A Cascade Model for Argument Mining in Japanese Political Discussions

Ramon Ruiz-Dolz
Universitat Politècnica de València

VRAIN
Valencian Research Institute for Artificial Intelligence

Budget Argument Mining

This work approaches the Budget Argument Mining (BAM) instance of the argument mining task. The BAM is aimed at improving the automatic argumentative analysis of political discussion transcripts through the use of NLP techniques. It includes the argumentative discourse segmentation and the argument component detection sub-tasks. For that purpose, monetary expressions are detected in the transcripts, and it must be determined if an expression belongs to an argument or not, and which is its argumentative role in the discussion. Furthermore, the required analysis is enriched with the relation of each monetary expression with a political budget item. This way, the resulting analysis will provide a set of argumentative components and their type detected in the transcripts of a discussion, and a set of relations between the arguments and budget items.

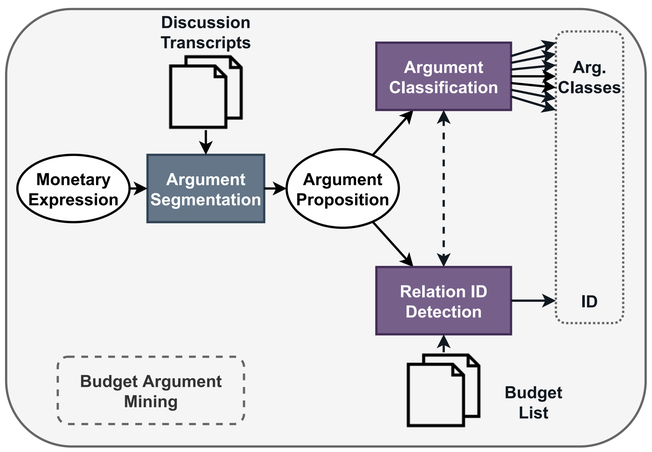

Fig. 1: Budget Argument Mining task diagram.

Therefore, the BAM consists of two different sub-tasks: the argument classification (AC) and the relation ID detection (RID) (see Figure 1).

Model Description

We propose a BERT-based cascade model to undertake the complete BAM process (see Figure 1). All the BERT-based classifiers integrated in our cascade model were fine-tuned from the *Inui Laboratory* (`https://github.com/cl-tohoku/bert-japanese/tree/v2.0`) pre-trained BERT-large Japanese Language Model. In our approach, each monetary expression is be treated as the input, and a class label and a related ID as the output. Thus, our model does the **argument segmentation** after identifying the monetary expressions in the poltical discussion transcripts. Then, **each argumentative proposition is classified, and related** to a specific item from the budget list. The proposed architecture aims to smooth the complexity of the classification task considering the size of the training corpus and the number of classes. Furthermore, it approaches independently the AC and the RID sub-tasks. Figure 2 synthesises the proposed architecture. The code implementation of the model architecture proposed in this paper is publicly available in GitHub: `https://github.com/raruidol/Budget-AM`.

Results

The *rVRAIN* achieved the 4th best performing position from a total of 6 participating teams. However, our approach was the best performing one from the teams that did not include task organisers.

Team	Score AC+RID	AC	RID
*JRIRD**	**0.51**	**0.58**	0.61
*OUC**	0.45	0.57	**0.66**
*fuys**	0.23	0.57	0.34
rVRAIN	0.17	0.48	0.21
rVRAIN (5BERT)	0.06	0.48	0.21
takelab	0.04	0.39	0.06
SMLAB	0.00	0.38	0.00
RB	0.00	0.13	0.00

Table 1: Formal-run evaluation of the different models for BAM. (*)The team contains task organisers.

Model Architecture

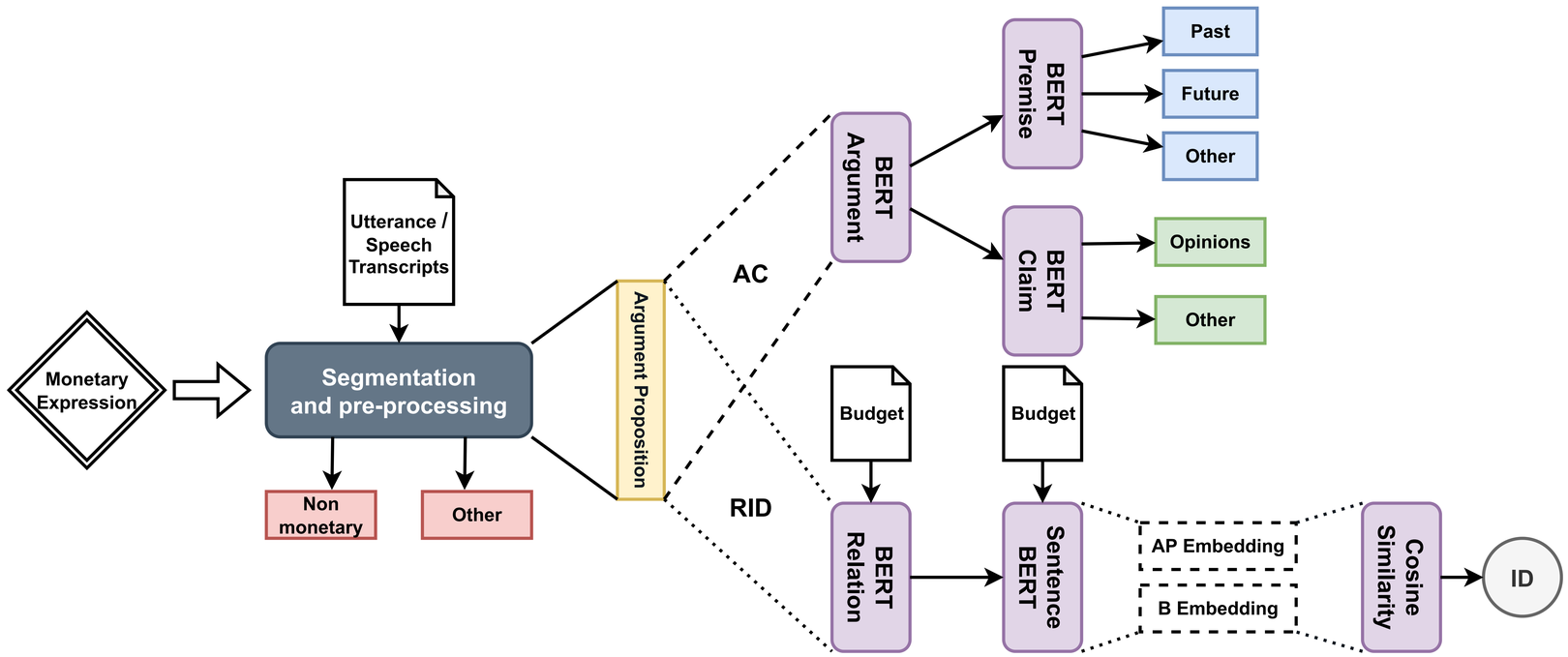

Fig. 2: *rVRAIN* model architecture.

Discussion

This poster represents the participation of *rVRAIN*'s team at the *Budget Argument Mining* task organised in the *QALab PoliInfo 3* and the *NTCIR-16*. The organisers proposed a new instance of the argument mining task, a classic in the NLP area of research. In this new instance, the main goal was to correctly classify arguments containing monetary expressions and relate them to items in a list of political budgets. In this paper, we have proposed a new approach to this task relying in the latest advances in NLP (i.e., Transformer-based architectures). The proposed cascade model architecture achieved the fourth position in the performance ranking, and it was the best among teams without task organisers.

- First, we have seen how when dealing with highly unbalanced corpora, a system can benefit from defining a set of handcrafted rules and relaxing the class complexity of the task. Instead of approaching the complete problem with the use of a unique classifier.
- Second, we have also observed that no improvement could be achieved by forcing the balance of the corpus. When using the balanced version, the score dropped significantly. This is most probably because the real distribution that the model has to predict is not balanced, but the corpus size limitation can also have a major role in this issue.

Acknowledgements

This work is supported by the Spanish Government project PID2020-113416RB-I00.