7C-01

The Significant Factors that Affect the Accuracy on Classifying English IBIS Datasets

Yihan Dong† Kyoto University†
Shiyao Ding‡ Kyoto University‡
Jawad Haqbeen§ Kyoto University§
Takayuki Ito¶ Kyoto University¶

1 Introduction

The issue-based information system (IBIS) was raised as an argumentation-based approach to solving complex problems [3]. The elements in IBIS can be divided into 4 types: *Issues* that need to be answered, *positions* towards issues, and arguments including *pros* and *cons* which are positive and negative respectively. To further promote the efficiency of IBIS, it is significant to classify the types of discussion contents by these labels, as a typical NLP problem.

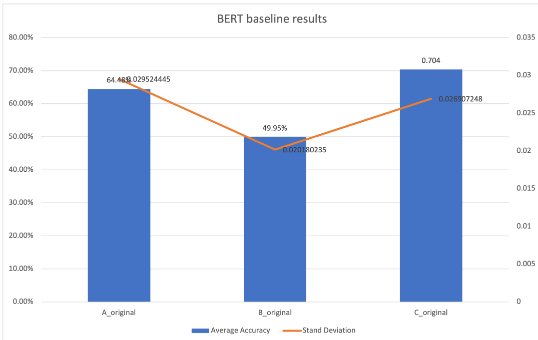
There are two classic and widely used models called BERT and fastText for tasks existing right now. BERT is an unsupervised, deep bidirectional system for pre-training NLP, resulting that it can be fine-tuned by just adding one additional output layer to be adjusted to some particular tasks [1]. Sakai et al. [4] combined it with Long Short-term Memory (LSTM) to classify the discussion contents in Japanese by IBIS labels with high accuracy. However, the experiments in English were not sufficient because of the lack of English datasets. Alternatively, fastText is a text classifier with similar performance compared with deep learning classifiers in terms of accuracy, but magnificently faster for training and evaluation [2].

The goal of this report is to investigate the key factors affecting the precision of the IBIS label classification by doing comparative experiments. According to the results, further detailed methods to improve or modify the process of IBIS label classification are raised at the end of the paper.

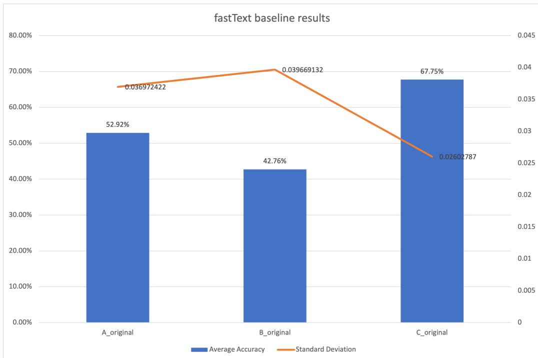
2 Experiments

The first two datasets were directly withdrawn from discussion recordings in English; By contrast, the third one was translated from a discussion recording in Japanese but strictly followed the IBIS structure as a comparison. 10-fold validation was taken as the verification method.

First and Foremost, the algorithm mentioned previously and the fastText classifier would be taken to classify the IBIS labels as the baseline results. The comparison between the baseline results would be illustrated in Figure 1.



(a) experiments results using BERT



(b) experiments results using fastText

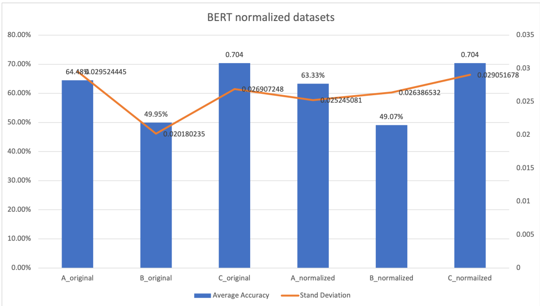
Figure1: baseline precision results

According to the baseline results, it is clear that the previous classification algorithm has overwhelming advantages in identifying the first two datasets, either the average precision or the stability of the identification. However, this advantage narrows when classifying the sentences in the third dataset, mainly because the third discussion recording strictly follows the IBIS structure.

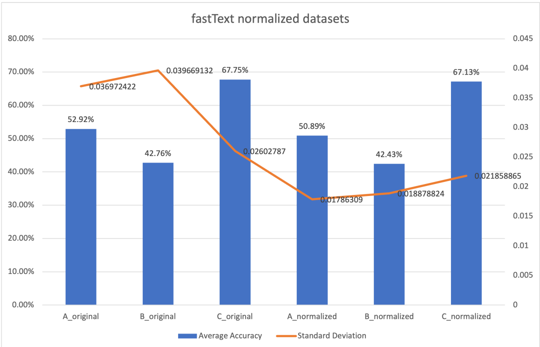
Furthermore, considering the major difference between English and Japanese, normalizing the

The Significant Factors that Affect the Accuracy on Classifying English IBIS Datasets
† Yihan Dong, Kyoto University
‡ Shiyao Ding, Kyoto University
§ Jawad Haqbeen, Kyoto University
¶ Takayuki Ito, Kyoto University

datasets by converting upper letters and punctuation is a common idea. In this experiment, normalized datasets were used instead of the original ones to figure out if they can improve the precision.



(a) experiments results using BERT



(b) experiments results using fastText

Figure2: comparative results generated by using normalized datasets

From the results given in Figure 2, the conclusion is that even though the previous normalization towards the datasets would not affect the precision of the label classification heavily, it still makes the classification more stable by decreasing the standard deviation of the results.

Last but not least, considering the sequence of words in sentences is another idea. It is relatively easy to test it by adding a parameter to the fastText classifier. Figure 3 illustrates how this factor affects precision.

As it illustrates, by considering the sequence of the words in sentences, the average precision of classification increased, but the increase was independent of the word number of a group.

3 Conclusion

From the results of the experiments, considering the sequence of words during training models in English could promote the precision of classification can be reached as a conclusion.

In addition, since normalizing sentences could decrease the fluctuation of the prediction without decreasing precision, it can be implemented as a process of the agent's behaviours in the future to improve stability.

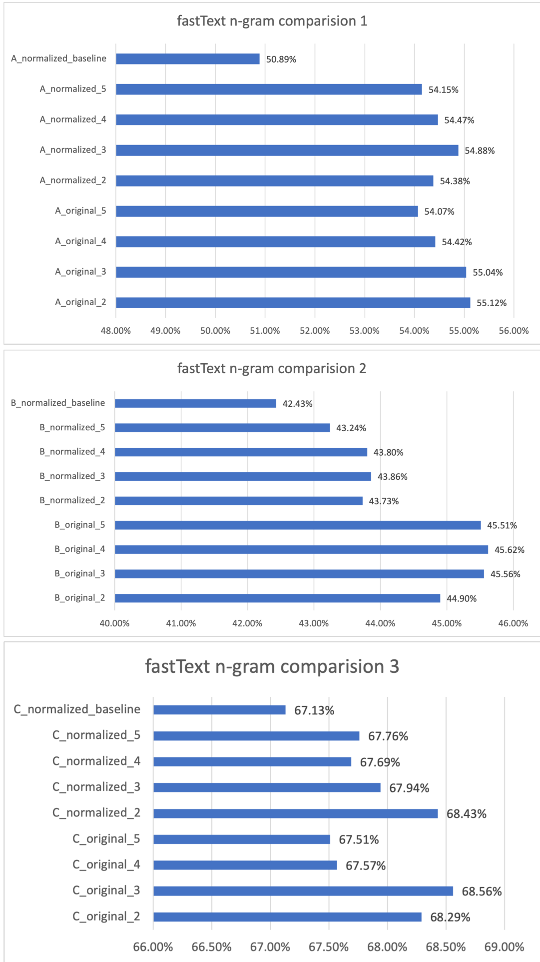


Figure3: comparative results using fastText, considering the sequence of words

Meanwhile, the results on the third dataset were the best even though it was translated from Japanese, which means that to some extent if native speakers attend the discussion does not affect the precision. It could make generating IBIS classifiers in different languages feasible.

References

[1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[2] Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. Bag of tricks for efficient text classification. *arXiv preprint arXiv:1607.01759*, 2016.

[3] Douglas Noble and Horst WJ Rittel. Issue-based information systems for design. 1988.

[4] 酒井敦也, 鈴木祥太, 佐藤拓実, 川村直輝, 伊藤孝行, et al. 議論における gated attention network を用いたノード分類. **人工知能学会論文誌**, 36(6):A–L24_1, 2021.