation is (score, AC, RID) = (0.0426, 0.3942, 0.0638). We attempted to create a neural sequence model that reflects the interaction between the utterances, but it did not work very well. So we tuned our implementation at the formal run phase to the model explained here that employs a sentiment dictionary to represent the utterances.

4 LINK IDENTIFICATION

As we have explained above, a given set of discussions (a set of discussion minutes) is analyzed to extract topics based on the utterance clause classification. We tried to extract topic words and opinion words based on a sequence model for utterance clause classification, but as explained in section 3, we changed the model to one based on a sentiment dictionary.

For this purpose, more than 50 rules for extracting words were created and examined. One example of those rules is that one extracts topicalized phrases from the attendant structure of an utterance.

We employ word embedding techniques to represent topics, and we calculate the link from a budget description to the candidate discussions based on the representation. We embed words from a small amount of domain data that consists of all of the given discussion minutes and budget descriptions. With such a small amount of data to obtain a domain-dependent distributed space, the dimension of the space to be low. The identification of links was performed based on the cosine similarity between all nouns in the budget description and the topic vectors of each discussion. We set a threshold value to determine the links.

Our implementation of link identification module has many points of adjustment, such as

- Topic word selection based on a speech classification model
- The number of dimensions for the distributed representation of topic words and budget descriptions
- Threshold to determine link

Since the settings of those parameters were manually controlled, adjusting the accuracy of the link was a trial and error process. The best performance of our implementation achieves (score, AC, RID) = (0.0426, 0.3942, 0.0638) at the formal run phase. We could not obtain good accuracy, especially the RID results, which reflect the difficulty of the adjustment.

5 CONCLUSION

This paper described the methods we employed on the budget argument mining task and our hypothesis behind the implementation. Although we encountered various problems in the actual implementation and could not obtain good results, we would like to improve the modules further.

REFERENCES

[1] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research* 3, Jan (2003), 993–1022.

[2] Masahiko Higashiyama, Kentaro Inui, and Yuji Matsumoto. 2008. 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*. 584–587.

[3] Yuya HIRAI, Yo AMANO, and Kazuhiro TAKEUCHI. 2020. NTCIR-15 QA LAB-PolIinfo-2 dialog topic detection based on discussion structure graph. In *Proceedings of the 15th NTCIR Conference on Evaluation of Information Access Technologies.*

[4] Emily Jamison and Iryna Gurevych. 2014. Adjacency Pair Recognition in Wikipedia Discussions using Lexical Pairs. In *Proceedings of the 28th Pacific Asia Conference on Language, Information and Computing*. 479–488.

[5] Yasutomo Kimura, Hideyuki Shibuki, Hokuto Ototake, Yuzu Uchida, Keiichi Takamaru, Madoka Ishioroshi, Masaharu Yoshioka, Tomoyoshi Akiba, Yasuhiro Ogawa, Minoru Sasaki, Kenichi Yokote, Kazuma Kadowaki, Tatsunori Mori, Kenji Araki, Teruko Mitamura, and Satoshi Sekine. 2022. Overview of the NTCIR-16 QA Lab-PoliInfo-3 Task. *Proceedings of The 16th NTCIR Conference* (6 2022).

[6] T. Daniel Midgley, Shelly Harrison, and Cara MacNish. 2006. Empirical Verification of Adjacency Pairs Using Dialogue Segmentation. In *Proceedings of the 7th SIGdial Workshop on Discourse and Dialogue*. Association for Computational Linguistics, Sydney, Australia, 104–108. https://aclanthology.org/W06-1315