# OUC at the NTCIR-16 QA Lab-PoliInfo-3 Budget Argument Mining

OUC team

Keiyu Nagafuchi[1], Rin Sasaki[2], Seiya Oki[2],
Yasutomo Kimura[2] and Kenji Araki[3]

[1] Graduate School of Information Science and Technology, Hokkaido University, Japan
[2] Otaru University of Commerce, Japan
[3] Faculty of Information Science and Technology, Hokkaido University, Japan

# 1. Our methods

The Budget Argument Mining task consists of two subtasks,
including Argument Classification (AC) and linking relatedIDs (RID).

We separately proposed several methods to perform AC or RID, and combined them.

## Budget Argument Mining

**AC**
**Argument Classification**

This subtask could be considered a simple sentence classification. We trained several classifiers with pairs of utterances containing monetary expressions and discussion labels.

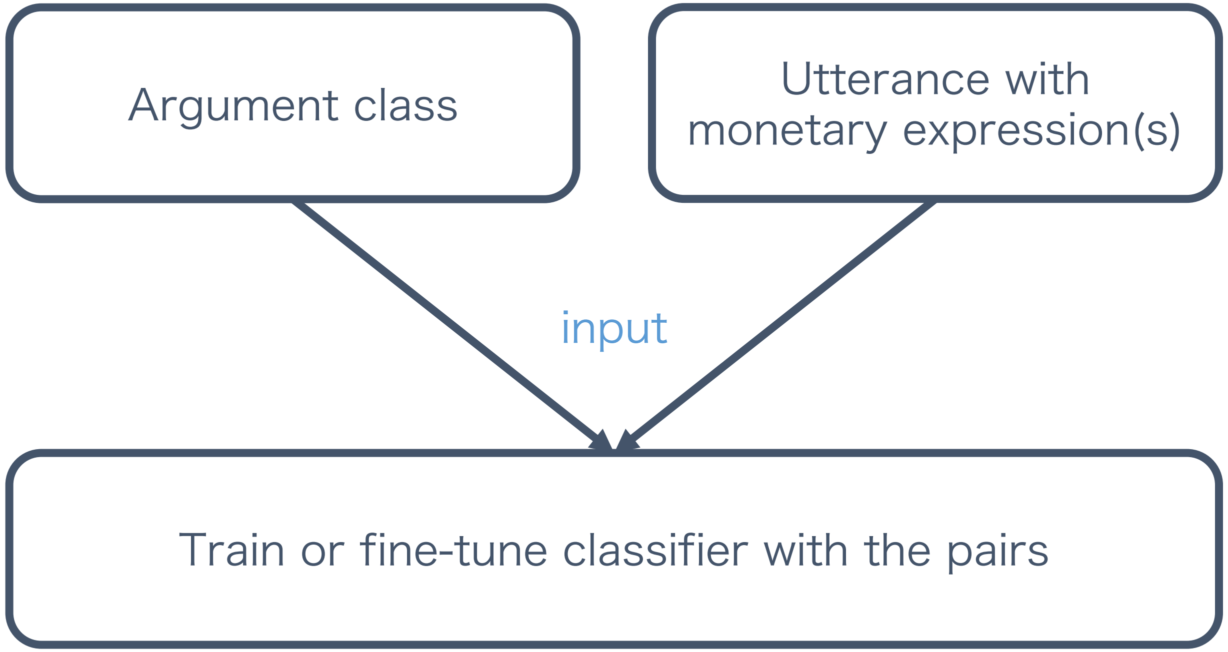
**RID**
**Linking relatedIDs**

This subtask could be considered a search for similar sentences. We calculated the cosine similarity between budget item descriptions and utterances containing monetary expressions with sentence vectors.

# 1.1. Our methods: AC

## Rule-based classifier

- If a specific keyword is included in a sentence, a corresponding argument class is selected.




## BoW-based classifier

- Calculated sentence vectors with Bag of Words or TF-IDF.
- Trained with scikit-learn[1]'s algorithms.

| Method name | Vectorizer | Tokenizer | Classifier |
|---|---|---|---|
| BoW_LSVC | BoW | MeCab IPADIC | LinearSVC |
| BoW_noun_LSVC | BoW(noun) | MeCab IPADIC | LinearSVC |
| TFIDF_LSVC | TF-IDF | MeCab IPADIC | LinearSVC |
| TFIDF_Sudachi_LSVC | TF-IDF | Sudachi Mode B | LinearSVC |
| BoW_SVC | BoW | MeCab IPADIC | SVC |
| BoW_RF | BoW | MeCab IPADIC | RandomForest |
| BoW_SGD | BoW | MeCab IPADIC | SGD |
| BoW_Ensenble | BoW | MeCab IPADIC | Ensenble |

## BERT[1] classifier

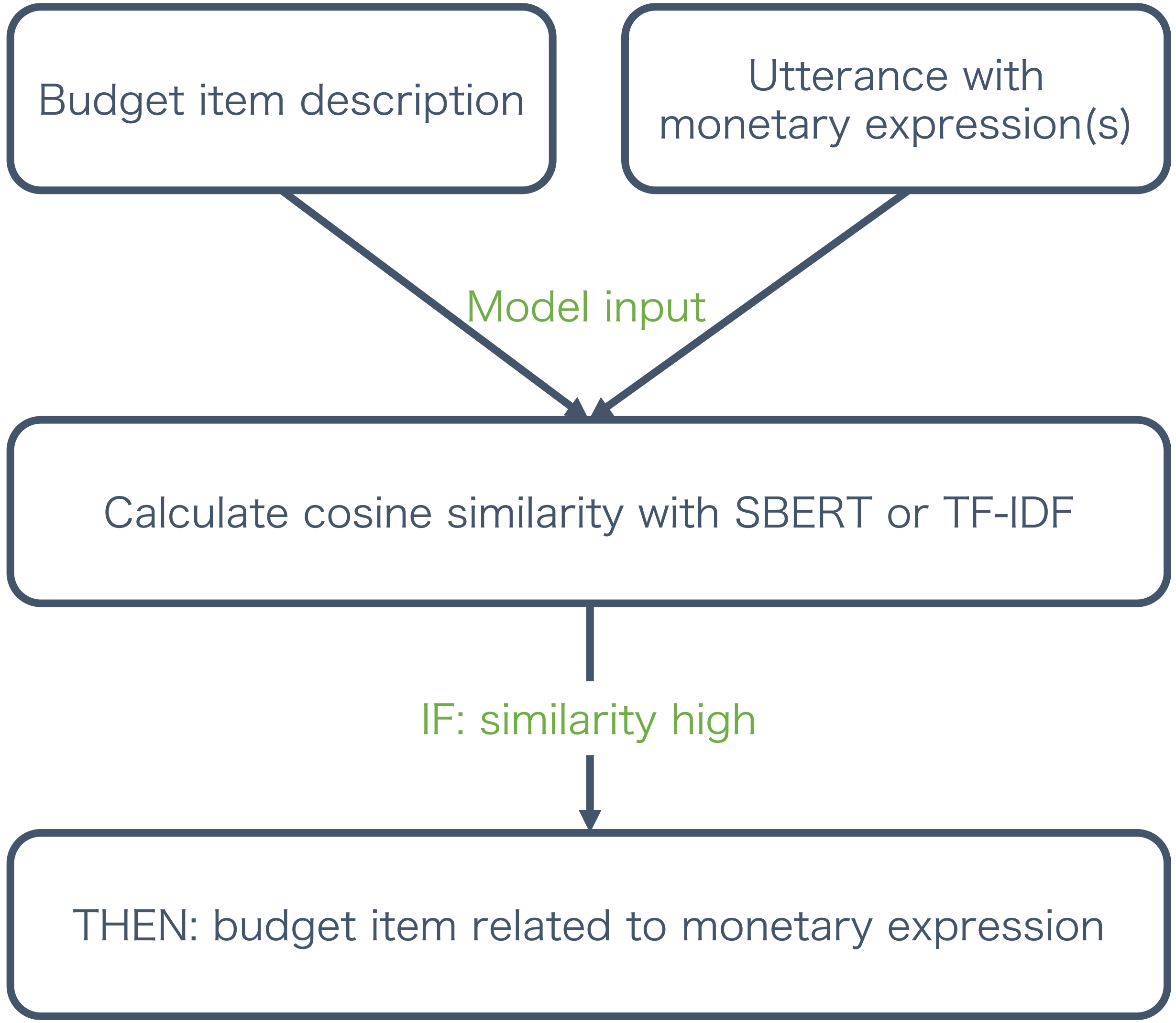
- Fine-tuned with Japanese BERT[2] on argument classification.

| Method name | Base model |
|---|---|
| BERT_base | bert-base-japanese-whole-word-masking |
| BERT_base_v2 | bert-base-japanese-v2 |
| BERT_large | bert-large-japanese |
| BERT_base_ml64 | bert-base-japanese-whole-word-masking |

1 https://scikit-learn.org/stable/
2 https://huggingface.co/cl-tohoku/

[1] Devlin et al. BERT: Pre-training of deep bidirectional transformers for language understanding. NAACL 2019.

# 1.2. Our methods: RID




## Sentence-BERT[2] (SBERT)

- Fine-tuned with Japanese BERT[2] on NLI and STS tasks.

## TF-IDF

- We tried several word segmentation modes by Sudachi[3].
  - Sudachi has three word segmentation modes: A (short), B (medium), and C (long).

[2] https://huggingface.co/cl-tohoku/

[2] Reimers et al. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. EMNLP 2019.
[3] Takaoka et al. Sudachi: a Japanese Tokenizer for Business. LREC 2018

# 2. Our results

## Overall (AC + RID) score

- Among our methods, ID300 obtained the highest score (3rd place on the leaderboard).

## AC score

- The BERT_base classifier obtained the highest score (2nd place on the leaderboard).
- BERT classifiers showed higher scores than rule-based and BoW-based ones.

## RID score

- The TFIDF_modeA model obtained the highest score (1st place on the leaderboard).
- TF-IDF models showed higher scores than SBERT ones.
- TF-IDF models with shorter word segmentation performed better.

| ID | Method name (AC+RID) | Score | AC | RID |
|---|---|---|---|---|
| 300 | BERT_base + TFIDF_modeA | **0.4468** | **0.5712** | **0.6596** |
| 309 | BERT_base_ml64 + TFIDF_modeA | 0.4255 | 0.5385 | **0.6596** |
| 263 | BoW_SVC + TFIDF_modeA | 0.4255 | 0.4827 | **0.6596** |
| 308 | BERT_large + TFIDF_modeA | 0.4043 | 0.5615 | **0.6596** |
| 305 | BERT_base_v2 + TFIDF_modeA | 0.4043 | 0.5577 | **0.6596** |
| 251 | BoW_LSVC + TFIDF_modeA | 0.4043 | 0.4904 | **0.6596** |
| 252 | TFIDF_Sudachi_LSVC + TFIDF_modeA | 0.4043 | 0.4885 | **0.6596** |
| 250 | BoW_LSVC + TFIDF_modeB | 0.4043 | 0.4904 | 0.5745 |
| 301 | BoW_Ensenble + TFIDF_modeA | 0.3830 | 0.4750 | **0.6596** |
| 277 | BoW_RF + TFIDF_modeA | 0.3830 | 0.4231 | **0.6596** |
| 248 | BoW_LSVC + TFIDF_modeC | 0.3830 | 0.4904 | 0.5532 |
| 278 | BoW_SGD + TFIDF_modeA | 0.2979 | 0.4615 | **0.6596** |
| 230 | BoW_LSVC + SBERT_NLI | 0.1489 | 0.4904 | 0.1702 |
| 177 | Rulebased + SBERT_NLI (dry run ver.) | 0.1277 | 0.3731 | 0.2128 |
| 234 | TFIDF_LSVC + SBERT_NLI | 0.0851 | 0.4750 | 0.1702 |
| 233 | BoW_noun_LSVC + SBERT_NLI | 0.0851 | 0.4231 | 0.1702 |
| 219 | Rulebased + SBERT_NLI | 0.0851 | 0.3731 | 0.1702 |
| 211 | Rulebased + SBERT_NLI | 0.0851 | 0.3731 | 0.1702 |
| 212 | Rulebased + SBERT_STS | 0.0851 | 0.3731 | 0.1489 |
| 183 | Rulebased + Doc2Vec | 0.0000 | 0.3731 | 0.1277 |
| 217 | Rulebased + miss | 0.0000 | 0.3731 | 0.0000 |

# 3.1 Discussion: AC

- BERT classifiers obtained higher scores than rule-based and BoW-based ones.
  → Considering contexts was effective for this subtask.
- We counted the number of misclassifications for all our methods.
  - Misclassification rates of "Premise" classes were low, but the rates of other classes were high.
    → Training and fine-tuning did not go well because datasets of this task were imbalanced.

| Argument class | Number of misclassifications for all our methods | Number of classes in GS data | Misclassification rate |
|---|---|---|---|
| Premise : Past and Decisions | 26 | 101 | 0.2574 |
| Premise : Current and Future / Estimates | 0 | 196 | 0.0000 |
| Premise : Other | 19 | 145 | 0.1310 |
| Claim : Opinions, suggestions, and questions | 25 | 42 | 0.5952 |
| Claim : Other | 4 | 4 | 1.0000 |
| It is not a monetary expression | 23 | 30 | 0.7667 |
| Other | 2 | 2 | 1.0000 |

# 3.2 Discussion: RID

- It is likely that poor results of SBERT were attributed to the fact that budget item descriptions were often omitted in utterances containing monetary expressions.
- Most utterances contained keywords related to budget items in preceding and following contexts of monetary expressions.
  → It might lead to that the TF-IDF obtained good results in linking RID.
- Utterances that were answered incorrectly with TF-IDF did not contained keywords.
  → Keywords were included in the preceding and following sentences.
- In the future, we should consider a system that also considers the surrounding sentences.

| Utterance | Related budget item |
|---|---|
| 例えば、雇用調整助成金の**一万五千円**への上限引上げや家賃支援給付金、学生支援給付金の創設などは、問題点はあるものの、賛成できるものです。 | 雇用調整助成金の抜本的拡充 |
| For example, raising the ceiling on employment adjustment subsidy to **15,000 yen** and establishing rent support benefits and student support benefits are all agreeable, although there are some problems. | Fundamental expansion of employment adjustment subsidy |

# 4. Conclusion

- We separately proposed several methods
  to perform AC or RID and combined them.
- Among our methods, the combination of
  BERT_base classifier and TF-IDF_modeA model
  obtaind the highest score (0.4468).
  - This method got 3rd place on the leaderboard of overall score (0.5712).
  - BERT_base classifier got 2nd place on the leaderboard of AC score (0.6596).
  - TF-IDF_modeA got 1st place on the leaderboard of RID score.
- Because only one utterance sentence was used as input for our systems in this work, it is necessary to develop a system that could consider the surrounding context in the future.