fuys at the NTCIR-16 QA Lab-PoliInfo-3 Budget Argument Mining Subtask

Daigo Nishihara
Fukuoka University
Japan
td222007@cis.fukuoka-u.ac.jp

Hokuto Ototake
Fukuoka University
Japan
ototake@fukuoka-u.ac.jp

Kenji Yoshimura
Fukuoka University
Japan
yosimura@fukuoka-u.ac.jp

ABSTRACT

This paper reports on the achievements of Budget Argument Mining subtask of the NTCIR-16 QA Lab-PoliInfo-3 task of fuys team. We have assigned ArgumentClass and RelatedID in different ways. ArgumentClass was assigned using BERT. We also thought that the accuracy could be improved by adding the flag indicating whether a speaker is a legislator or not ("giin-flag"). RelatedID was assigned using keyword extraction with TFIDF. The results showed good results for ArgumentClass, but the improvement in accuracy of RelatedID could not be confirmed. Although there was a difference in the results for the "giin-flag", the difference was small, and no advantage was found with or without the "giin-flag".

KEYWORDS

BERT, TFIDF, sentence splitting

TEAM NAME

fuys

SUBTASKS

Budget Argument Mining subtask (Japanese)

1 INTRODUCTION

We, fuys team, participated in Budget Argument Mining (BAM) subtask of the NTCIR-16 QA Lab-PoliInfo-3 task [2].

We thought that BAM could be divided into two sub-sub-tasks: ArgumentClass and RelatedID. When the author observed the dataset ourselves and actually tried to manually annotate the ArgumentClass and RelatedID, we assigned them in different ways in each. We determined ArgumentClass by reading the content of the minutes, and RelatedID by looking for keywords in the minutes that could be used as keywords in the budget item. Therefore, we thought it would be better to assign ArgumentClass and RelatedID in different ways.

2 ANNOTATION

The author participated in the annotation of test and training data for this task. In this section, I will show you how I determined the ArgumentClass and RelatedID in the annotation.

2.1 ArgumentClass

We determined ArgumentClass by reading the content of the minutes. We read the utterances where a monetary expression was present and determined the argument labels for that monetary expression. When determining the argument labels for monetary expressions, we judged them from the sentence or clause in the utterance in which the monetary expression is found. If we could not determine it, we read the preceding and following sentences in the utterance. We will explain how made those decisions, using the minutes of the first regular meeting of Fukuoka City in 2nd year of the Reiwa period as an example.

We labeled the monetary expressions that we could determine from the utterance with the monetary expression as amounts that had been decided or used in the past as "Premise : Past and Decisions". In the example on row of "Premise : Past and Decisions" of the Table 1, we determined that the monetary expression in this sentence is money spent in the past, since it says "平成30年度決算(Fiscal Year 2018 Settlement of Accounts)". We labeled monetary expressions that appear in the utterance talking about estimates or the proposed budget for the year as "Premise : Current and Future / Estimates". In the example on row of "Premise : Current and Future / Estimates" of the Table 1, we determined that they were talking about the year in question, given that they say "今回の補正規模は(Scale of this budget amendment is ...)". We labeled the monetary expressions in the factual and documented utterances as "Premise : Other". In the example on row of "Premise : Other" of the Table 1, we labeled it "Premise : Other" because we determined that it was a monetary expression that came up when talking about actual income.

We labeled the monetary expressions as "Claim : Opinions, suggestions, and questions" if we could determine that the utterance with the monetary expression was the speaker's claim. In the example utterance on row of 'Claim : Opinions, suggestions, and questions" of the Table 1, we determined that it was the speaker's claim based on words such as "べきだと思う(I think it should be ...)". In the example utterance on row of "Claim : Other" of the Table 1, we labeled the statement "Claim : Other" because it is not the speaker's claim, but it reads as a development that the speaker will now reject this proposal.

When a monetary expression is part of another noun or not a monetary unit, as in the example on row of "It is not a monetary expression" of the Table 1, it is labeled "It is not a monetary expression". Monetary expressions that appeared as idiomatic expressions were labeled "Other", as in the example on row of "Other" of the Table 1.

Thus, in our annotations, ArgumentClass determined the discussion labels based on the content of the utterances in the minutes. Therefore, we thought it would be better to design the system's methodology so that the system automatically determines the content of the minutes and label the monetary expressions.

Table 1: Example of annotation(ArgumentClass)

argument label	example
Premise : Past and Decisions	法定外繰入金のうち保険料負担緩和などの決算補填等を目的とするものが赤字対象とされておりまして、平成30年度決算では約17億円でございます。
Premise : Current and Future / Estimates	今回の補正規模は、一般会計135億5,628万円の追加、特別会計28億2,111万円の追加、企業会計３億5,916万円の追加、合計167億3,656万円の追加となっております。
Premise : Other	本市の保険料は、年所得233万円の３人家族で42万7,100円という所得の２割近い保険料に対し
Claim : Opinions, suggestions, and questions	21億円を活用して市民の切実な願いである保険料の大幅な引下げを図るべきだと思いますが、最後に高島市長の答弁を求めて、私の質問を終わります。
Claim : Other	福祉の分野では、国民健康保険については21億円もの剰余金を保険料引下げに充てることなく、全額基金に積み立てる一方で、３万2,000筆を超える署名を無視して、１人当たりの保険料を2,000円、介護分を含めると4,300円余も引き上げようとしております。
It is not a monetary expression	そこで、地下鉄無料パスの福祉乗車証は、制度の変更に伴い、
Other	もしも１円でも安く入札されればそちらがとれたわけであります。

2.2 RelatedID

We searched for relevant budget items using words from the minutes that might be keywords. As before, we will use the minutes of the first regular meeting of Fukuoka City in 31st year of the Heisei period as an example.

In the case of the example in Table 2, we were able to determine that the word “保育(Childcare)” was the keyword. In the “description” of the budget item, we searched for this word as a keyword and found the budget item “ID-2019-401307-00-000031” and were able to determine that this could be tied to it. In the case of the example in Table 3, we were able to determine that the words “高速道路(expressway)” and “道路整備(road improvement)” were keywords. Using these words, we were able to determine that the budget items “ID-2019-401307-00-000094” and “ID-2019-401307-00-000096” could be linked from the “description” and “budgetItem” of the budget item. We were also able to determine from the “categories” and “departments” that these budget items could be tied together.

If a keyword could not be found, it was searched from the preceding and following sentences in the utterance. We also tied the same budget items to all the monetary expressions that had the same meaning in the minutes. The “8,823億円(882.3 billion yen)” in Table 3 appeared several times in the minutes, all of them with the same budget line item.

Table 2: Example of annotation(Cases with one RelatedID)

utterances	保育士の処遇につきましては、保育所の運営に通常必要とする費用として、国が定めます公定価格の引き上げにより、平成25年度から29年度の５年間で、月額約３万5,000円の改善に加え、技能、経験に応じた月額最大４万円の追加的な改善が行われております。
keyword	保育
RelatedID	ID-2020-401307-00-000031(幼児教育・保育の充実)

Table 3: Example of annotation(Cases with multiple RelatedID)

utterances	福岡市道福岡高速１号線から人工島延伸に係る６号線の項目について定めている福岡高速道路整備計画の中の新設または改築に要する費用の概算額は、現在8,823億円となっています。
keyword	高速道路,道路整備
RelatedID	ID-2019-401307-00-000094(道路橋りょう整備) ID-2019-401307-00-000096(都市計画道路整備)

When searching based on keywords, it was determined that the budget item “description” and “budgetItem” would be necessary in finding the relevant budget item. From this, we thought it would be a good idea to extract feature words from the utterances in the minutes, find those words in the budget items, and link them to RelatedID.

3 PROPOSED METHOD

The system’s methodology was based on our experience with annotation. Thereby, we thought it would be better to assign ArgumentClass and RelatedID using separate methods.

3.1 ArgumentClass

We will use BERT to assign argument labels to ArgumentClass. BERT is a model proposed by Google in 2018 that outperforms existing models on a variety of language tasks [1].

We extract the utterances with monetary expressions from the utterances in each the minutes, analyze them, and assign discussion labels to ArgumentClasses. We divide each utterance using “、(comma)”, “。(period)” and space as separators in the minutes, as shown in Table 4, and define a “divided-sentence” as one in which a monetary expression appears. If a “divided-sentence” is too short (15 characters or less), it should be combined with the preceding

Table 4: Example of splitting sentence for ArgumentClass

Before division	これより本日の会議を開きます。会議録署名議員に篠原達也議員、倉元達朗議員を指名いたします。日程に入るに先立ち、この際、報告いたします。まず、市長から別紙報告書類一覧表に記載の書類が提出されましたので、その写しを去る２月８日お手元に送付いたしておきました。次に、監査委員から監査報告第１号及び第２号が提出されましたので、その写しをお手元に送付いたしておきました。次に、地方自治法第100条第13項及び会議規則第125条第２項の規定により、お手元に配付いたしております議員派遣報告一覧表のとおり議長において議員の派遣を決定いたしておきました。以上で報告を終わります。
After division	これより本日の会議を開きます
	会議録署名議員に篠原達也議員
	倉元達朗議員を指名いたします
	日程に入るに先立ち、この際、報告いたします
	まず、市長から別紙報告書類一覧表に記載の書類が提出されましたので
	その写しを去る２月８日お手元に送付いたしておきました
	次に、監査委員から監査報告第１号及び第２号が提出されましたので
	その写しをお手元に送付いたしておきました
	次に、地方自治法第100条第13項及び会議規則第125条第２項の規定により
	お手元に配付いたしております議員派遣報告一覧表のとおり議長において議員の派遣を決定いたしておきました
	以上で報告を終わります

"divided-sentence" to make it new "divided-sentence".

3.1.1 About the model.

Our model is based on the argument labels as labels for document classification, fine-tuned by BERT. The training data will be used to create our model. We use argument label attached to each the monetary expression as the label for the corresponding "divided-sentence". We used a Japanese language model for BERT pre-trained called "cl-tohoku/ bert-base-japanese-whole-word-masking[1]" created by the Inui-Suzuki Laboratory at Tohoku University. To implement the BERT-related part of the proposed method, we used "BeltForSequenceClassification", a class for document classification in Transformers, an open-source library developed by Hugging Face. Our model was created with an epoch of 10, a batch size of 20, and a learning rate of 10^{-5} and maximum input length of 430.

[1]https://github.com/cl-tohoku/bert-japanese

By fine-tuning against this pre-trained model, we have created our model that classifies into seven argument labels.

3.1.2 About the "giin-flag".

Our model described in the previous section is a classification based only on the content of the utterances. However, when we actually read the minutes and assigned ArgumentClass to them, we sometimes judged the label based on the speaker's position as well as the content of the utterance. In addition, the ratio of labels sometimes changed depending on the speaker's position.

For example, in the first regular meeting of Fukuoka City in Heisei 31, special positions such as mayor and director were not labeled with Claim. Therefore, we want to add a dimension called the "giin-flag", which determines whether the speaker is a legislator or not, in order to find Claims. In the local assembly, we determined that a person whose "speakerPosition" was "議員(legislator)", and in the National Diet, a person whose "speakerPosition" was "NULL" and whose speaker was not the chairperson of the committee.

We combined the "giin-flag" with the dimensions of the output vector of the final layer of BERT corresponding to the CLS tokens. We then created a different model than our model described earlier.

3.1.3 Granting method.

Before we used our model, we went through the process of looking at the pre-extracted monetary expressions and labeling those that were not monetary expressions as "It is not a monetary expression". The term "not monetary expressions" here refers to monetary expressions with units that are not in "円(yen)", such as "kw" and "TEU", or that do not have a circle at the end, such as "ワークライフバランス(work-life balance)". However, "無料(free)" was excluded.

We then used our model created for the "divided-sentence" corresponding to each the money expression to perform inference and assigned argument labels to ArgumentClass. However, if there is more than one monetary expression in a "divided-sentence", all monetary expressions that appeared in the "divided-sentence" were given the inferred label for that "divided-sentence".

3.2 RelatedID

We assign relevant budget items to RelatedID by keyword extraction using TFIDF. We used "genism[2]" library to calculate the TFIDF. TFIDF is used to extract keywords in units of sentences that further divide the utterance in the minutes, as shown in the red frame in Figure 1.

We divided the utterances in the minutes using "。(period)" and space as separators, as shown in Table 1. If a sentence is too short (less than 200 characters), it should be combined with the subsequent sentences to make it a single sentence. We used "sudachipy[3]" for the word segmentation. The tokenizer used the C mode. The candidate words for keywords were nouns that appeared in the minutes, were not numerals, and had at least three letters.

The method of assignment is to calculate TFIDF for all utterance sentences in the units described above. The words with the highest TFIDF value in the sentence corresponding to each monetary expression are then sequentially examined for their presence in the

[2]https://radimrehurek.com/gensim/
[3]https://github.com/WorksApplications/SudachiPy

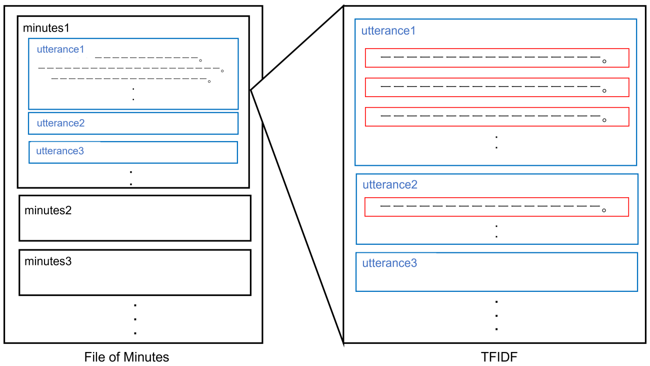

Figure 1: Scope of TFIDF

"budgetItem" and "description" of the budget item for that meeting, as shown in Figure 2. The budget items to look for are looked at from the youngest number in the "budgetId". If the keyword exists in a budget item, assign the "budgetId" to the monetary expressions as the related budget item.

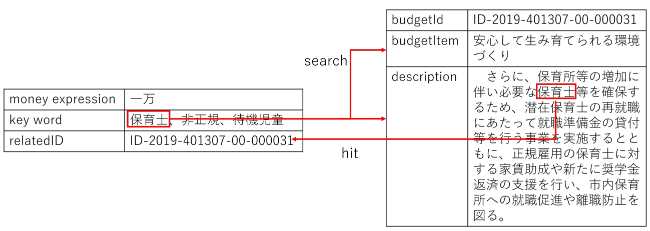

Figure 2: How to search for RelatedID

4 RESULTS

We will evaluate each of the test and training data for this task. We will also evaluate the results with and without the "giin-flag".

4.1 Calculation Method

The evaluation index is the score shown in the following equation, which is described in the overview paper of Budget Argument Mining subtask [2].

$$\text{Score} = \frac{1}{|S_{RID}|}\sum_{x,y}\{ACC(x,y) \times RIDC(x,y)\}$$

x and y mean the same monetary expressions of input and the gold standard data respectively. S_{RID} means a set of monetary expressions in the gold standard data whose RIDs is not null, as shown in the following equation.

$$S_{RID} = \{y|y.RIDs \neq null\}$$

ACC means whether an AC of monetary expressions is correct or not, as shown in the following equation.

$$ACC(x,y) = \begin{cases} 0 & (x.AC \neq y.AC) \\ 1 & (x.AC = y.AC) \end{cases}$$

Table 5: Example of splitting sentence for RelatedID

Before division	これより本日の会議を開きます。会議録署名議員に篠原達也議員、倉元達朗議員を指名いたします。日程に入るに先立ち、この際、報告いたします。まず、市長から別紙報告書類一覧表に記載の書類が提出されましたので、その写しを去る２月８日お手元に送付いたしておきました。次に、監査委員から監査報告第１号及び第２号が提出されましたので、その写しをお手元に送付いたしておきました。次に、地方自治法第100条第13項及び会議規則第125条第２項の規定により、お手元に配付いたしております議員派遣報告一覧表のとおり議長において議員の派遣を決定いたしておきました。以上で報告を終わります。
After division	これより本日の会議を開きます。会議録署名議員に篠原達也議員、倉元達朗議員を指名いたします
	日程に入るに先立ち、この際、報告いたします。まず、市長から別紙報告書類一覧表に記載の書類が提出されましたので、その写しを去る２月８日お手元に送付いたしておきました
	次に、監査委員から監査報告第１号及び第２号が提出されましたので、その写しをお手元に送付いたしておきました。次に、地方自治法第100条第13項及び会議規則第125条第２項の規定により、お手元に配付いたしております議員派遣報告一覧表のとおり議長において議員の派遣を決定いたしておきました
	以上で報告を終わります

$RIDC$ means whether an input RID is included in the RIDs of the gold standard data or not.

$$RIDC(x,y) = \begin{cases} 0 & (x.RID \notin y.RIDs) \\ 1 & (x.RID \in y.RIDs) \end{cases}$$

4.2 What the results show

The results for the case without the "giin-flag" in the test data are shown in Table 6, and the case with the "giin-flag" is shown in Table 7. The results for the training data without the "giin-flag" are shown in Table 8, and the results for the data with the "giin-flag" are shown in Table 9.

We made comparisons between results with and without the addition of the "giin-flag".

In the test data, the score was 23.40% for "all" both with and without the "giin-flag", with no difference. In all cases, only a difference of less than 5% points could be confirmed in the percentage of

Table 6: results of test data (no giin-flag)

	local	diet	all
score	23.91%	0%	23.40%
ACC	59.56%	38.46%	56.92%
RIDC	34.78%	0%	34.04%

Table 7: results of test data (giin-flag)

	local	diet	all
score	23.91%	0%	23.40%
ACC	55.38%	41.54%	53.65%
RIDC	34.78%	0%	34.04%

Table 8: results of training data (no giin-flag)

	local	diet	all
score	21.82%	72.73%	23.43%
ACC	87.90%	87.27%	87.82%
RIDC	24.19%	73.73%	25.71%

Table 9: results of training data (giin-flag)

	local	diet	all
score	21.24%	72.73%	22.86%
ACC	88.00%	89.09%	88.14%
RIDC	24.19%	73.73%	25.71%

correct answers in the ArgumentClass, and both “local” and “diet” did not work.

In the training data, the score is higher when the “giin-flag” is not present, but the difference is only 0.57% points for “all”. In terms of the percentage of correct answers for ArgumentClass, the correct answer was higher with the “giin-flag” in all cases of “local,” “diet,” and “all,” but by less than 2% points in all cases.

5 CONSIDERATION

We will consider ArgumentClass and RelatedID separately.

5.1 ArgumentClass

We used the “giin-flag” to improve the accuracy of ArgumentClass. As noted in Section 4, there was little difference in the percentage of correct answers with and without the “giin-flag”. When we examined the post-experimental test data, we found that models with the “giin-flag” slightly increased the number of Claims attached. However, there is no difference in the number of correct answers in the Claim, which is not in line with the intention of being able to correctly extract the Claim from the speaker.

We looked at monetary expressions that have different correct and incorrect answers depending on the presence or absence of a “giin-flag”. Examining the cases in which the correct answer was given only when the “giin-flag” was not present, we found that many of the cases with the “giin-flag” made mistakes by outputting

Table 10: Correct only if there is no “giin-flag”

speaker	議員
utterances	これらの訪日外国人旅行者の2018年の消費額が４兆5,064億円
monetary expression	４兆5,064億円
giin-flag (wrong)	Premise : Past and Decisions
no giin-flag (correct)	Premise : Current and Future / Estimates

Table 11: Correct only if there is “giin-flag”

speaker	議員
utterances	平成31年度の当初一般会計予算総額は8,666億円となりました
monetary expression	8,666億円
giin-flag (correct)	Premise : Current and Future / Estimates
no giin-flag (wrong)	Premise : Past and Decisions

“Premise : Current and Future / Estimates”, as shown in Table 10. When we examined the cases in which the correct answer was given only when the “giin-flag” was present, the opposite of the aforementioned case was found, where the output “Premise : Current and Future / Estimates” was often the correct answer, as shown in Table 11. The most common training data used to create the model was “Premise : Current and Future / Estimates”. We attributed this result to the fact that models with congressional flags were more affected by this effect.

5.2 RelatedID

RelatedID was not able to achieve high accuracy in both training and test data. Based on the results, we thought that the low score was due to the low accuracy of RelatedID. We believe that there are two main reasons for the failure to achieve high accuracy with RelatedID.

The first reason is that the words split by word segmentation may not appear in the budget items. As an example, words such as “高速道路(expressway)” and “人工島(Artificial island)”, which appear in the minutes of the first regular meeting of Fukuoka City in Reiwa 2, do not appear in the budget for that year. Our system looks for the extracted words directly from the budget items, so we cannot get to the correct budget item.

The second reason is that the TFIDF values of words we can determine to be keywords are not high enough or words cannot be extracted. As an example, words that are commonly used in budget meetings, such as “事業費(project cost)” and “給付金(payment)”, are extracted as words with high TFIDF values. In addition, the TFIDF values for words that we can determine to be keywords, such as “雇用調整助成金(employment adjustment subsidy)”, are sometimes the lowest. There are times when a word that we can see as a keyword is not in the sentence in which the monetary expression appears. As a result, the correct keywords are not extracted, and unsuitable budget items are linked.

5.2.1 Solution.

As a solution to the first cause, we believe that linking synonyms will improve accuracy. As mentioned above, there is no word "高速道路(expressway)" in the budget item. However, words like "幹線道路(main road)" and "道路の整備(Road maintenance)" do appear. Therefore, we believe that if we can link them as the same meaning, we can link the budget items that were not linked in this study.

As a solution to the second cause, we believe that dividing utterances into categories will improve accuracy. The meetings of the Congress take the form of all-at-once questions and answers system. It is characterized by the fact that a single utterance may contain several agenda items, and that people rarely talk about the next agenda item until they have finished talking about one. Therefore, we believe that categorizing utterances by ministry, department, or office will increase the scope of the TFIDF and increase the TFIDF value of words that we can determine to be feature words. In our method, the budget items were searched in order of the youngest budgetId number, so if the TFIDF value of a word that was not a keyword was high, it was easy to link unrelated budget items. As shown in the example in Section 2, budget items may also be determined from "categories" and "departments". Therefore, we believe that if the categories of statements are known, it is easier to find the correct budget item, as many budget items can be narrowed down by category.

6 CONCLUSIONS

We proposed a BERT-based method for ArgumentClass and a TFIDF-based keyword extraction method for RelatedID.

The method using BERT gave relatively good results. In addition, a comparison was made between those with and without the added dimension of the "giin-flag", which is attached to determine whether a person is a legislator or not. Although there was a difference in results with and without the "giin-flag", the difference was small and no advantage was found in either case.

The method using TFIDF showed sluggish accuracy and low results. There are two reasons for this: the words that we split by word segmentation may not appear in the budget items, and the TFIDF value of the word we can determine to be a keyword is low or cannot be extracted. We believe that these causes can be solved by treating synonyms as the same word and separating utterances by agenda.

ACKNOWLEDGMENTS

This work was supported by JSPS KAKENHI Grant Numbers JP-21H03769, JP20K00576.

REFERENCES

[1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT : Pre-training of Deep Bidirectional Transformers for Language Understanding. *NAACL-HILT* (2019).

[2] 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).