NTCIR 17 Conference: Proceedings of the 17th NTCIR Conference on Evaluation of Information Access Technologies, December 12-15, 2023, Tokyo, Japan
DOI: https://doi.org/10.20736/0002001276

IDEA at the NTCIR-17 FinArg-1 Task: Argument-based Sentiment Analysis

Shaopeng Tang
School of Computer and Artificial Intelligence,
Wuhan University of Technology
Wuhan, China
karitown@whut.edu.cn

Lin Li
School of Computer and Artificial Intelligence,
Wuhan University of Technology
Wuhan, China
cathylilin@whut.edu.cn

ABSTRACT

Although argument mining has been discussed for several years, financial argument mining is still in the early stage. The IDEA team participates in Argument Unit Identification (for Earnings Conference Call) and Argument Relation Identification (for Earnings Conference Call) subtasks of the NTCIR-17 FinArg-1 Task. This paper presents our work on the two subtasks. For Argument Unit Identification subtask, we successively construct the models based on BERT and Roberta to classify a given argumentative sentence. To better extract the semantic features, we combine the pre-trained model with CNN. Micro-F1 and Macro-F1 achieve 76.47% and 76.46% in official evaluation results of the first run (i.e., IDEA-1) , respectively, outperforming most approaches of other teams. For Argument Relation Identification subtask, we classify sentence pairs based on the pre-trained model and Prompt-Tuning. And Micro-F1 and Macro-F1 achieve 81.74% and 51.85% in official evaluation results of the third run (i.e., IDEA-3), respectively.

KEYWORDS

Sentiment Analysis, Argument Mining, Pre-trained Model, Prompt-Tuning

TEAM NAME

IDEA

SUBTASKS

Argument Unit Identification (English)
Argument Relation Identification (English)

1 INTRODUCTION

Although argument mining has been discussed for several years, financial argument mining is still in the early stage. In FinNum-3, the concept was proposed for identifying the arguments in financial narratives [3]. To perform a more fine-grained analysis, an argument-based sentiment analysis task is proposed in NTCIR-17 FinArg-1 Task [4].

The FinArg-1 Task has three sources of data: i) professional analysts' reports written in English, ii) earning conference calls transcribed in English and iii) social media writen in Chinese. Due to time constraints, our team only participated in two subtasks corresponding to the earning conference calls data [1], i.e., Argument Unit Identification and Argument Relation Identification.

The problem in the Argument Unit Identification subtask is a binary classification, i.e., classifying sentences from the Earnings Conference Call to either Premise or Claim. Premise and Claim are related to statements made by individuals or entities, but they have distinct meanings and implications. A premise involves a commitment to take specific actions in the future, while a claim is a statement asserting that something is true without necessarily implying any commitment to act. In finance field, the two have specific meanings. Premise tends to suggest the cause of financial development, while claim tends to indicate the effect of economic development. Traditional text classification approaches include SVM, Random Forest and Logistic Regression etc [2, 9, 10].

In recent years, pre-trained models have shone in various tasks. BERT and Roberta, as two well-known pre-trained models, once proposed, have achieved better results on various tasks [5, 8]. Yoon Kim et al. applied the convolutional neural network CNN to a text classification task, utilizing multiple kernels of different sizes to extract key information in a sentence, allowing for better capture of local relevance [6]. We combine the pre-trained model and CNN for this task to fully utilize the advantages of both.

The Argument Relation Identification subtask is to identify the relations (support/attack/none) of the given two sentences from the Earnings Conference Call. Simply concatenating two sentences together would reduce this task to a text classification task without taking into account the relationship between the sentences. The recently popular prompt learning can stimulate the potential knowledge of pre-trained models and can be more adaptive to downstream tasks [7]. Therefore we accomplish this task by designing a suitable prompt in the form of an MLM task.

The remainder of this paper is structured as follows. Section 2 describes how to combine BERT with CNN for the Argument Unit Identification subtask and how to use Prompt Learning for the Argument Relation Identification subtask. This section includes which model we applied, the CNN convolutional layer design, the design of Prompt, the experimental configuration. Section 3 provides the experimental results and the analysis of the results. Finally, Section 4 presents conclusions and future work.

2 PROPOSED APPROACHES

Considering that the datasets and forms of these two tasks are quite different, we designed different approaches for them to be accomplished, respectively. In particular, Section 2.1 is about the approach of the Argument Unit Identification task and Section 2.2 is about the approach of the Argument Relation Identification task.

2.1 Argument Unit Identification

The overview of our approach for this task is shown in Figure 1. BERT, as a pre-trained model, already contains a lot of prior knowledge, so we use BERT as part of our approach to provide a good

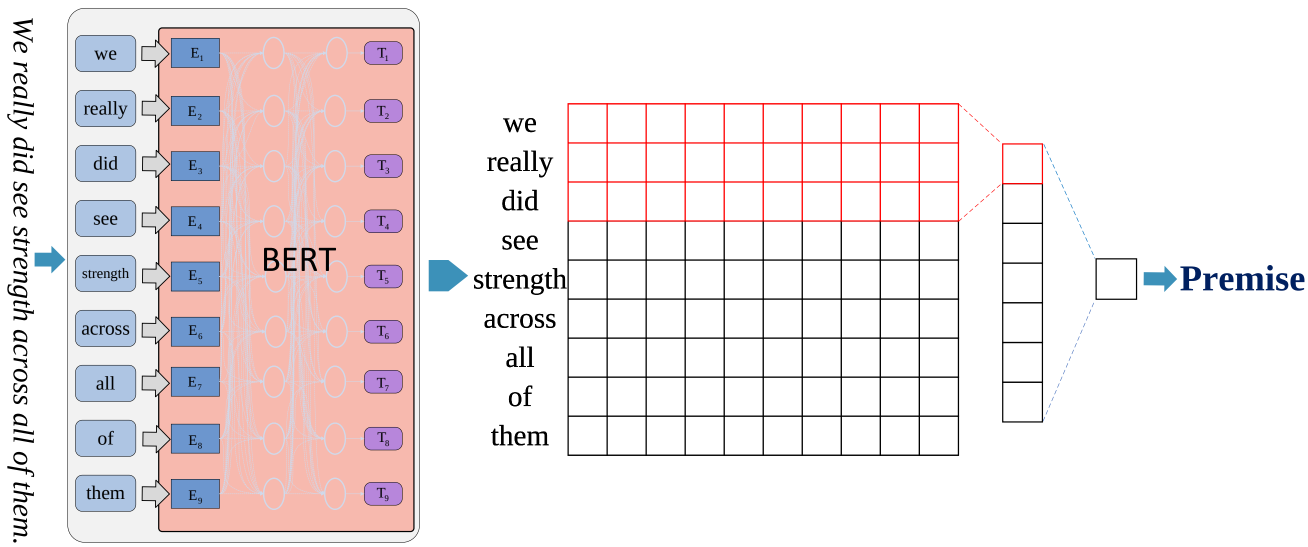

Figure 1: The Overview of Our approach for Argument Unit Identification

starting point in the IDEA-1 run. Specifically, the last_hidden_state embedding generated by BERT is used as the initial state of the CNN.

First, we perform a convolution operation on the embedding to extract key features. This process is as follows:

$$c = f\ (w_1, e, b_1) \tag{1}$$

where e is the last_hidden_state embedding generated by BERT, w_1 is the convolution kernel, b_1 is a bias term and f is the convolution operation. This filter is applied to each possible window of words in the sentence to produce a feature map c.

Similarly, a further convolution operation is performed based on the output of the last layer. The difference is that we add an activation function to regularize the convolution result of this layer. The operation is represented as follows:

$$output = activate(g(w_2, c, b_2)) \tag{2}$$

where w_2 is the convolution kernel, b_2 is a bias term, g is the convolution operation and *activate* is the **sigmoid** activation function. Finally, $output \in [0, 1]$.

We use the BERT uncased implementation from Huggingface [1]. The used parameters for our fine-tuned BERT model for Argument Unit Identification are stated in Table 1.

In addition, we compare the effectiveness of Roberta [2] for this task in the IDEA-3 run. We utilize Roberta to generate embedding, and then feed them into the CNN. The used parameters are likewise shown in Table 1.

2.2 Argument Relation Identification

BERT mainly uses MLM (Masked Language Model) tasks in the pre-training phase [5]. Therefore, in order to better adapt the downstream tasks to the pre-trained model to stimulate the potential knowledge of the pre-trained model, we design appropriate prompt for the pre-trained model in the form of the MLM task.

In the case of the Argument Relation Identification task, we conduct template engineering and answer engineering to select appropriate template and answer mapping space. The templates and answer spaces we used are shown in Table 2

We construct all the data according to the template, then encode them and feed them to BERT, which predicts the [MASK] positions. The Levenshtein Distance [11] between two strings refers to the minimum times of editing operations required to change one string into another, and can be used to measure the similarity between two strings. We map the predictions to the specific answer (support/attack/none) via Levenshtein Distance. The used parameters for Argument Relation Identification are stated in Table 3.

Table 1: The Parameters for The Model of Argument Unit Identification

Parameter	Value
Optimizer	AdamW
Batch Size	64
Learning Rate	1e-5
Max Length	128
Total Epoch	5
Loss Function	Cross Entropy

3 EXPERIMENTS & ANALYSIS

We submitted our predictions to the officials, being allowed to submit three runs per subtask. The official reports both Micro-F1 and Macro-F1 scores of all tasks, and use Macro-F1 to rank the

[1]https://huggingface.co/bert-base-uncased
[2]https://huggingface.co/roberta-base

Table 2: The Template and Answer for Argument Relation Identification

Input	Team-Run	Template	Answer([MASK])
S1, S2	IDEA-1	There are two sentences: S1 and S2, and the relation between them is [MASK].	support attack none
	IDEA-3	Sentence 1: S1, Sentence 2: S2. There is [MASK] relation from sentence 1 to sentence 2.	support attack none

Table 3: The Parameters for The Model of Argument Relation Identification

Parameter	Value
Optimizer	AdamW
Batch Size	32
Learning Rate	1e-5
Max Length	256
Total Epoch	5
Loss Function	Cross Entropy

Table 4: The results of all teams for Argument Unit Identification. The underlined runs are that we submitted. The bolded one is the one that yielded the best result out of the runs we submitted.

Team-Run	Micro-F1	Macro-F1	Weight-F1
TMUNLP-1	76.57%	76.55%	76.59%
IDEA-1	**76.47%**	**76.46%**	**76.48%**
TUA1-1	76.37%	76.36%	76.38%
IMNTPU-2	76.06%	76.05%	76.07%
TMUNLP-3	76.06%	76.04%	76.07%
TMUNLP-2	75.95%	75.94%	75.97%
MONETECH-3	75.54%	75.53%	75.56%
IMNTPU-1	75.44%	75.31%	75.40%
MONETECH-1	75.13%	75.13%	75.12%
MONETECH-2	75.03%	75.02%	75.04%
TUA1-0	74.61%	74.56%	74.62%
WUST-1	74.41%	74.41%	74.41%
LIPI-3	73.89%	73.86%	73.90%
IDEA-3	73.68%	73.68%	73.69%
LIPI-1	73.48%	73.47%	73.49%
LIPI-2	73.27%	73.27%	73.28%
SCUNLP-1-2	71.10%	71.07%	71.02%
SCUNLP-1-3	71.10%	70.53%	70.73%
SCUNLP-1-1	68.73%	68.62%	68.53%
WUST-2	69.04%	67.76%	68.07%
IMNTPU-3	56.97%	56.82%	56.70%

results. Section 3.1 and Section 3.2 describes and analyzes the effects of our predictions for the two tasks, respectively.

3.1 Argument Unit Identification

In the Argument Unit Identification task, we submitted a total of two runs (IDEA-1 and IDEA-3) and the results of all teams are shown in Table 4.

From the experimental results in Table 4, our approach (corresponding to IDEA-1 run) is able to achieve better prediction results than most approaches of other teams (both Micro-F1 and Macro-F1). Meanwhile, our additional run (IDEA-3) only uses BERT for text classification, which is not as effective as when we introduced CNN. This demonstrates that CNN have a unique advantage in feature extraction.

3.2 Argument Relation Identification

In the Argument Relation Classificatio task, we submitted a total of three runs (IDEA-1, IDEA-2 and IDEA-3) and the results of all teams are shown in Table 5.

From the experimental results in Table 5, it is shown that for the Argument Relation Identification task, our approach (corresponding to IDEA-3 run) does not have a significant advantage. We can find that Micro-F1 is barely satisfactory in IDEA-3, but Macro-F1 is far behind the other teams'. This is due to data imbalance in the second task. After analyzing the official test set, we have not taken enough account of the data imbalance that caused the huge difference between Macro-F1 and Micro-F1.

Next, we use different templates in IDEA-1 and IDEA-3, which results in an improvement of 1.44% and 1.43% for Micro-F1 and Macro-F1 in IDEA-3 over IDEA-1, respectively. It can be seen that the different templates have an effect on the results of the model. In addition, we concatenate two sentences to convert this task into a text classification task in IDEA-2. However, experimental results show that such an approach is simply crude and does not yield good results.

4 CONCLUSION & FUTURE WORK

This paper illustrates the relevant work of IDEA team at the NTCIR-17 FinArg-1 Task. We participate in a total of two subtasks, the first of which is a text classification task and the second was a sentence pair relationship judgment task. In these tasks, we introduce pre-trained models, CNN and Prompt Tuning. In the first task, our approach achieves relatively good results, but in the second task, our approach can not work as well.

Table 5: The results of all teams for Argument Relation Identification. The underlined runs are that we submitted. The bolded one is the one that yielded the best result out of the runs we submitted.

Team-Run	Micro-F1	Macro-F1	Weight-F1
TUA1-1	85.65%	61.50%	84.86%
LIPI-3	79.42%	60.22%	78.90%
TMUNLP-2	82.03%	57.90%	81.57%
TMUNLP-1	81.88%	57.36%	81.45%
TMUNLP-3	81.88%	56.72%	81.52%
TUA1-2	81.30%	56.26%	80.76%
TUA1-0	85.94%	55.36%	85.13%
SCUNLP-1-3	72.17%	54.06%	72.35%
WUST-1	78.70%	53.97%	77.93%
IMNTPU-2	82.61%	52.97%	82.14%
IDEA-3	**81.74%**	**51.85%**	**80.88%**
LIPI-1	80.72%	51.35%	80.09%
IDEA-1	80.58%	51.12%	79.89%
LIPI-2	80.29%	51.08%	79.79%
IMNTPU-3	80.72%	50.73%	79.67%
SCUNLP-1-2	68.55%	49.00%	68.57%
IMNTPU-1	78.99%	47.36%	76.54%
SCUNLP-1-1	68.70%	45.68%	68.05%
IDEA-2	57.10%	29.18%	59.39%

Our approach for the Argument Relation Identification task is limited by data imbalance, which leads to unsatisfactory evaluation results. In the future, we'll dive deeper into Sentiment Analysis and Argument Mining to perform some fine-grained tasks. Also, we will focus on the problem of data imbalance and explore mitigation.

REFERENCES

[1] Alaa Alhamzeh, Romain Fonck, Erwan Versmée, Elöd Egyed-Zsigmond, Harald Kosch, and Lionel Brunie. 2022. It's Time to Reason: Annotating Argumentation Structures in Financial Earnings Calls: The FinArg Dataset. In *Proceedings of the Fourth Workshop on Financial Technology and Natural Language Processing (FinNLP)*. Association for Computational Linguistics, Abu Dhabi, United Arab Emirates (Hybrid), 163–169. https://doi.org/10.18653/v1/2022.finnlp-1.22

[2] Leo Breiman. 2001. Random forests. *Machine learning* 45 (2001), 5–32.

[3] Chung-Chi Chen, Hen-Hsen Huang, Yu-Lieh Huang, Hiroya Takamura, and Hsin-Hsi Chen. 2022. Overview of the ntcir-16 finnum-3 task: investor's and manager's fine-grained claim detection. In *Proceedings of the 16th NTCIR conference on evaluation of information access technologies, Tokyo, Japan (forthcoming)*.

[4] Chung-Chi Chen, Chin-Yi Lin, Chr-Jr Chiu, Hen-Hsen Huang, Alaa Alhamzeh, Yu-Lieh Huang, Hiroya Takamura, and Hsin-Hsi Chen. 2023. Overview of the NTCIR-17 FinArg-1 Task: Fine-Grained Argument Understanding in Financial Analysis. In *Proceedings of the 17th NTCIR Conference on Evaluation of Information Access Technologies, Tokyo, Japan*.

[5] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*. Association for Computational Linguistics, 4171–4186.

[6] Yoon Kim. 2014. Convolutional Neural Networks for Sentence Classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL*. ACL, 1746–1751.

[7] Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. 2023. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *Comput. Surveys* 55, 9 (2023), 1–35.

[8] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *CoRR* abs/1907.11692 (2019).

[9] John Platt. 1998. Sequential minimal optimization: A fast algorithm for training support vector machines. (1998).

[10] Raymond E Wright. 1995. Logistic regression. (1995).

[11] Li Yujian and Liu Bo. 2007. A normalized Levenshtein distance metric. *IEEE transactions on pattern analysis and machine intelligence* 29, 6 (2007), 1091–1095.