System Report for CCL23-Eval Task 3: UIR-ISC Pre-trained Language Model for Chinese Frame Semantic Parsing

Yingxuan Guan, Xunyuan Liu, Lu Zhang, Zexian Xie, Binyang Li*
Lab of Intelligent Social Computing
University of International Relations
Beijing, China
{guanyingxuan, xyliu, zhanglu, zxxie, byli}@uir.edu.cn

Abstract

Chinese Frame Semantic Parsing (CFSP) is a semantic parsing task based on Chinese FrameNet (CFN). This paper presents a solution for CCL2023-Eval Task 3. We first attempt various pre-trained models for different sub-tasks. Then, we explore multiple approaches to solving each task from the perspectives of feature engineering, model structure, and other tricks. Finally, we provide prospects for the task and propose potential alternative solutions. We conducted extensive comparative experiments to validate the effectiveness of our system.

1 Introduction

Chinese Frame Semantic Parsing (CFSP) is a semantic parsing task based on Chinese FrameNet (CFN) (Gildea and Jurafsky, 2002). To gain a thorough understanding of the events included in the sentence, it aims to extract the frame semantic structure from the sentence, including identifying the frame activated by the target word, Frame elements, etc. In downstream tasks like relation extraction (Zhao et al., 2020) and reading comprehension (Guo et al., 2020b; Guo et al., 2020a), FSP is extremely important.

1.1 Task Definition

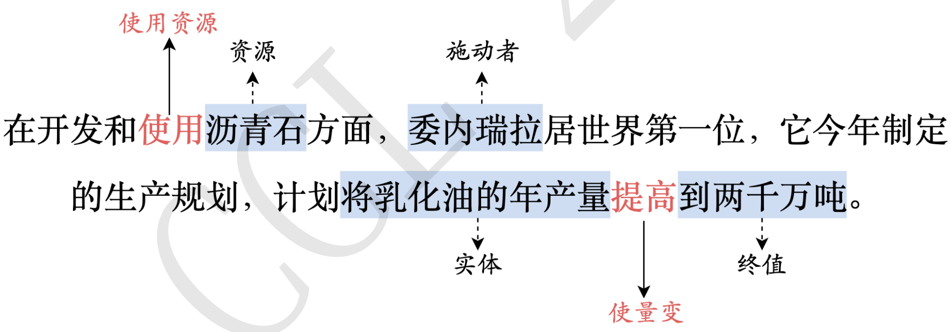

Figure 1: Example definition of three sub-tasks

Frame Identification is a large-scale text classification task that seeks to identify the activated frame based on the target word in the sentence. There are numerous frame types, but there are few sentences that fit into each class, which presents a challenge. Given a sentence including a target word, this task requires to identify the activated frame. The target word is showed as 目标词 in the Figure 1. The input is the sentence, and the output is the identification result of the activated frame 框架 indicated by the solid arrow in Figure 1.

The goal of Argument Identification is to indentify the range of argument spans in a sentence that are controlled by a certain target word. We model this task as a sequence labeling task, which is to identify the arguments that are present in the sentence, namely their start and end positions. The input is the

*Corresponding author

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 124-138, Harbin, China, August 3 - 5, 2023.
(c) Technical Committee on Computational Linguistics, Chinese Information Processing Society of China

sentence and the target word, the output is the starting position and ending position of every identified argument span , which is showed as 论元 in Figure 1, and each sentence may have one argument at least.

The goal of Role Classification task is to identify the semantic role of the argument that was identifid in the Argument Identification task. Additionally, we model the Role Classification as a large-scale text classification task. There are several potential argument spans and semantic roles inside various frames, and each argument may take on different semantic roles in different sentences. The inputs include the sentence, the target word, frame information, and the positions of the argument spans. The output is the argument Role Classification result indicated by the dotted arrow in Figure 1.

1.2 Contribution

Our main contributions can be summarized in the following two points:

1. We propose our own solution for the three sub-tasks, which improved more than the baseline in evaluation metrics like accuracy and F1-score.

2. We explore a number of approaches to the three sub-tasks from the perspectives of feature engineering, model structure, and tricks. We also provide an analysis of the experimental results.

2 Related Work

The FrameNet is a taxonomy of manually identified semantic frames for English(Fillmore et al., 2012). Listed in the FN with each frame are a set of lemmas with part of speech that can evoke the frame, which are called lexical units (LUs). Accompanying most LUs in the FN is a set of exemples annotated for them. Moreover, there are a set of labeled relations between frames(Liu et al., 2016).

Frame Identification can be abstracted as an event extraction task. Certain approaches employ pattern matching techniques by learning patterns from annotated corpora to extract information from documents(Riloff and others, 1993; Kim and Moldovan, 1995). Subsequently, machine learning-based event extraction treats event category classification and event element extraction as classification problems, aiming to categorize event-triggering words into their respective event types(Chieu and Ng, 2002; Ahn, 2006; Llorens et al., 2010). In recent years, the remarkable feature representation capabilities demonstrated by deep learning methods(Nguyen and Grishman, 2015; Chen et al., 2015; Nguyen et al., 2016) have yielded impressive results in event extraction. By fine-tuning pretrained models on downstream tasks(Peters et al., 2018), these methods have significantly enhanced overall performance.

Since event detection benefits many NLP applications (Cheng and Erk, 2018), intensive efforts have been devoted to identifying their arguments. Several studies use dependency parsers to obtain features (Li et al., 2019), or use sequence labeling as a viable solution (Strzyz et al., 2019; Du and Cardie, 2020; Veyseh et al., 2021). Meanwhile, argument Identification has been recently addressed by end-to-end event extraction models, (Wadden et al., 2019; Lin et al., 2020; Li et al., 2021). Lately, some works reformulated the task as a Question Answering problem (Wei et al., 2021; Lyu et al., 2021; Sulem et al., 2022; Du and Ji, 2022))or as a constrained text generation problem (Dai et al., 2022) using predefined prompts or templates. Role Classification is a subtask of argument extraction, the goal is to identify the semantic role assumed by argument. Researchers usually model this task as a text classification problem. Meanwhile, Using the interaction between argument roles can improve the performance of argument extraction (Ding et al., 2022).

FrameNet is a typical method for frames emantic parsing. It consists of mapping a predicate into a frame, and analysis of the frame's elements. FrameNet has been applied to many downstream tasks, such as Machine Reading Comprehension and Text Summarization (Guan et al., 2021b; Guan et al., 2021a). Chinese FrameNet is a frame semantic resource refer to FrameNet, and based on Chinese corpus, including frames, frame elements, lexical units and frame relations. Chinese framenet is increasingly important in Chinese information processing.

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 124-138, Harbin, China, August 3 - 5, 2023.

3 Background

3.1 Dataset

The dataset utilized in this evaluation is the Chinese FrameNet (CFN) dataset. It is a frame semantic resource constructed by Shanxi University based on Chinese real-language corpus. The data consists of frame knowledge and annotated example sentences, including nearly 700 semantic frames and 20,000 annotated examples.

In the annotated example sentences, the information includes example sentence ID, frame element annotations, activated frame name, target word and its position, part-of-speech, annotated text, tokenized text with part-of-speech information.

In the frame information, it includes the English and Chinese names of the frame, frame definition, English and Chinese names of frame roles, abbreviations, definitions, and other information.

In the CFN dataset, there is a significant long-tail effect in the frame information. We sorted the frames in descending order of their frequencies and collected one sample every 8 frames to plot a bar graph. As shown in Figure 2, the most frequent frame is “陈述” (statement), which appears 343 times. The least frequent frames, such as “不复存在” (no longer exist), “能否使用” (whether it can be used), “同时性” (simultaneity), etc., only appear once. Frames with frequencies below 20 account for more than half of all categories. This reflects the difference between the head and the tail, indicating that a few high-frequency frames occur repeatedly, while the occurrences of numerous tail frames are more scattered.

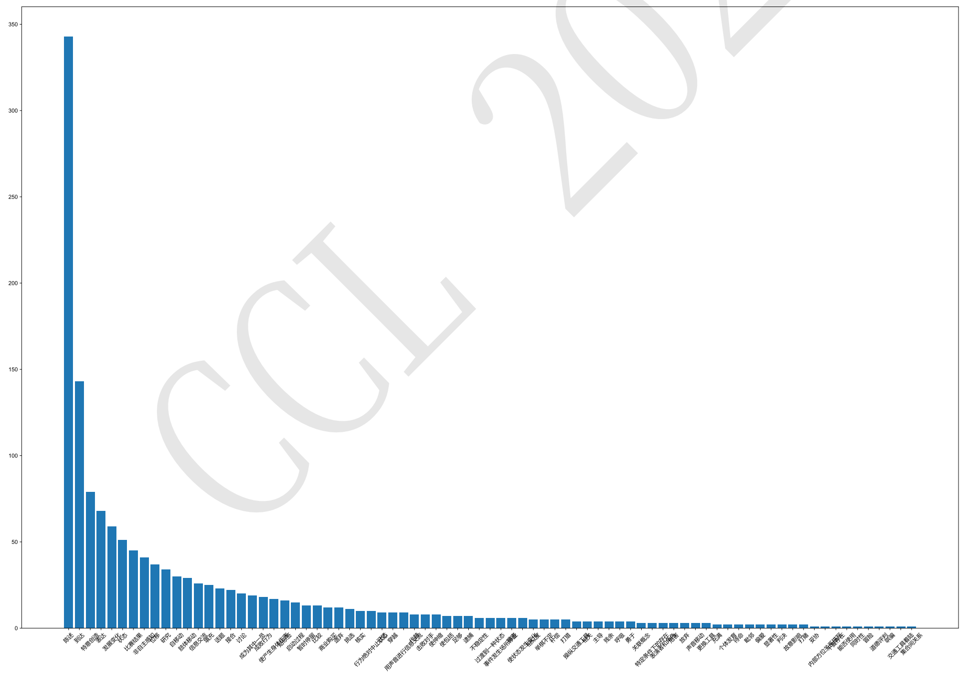

Figure 2: Long-tail effect of the frame distribution of the original dataset

Due to the scarcity of data samples for low-frequency frames compared to high-frequency frames, they are difficult to classify accurately. Therefore, we consider enhancing the dataset to mitigate the problem of poor classification performance caused by the long-tail effect.

3.2 Evaluation Metrics

- Frame Identification

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 124-138, Harbin, China, August 3 - 5, 2023.

The evaluation metric for the framework identification task is accuracy.

$$task1\_acc = \frac{number\ of\ correctly\ identified\ frames}{total\ number\ of\ frames} \tag{1}$$

- Argument Identification

  The evaluation metrics for argument span identification are precision (P), recall (R), and F1-score (F1).

$$task2\_precision = \frac{InterSec(gold, pred)}{Len(pred)} \tag{2}$$

$$task2\_recall = \frac{InterSec(gold, pred)}{Len(gold)} \tag{3}$$

$$task2\_f1 = \frac{2 * task2\_precision * task2\_recall}{task2\_precision + task2\_recall} \tag{4}$$

In this context: *gold* and *pred* represent the ground truth and predicted results, respectively. *InterSec()* refers to the calculation of the number of tokens shared by both the ground truth and predicted results. *Len()* refers to the calculation of the total number of tokens.

- Role Classification

  The evaluation metrics for argument Role Classification are precision (P), recall (R), and F1-score (F1).

$$task3\_precision = \frac{Count(gold \cap pred)}{Count(pred)} \tag{5}$$

$$task3\_recall = \frac{Count(gold \cap pred)}{Count(gold)} \tag{6}$$

$$task3\_f1 = \frac{2 * task3\_precision * task3\_recall}{task3\_precision + task3\_recall} \tag{7}$$

In this context: *gold* and *pred* respectively represent the ground truth and predicted results. *Count(*)* indicates the calculation of the number of elements in a set.

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 124-138, Harbin, China, August 3 - 5, 2023.

3.3 Model Overview

- **BERT**

  BERT (Devlin et al., 2018) is a deep pre-trained language model based on the Transformer architecture. It leverages large amounts of unlabeled data for pre-training through tasks such as Masked Language Model (MLM) and Next Sentence Prediction (NSP). This allows the model to learn rich language representations. After pre-training, BERT can be fine-tuned on specific downstream tasks, enabling training for various specific tasks.

- **LERT**

  LERT (Cui et al., 2022) (Linguistically-motivated bidirectional Encoder Representation from Transformer) employs the Linguistically-motivated Information Pre-training (LIP) strategy, which incorporates three types of linguistic features and the original MLM pre-training task. These linguistic features include Part-of-Speech (POS) tagging, Named Entity Recognition (NER), and Dependency Parsing (DEP), among others. This strategy enables faster learning of foundational language knowledge. Experimental results on ten Chinese natural language understanding tasks demonstrate that the LERT algorithm significantly improves the performance of various pre-trained language models.

- **ERNIE3.0**

  ERNIE 3.0 (Sun et al., 2021) (Enhanced Representation through Knowledge Integration) is pre-trained on a 4TB corpus that includes both plain text and a knowledge graph to enhance knowledge integration. To address language understanding and generation tasks through zero-shot learning, few-shot learning, and fine-tuning, ERNIE 3.0 introduces a unified pre-training framework that integrates autoencoder networks and autoregressive networks. Experimental results demonstrate that ERNIE 3.0 consistently outperforms state-of-the-art models across 54 benchmark tests and achieves first place in the SuperGLUE benchmark test.

3.4 Loss Function

Loss function is used to evaluate the extent to which the predicted and true values of the model are not the same. In this task, the Focal Loss (Lin et al., 2017) function is used to better alleviate the problem of unbalanced number of sample categories.

The goal of Focal Loss is to address the issue where traditional cross-entropy loss contributes less to the loss of positive samples when there is a large number of easily classified negative samples. Traditional cross-entropy loss tends to focus on the majority of negative samples and neglects the minority of positive samples when dealing with highly imbalanced datasets. By introducing Focal Loss, the model can better handle class imbalance problems and pay more attention to difficult-to-classify samples, thereby improving the performance of classification tasks.

$$loss(o,t) = -\frac{1}{n}(\sum_i (t[i] * log(o[i]) + (1 - t[i]) * log(1 - o[i]))) \tag{8}$$

As shown in formula 8, we use balance factor to deal with unbalanced samples in Balance Cross Entropy loss(BCEloss).

$$\mathrm{FL}\left(p_{\mathrm{t}}\right) = -\alpha_{\mathrm{t}}\left(1 - p_{\mathrm{t}}\right)^{\gamma} \log\left(p_{\mathrm{t}}\right) \tag{9}$$

Focal loss is specially designed for the one-stage detection algorithm, which reduces the loss weight of easy-to-distinguish negative examples. It increases the dynamic adjustment factor based on BCEloss to achieve the effect of difficult sample mining. We make the model more focused on hard-to-learn samples by setting γ value as 2 in the formula 9, thus the network will not be biased by too many negative examples.

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 124-138, Harbin, China, August 3 - 5, 2023.

4 Model

4.1 Dataset processing

In our approach to data augmentation, we treat all fields at the same level as sentence_id, along with their corresponding subfields (such as sentence_id, cfn_spans, frame, target, text, word and their respective subfields), as a single data unit. Since there are 695 frame categories shared among 10,000 training data, the frame field may be the same across different data units. Hence, we tally the occurrences of the frame field. If a data unit exceeds a certain frequency threshold, we create duplicates of that unit. Specifically, for frame frequencies between 10 and 20, we replicate the data unit three times, while for frequencies below 10, we replicate it ten times. Consequently, the augmented dataset comprises around 26,000 data units, compared to the original dataset of 10,000. Following data augmentation, the dataset becomes smoother, alleviating the long-tail effect that was pronounced in the pre-augmented dataset(Karimi et al., 2021). This, to some extent, reduces the difficulty of the classification task(Wei and Zou, 2019).

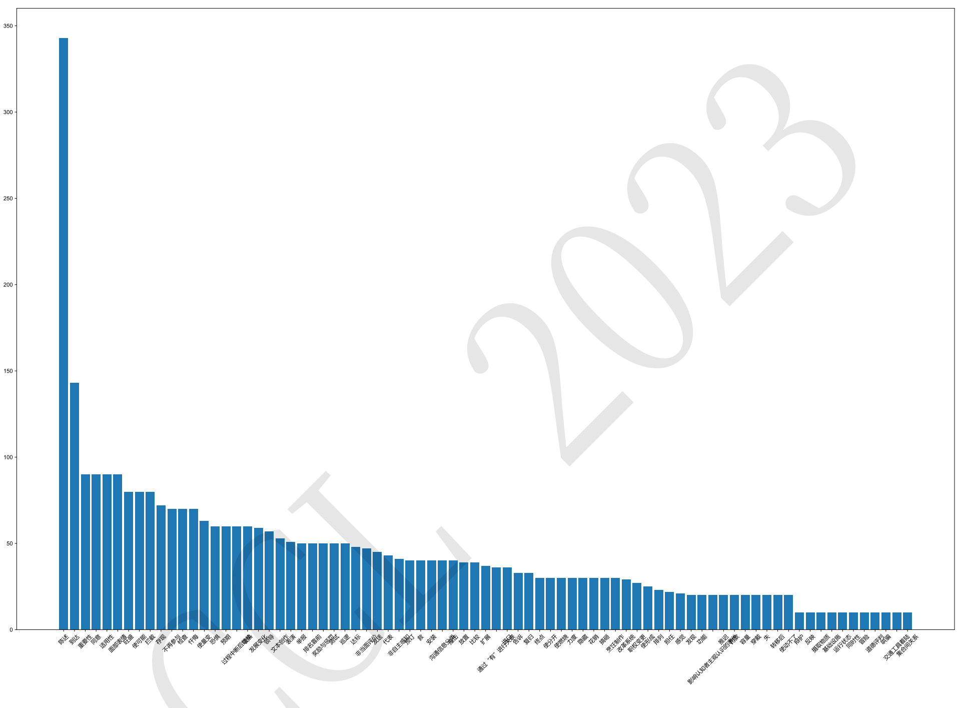

Figure 3: The frame distribution after data augmentation

4.2 Frame Identification

The long-tail effect of the dataset influences the performance in large-scale text classification tasks. The class with few examples is inadequately modeled, even be suppressed, and the model can only learn the properties of the category with enough samples. To reduce the long-tail effect, we choose to use the dataset after data augmentation and focal loss as the loss function.

We choose BERT as the pre-trained model, and the model input is $[CLS] + sentence + [SEP]$. After extracting features from the hidden layer of BERT, the $[CLS]$ vector is concatenated after the average-pooling target word vector, and then classified by MLP. A special $[CLS]$ vector will help the model for classification. The model process is shown in Figure 4.

We use the training dataset as an extra-domain knowledge base at the same time because the target word and the activated frame are a one-to-one single-label text classification task. After getting the predictions of the trained model, we integrate them using the rule matching method, mapping the target

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 124-138, Harbin, China, August 3 - 5, 2023.

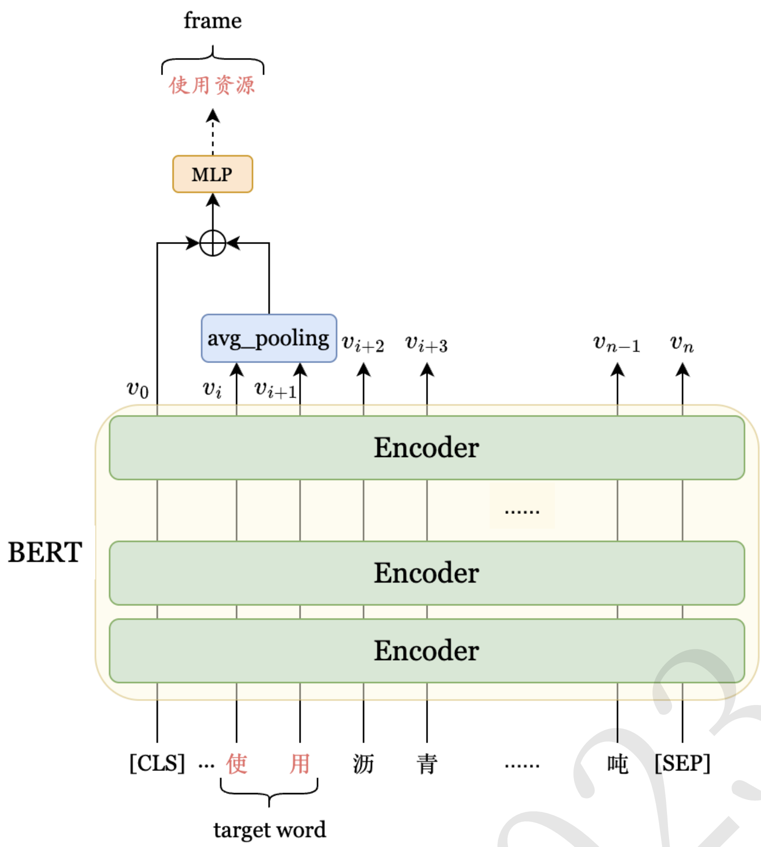

Figure 4: Frame Identification process

words into the appropriate frames. The same target word can be directly matched if it appears during the test. This method can enhance and correct the model's predictions and increase the model performance.

4.3 Argument Identification

We model the Argument Identification as a sequence labeling task, using LERT integrated with linguistic knowledge as the pre-trained model. And we use the original dataset with focal loss as the loss function. The model process is shown in Figure 5.

The model input is $[CLS] + sentence + [SEP]$ and the target word with special token $< t >< /t >$. This allows the model to learn where the target word is located. The model classifies each token and identifies whether the token is part of the argument span, that is, to identify whether the token is the start position, end position or middle position of the argument span.

As shown in Figure 5, when the classification result is 0, it means that the token is not part of the argument; when the classification result is 1, it means that the token is the start position of the argument, and when the classification result is 2; it means that the token is in other positions of the argument except the start position.

4.4 Role Classification

The goal of the Role Classification task is to identify the semantic role of the argument in the sentence based on the argument extracted by Argument Identification. We also model Role Classification as a large-scale text classification task. The model process is shown in Figure 6.

We use ERNIE as the pre-trained model, and use the augmentation dataset. The model input is $[CLS] + sentence + [SEP]$, the target word with special token $< t >< /t >$, frame information with special token $< f >< /f >$ and argument spans. This allows the model to learn the location of the target word and the frame that it activates.

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 124-138, Harbin, China, August 3 - 5, 2023.
(c) Technical Committee on Computational Linguistics, Chinese Information Processing Society of China

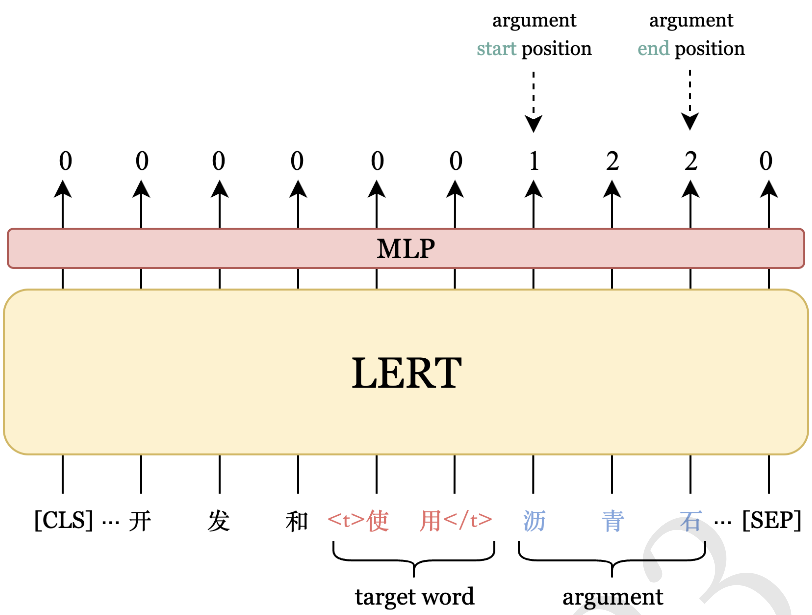

Figure 5: Argument Identification process

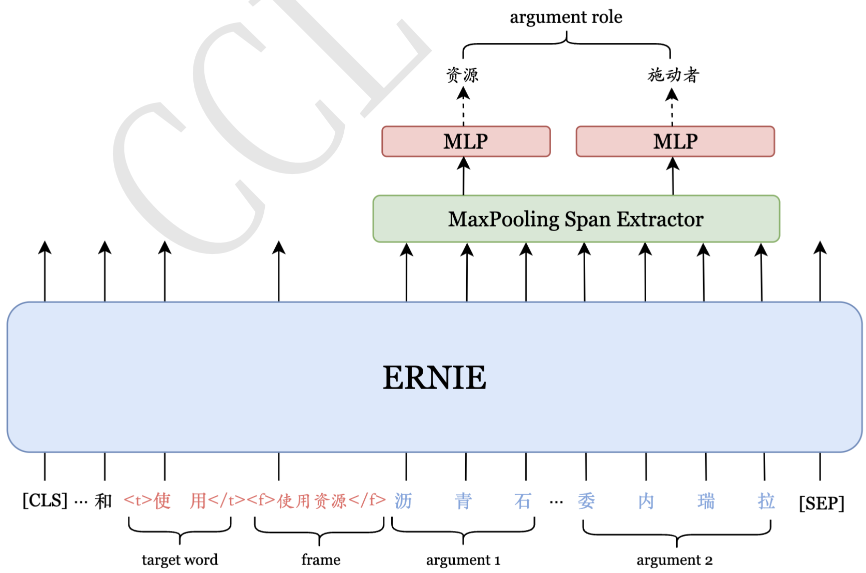

Figure 6: Role Classification process

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 124-138, Harbin, China, August 3 - 5, 2023.

As shown in Figure 6, the Maxpooling Span extractor is to max-pooling the argument span vectors, and get the arguments features. The model classifies each argument, and identifies the semantic role that the argument assumes.

5 Experiments

5.1 Environmental Setup

We train on Nvidia RTX3090-24G. It takes about 2 hours to train 20 epochs for the Frame Identification task; about 1 hour for 30 epochs for the Argument Identification task; about 1.5 hours for 30 epochs for the Role Classification task.

5.2 Results

We give the result of different models and methods including the two ranking lists of A and B.

Table 1: Frame Identification Experimental Results

Ranking List	Model	Method	Acc(%)
A	BERT	dataset_aug	62.35
		Focal loss BiLSTM	63.05
		dataset_aug concat_CLS	63.35
		concat_CLS Focal loss BiLSTM	63.45
		dataset_aug concat_CLS R-drop	64.40
		concat_CLS	64.55
		baseline	65.10
		concat_CLS Focal loss	66.15
		dataset_aug concat_CLS Focal loss	66.30
		dataset_aug concat_CLS Focal loss rule matching	**69.70 (↑4.6)**
	ERNIE	/	65.90
B	BERT	dataset_aug concat_CLS Focal loss	60.12
		dataset_aug concat_CLS Focal loss rule matching	**65.14**

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 124-138, Harbin, China, August 3 - 5, 2023.

As shown on Table 1, in the Frame Identification task, we compared the BERT and ERNIE models. Since ERNIE is not significantly improved compared to BERT, in order to reduce model parameters, we choose to use BERT as the pre-trained model.

At the same time, we adopt a data augmentation strategy, introduce the $[CLS]$ vector, use focal loss as the loss function, use the rule matching algorithm, add BiLSTM layer, R-Drop and other tricks to improve the classification accuracy.

Table 2: Argument Identification Experimental Results

Ranking List	Model	Method	F1-score
A	BERT	baseline	87.55
	LERT	dataset_aug	87.67
		BiLSTM Focal loss	87.93
		dataset_aug Focal loss	88.12
		/	88.13
		CRF	88.46
		Focal loss	**89.03(↑1.48)**
B	LERT	Focal loss	**87.94**

As shown on Table 2, in the Argument Identification task, we compared the BERT and LERT models. Since LERT has a greater improvement than BERT without any other tricks, we make a series of attempts on LERT, including data augmentation strategies, focal loss, adding BiLSTM layers, adding CRF layers, etc.

Table 3: Role Classification Experimental Results

Ranking List	Model	Method	F1-score
A	BERT	baseline	54.08
	ERNIE	/	55.56
		BiLSTM Focal loss	54.79
		dataset_aug Focal loss	56.09
		Focal loss	56.51
		dataset_aug	**56.71(↑2.63)**
	LERT	/	55.18
B	LERT	/	50.65
	ERNIE	/	**50.85**

As shown on Table 3, in the Role Classification task, we compared the three models of BERT, ERNIE and LERT. Since ERNIE and LERT have a higher improvement than BERT without any other tricks, and ERNIE has the best improvement effect, we have made a series of attempts based on the ERNIE,

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 124-138, Harbin, China, August 3 - 5, 2023.

including data augmentation strategies, focal loss, and adding BiLSTM layer, etc.

5.3 Analysis

In the Frame Identification task, we found that the method of choosing BERT as pre-trained model, using the augmentation dataset, concatenating $[CLS]$ vectors, using focal loss as the loss function, and using the rule matching algorithm achieve the best performance.

After data augmentation and using focal loss as the loss function, the unbalanced distribution of dataset problem is well alleviated. The introduction of the $[CLS]$ vector, that is, the introduction of the overall features of the sentence, and the integration of sentence-level features will be useful for the model to classify tokens. It is worth noting that adding the BiLSTM layer will reduce the accuracy of the model. We guess that it is because for the Frame Identification task, the model needs to distinguish the target word features, and the BiLSTM layer will instead make the model pay attention to the semantic information of other words in the sentence, introducing unnecessary noise. In order to reduce the influence of dropout on the model, we try to enhance the robustness of the model through R-Drop. R-Drop limits the KL divergence between the output distributions of the two sub-models sampled by dropout, so as to alleviate the problem of inconsistency between prediction and training, but from the results it seems that the performance is not as expected.

In the Argument Identification task, we found that choosing LERT as pre-trained model and using the original dataset with focal loss works best. Since the sequence labeling task often requires the model to pay attention to the semantic information of the context, we add the BiLSTM layer and the CRF layer. Although there is a certain improvement compared to the baseline, it does not improve greater than directly using focal loss as the loss function.

In the Role Classification task, we found that choosing ERNIE as pre-trained model works best with only the augmentation dataset. Similar to the Frame Identification task, the Role Classification task is also a large-scale text classification task, so adding the BiLSTM layer and CRF layer will not help the model much, and may introduce unnecessary noise.

Sub-tasks 1 and 3 can be abstracted as classification tasks, and the utilization of data augmentation methods can be highly effective in improving the performance of these text classification tasks. The reason behind this lies in the fact that data augmentation enhances the model's capacity to generalize by introducing a broader diversity of training data(Shorten et al., 2021). By augmenting the dataset, models can expose themselves to a wider spectrum of language patterns, enabling them to gain a better understanding and proficiency in handling diverse types of textual data(Bayer et al., 2022). As a result, the application of data augmentation leads to a significant improvement in classification accuracy.

6 Future Work

Due to limited evaluation time, we have some unfinished attempts on the three sub-tasks.

- For the Frame Identification task, we think that we can try to use methods such as copying the target word and adding synonyms to make the model pay more attention to the target word in the sentence and increase the semantic information of the target word.

- For the Argument Identification task, we think that we can try to model the task as an Machine Reading Comprehension task, so that the model can better understand the semantic information of the sentence.

- For the Role Classification task, since there are only several possible roles in the frame activated by target word, we think that we can try to let the model only classify the role of the argument under the activated frame to reduce the interference of other classes. Furthermore, since different roles have different meanings under different frameworks, we can also try to integrate the role description information of arguments, so that the model can better understand the meaning of argument roles under different frames, thereby improving Role Classification performance.

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 124-138, Harbin, China, August 3 - 5, 2023.

7 Conclusion

Experimental results have proved that our proposed solution has achieved greater improvement compared with the baseline in the three sub-tasks of Frame Identification, Argument Identification, and Role Classification. At the same time, we have made a variety of attempts to solve each task from different perspectives, and give an analysis of the experimental results. Modeling the task as one certain problem is the key to solving the task. When using complex methods to solve traditional tasks, returning to the essence of the task and simplifying it is often a good way to break through the bottleneck.

Acknowledgements

We would like to thank Zhengyi Zhao, who offered us invaluable advice on improving the experiments. Additionally, we extend our gratitude to Yan Li for supporting the experimental GPU resources. This project was partially supported by National Natural Science Foundation of China (Grant number: 61976066), Beijing Natural Science Foundation (Grant number: 4212031), the Fundamental Research Fund for the Central Universities (Grant numbers: 3262023T19), and Research Funds for NSD Construction, University of International Relations (Grant numbers: 2021GA07).

References

David Ahn. 2006. The stages of event extraction. In *Proceedings of the Workshop on Annotating and Reasoning about Time and Events*, pages 1–8.

Markus Bayer, Marc-André Kaufhold, and Christian Reuter. 2022. A survey on data augmentation for text classification. *ACM Computing Surveys*, 55(7):1–39.

Yubo Chen, Liheng Xu, Kang Liu, Daojian Zeng, and Jun Zhao. 2015. Event extraction via dynamic multi-pooling convolutional neural networks. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 167–176.

Pengxiang Cheng and Katrin Erk. 2018. Implicit argument prediction with event knowledge. In Marilyn A. Walker, Heng Ji, and Amanda Stent, editors, *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2018, New Orleans, Louisiana, USA, June 1-6, 2018, Volume 1 (Long Papers)*, pages 831–840. Association for Computational Linguistics.

Hai Leong Chieu and Hwee Tou Ng. 2002. A maximum entropy approach to information extraction from semi-structured and free text. *Aaai/iaai*, 2002:786–791.

Yiming Cui, Wanxiang Che, Shijin Wang, and Ting Liu. 2022. Lert: A linguistically-motivated pre-trained language model. *arXiv preprint arXiv:2211.05344*.

Lu Dai, Bang Wang, Wei Xiang, and Yijun Mo. 2022. Bi-directional iterative prompt-tuning for event argument extraction. In Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang, editors, *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, EMNLP 2022, Abu Dhabi, United Arab Emirates, December 7-11, 2022*, pages 6251–6263. Association for Computational Linguistics.

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

Nan Ding, Chunming Hu, Kai Sun, Samuel Mensah, and Richong Zhang. 2022. Explicit role interaction network for event argument extraction. In Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang, editors, *Findings of the Association for Computational Linguistics: EMNLP 2022, Abu Dhabi, United Arab Emirates, December 7-11, 2022*, pages 3475–3485. Association for Computational Linguistics.

Xinya Du and Claire Cardie. 2020. Document-level event role filler extraction using multi-granularity contextualized encoding. In Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel R. Tetreault, editors, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, pages 8010–8020. Association for Computational Linguistics.

Xinya Du and Heng Ji. 2022. Retrieval-augmented generative question answering for event argument extraction. In Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang, editors, *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, EMNLP 2022, Abu Dhabi, United Arab Emirates, December 7-11, 2022*, pages 4649–4666. Association for Computational Linguistics.

Charles J Fillmore, Russell Lee-Goldman, and Russell Rhodes. 2012. The framenet constructicon. *Sign-based construction grammar*, (193):309–372.

Daniel Gildea and Daniel Jurafsky. 2002. Automatic labeling of semantic roles. *Comput. Linguistics*, 28(3):245–288.

Yong Guan, Shaoru Guo, Ru Li, Xiaoli Li, and Hongye Tan. 2021a. Frame semantic-enhanced sentence modeling for sentence-level extractive text summarization. In Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih, editors, *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 7-11 November, 2021*, pages 4045–4052. Association for Computational Linguistics.

Yong Guan, Shaoru Guo, Ru Li, Xiaoli Li, and Hu Zhang. 2021b. Integrating semantic scenario and word relations for abstractive sentence summarization. In Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih, editors, *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 7-11 November, 2021*, pages 2522–2529. Association for Computational Linguistics.

Shaoru Guo, Yong Guan, Ru Li, Xiaoli Li, and Hongye Tan. 2020a. Incorporating syntax and frame semantics in neural network for machine reading comprehension. In Donia Scott, Núria Bel, and Chengqing Zong, editors, *Proceedings of the 28th International Conference on Computational Linguistics, COLING 2020, Barcelona, Spain (Online), December 8-13, 2020*, pages 2635–2641. International Committee on Computational Linguistics.

Shaoru Guo, Ru Li, Hongye Tan, Xiaoli Li, Yong Guan, Hongyan Zhao, and Yueping Zhang. 2020b. A frame-based sentence representation for machine reading comprehension. In Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel R. Tetreault, editors, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, pages 891–896. Association for Computational Linguistics.

Akbar Karimi, Leonardo Rossi, and Andrea Prati. 2021. Aeda: an easier data augmentation technique for text classification. *arXiv preprint arXiv:2108.13230*.

Jun-Tae Kim and Dan I. Moldovan. 1995. Acquisition of linguistic patterns for knowledge-based information extraction. *IEEE transactions on knowledge and data engineering*, 7(5):713–724.

Diya Li, Lifu Huang, Heng Ji, and Jiawei Han. 2019. Biomedical event extraction based on knowledge-driven tree-lstm. In Jill Burstein, Christy Doran, and Thamar Solorio, editors, *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)*, pages 1421–1430. Association for Computational Linguistics.

Manling Li, Sha Li, Zhenhailong Wang, Lifu Huang, Kyunghyun Cho, Heng Ji, Jiawei Han, and Clare R. Voss. 2021. The future is not one-dimensional: Complex event schema induction by graph modeling for event prediction. In Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih, editors, *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 7-11 November, 2021*, pages 5203–5215. Association for Computational Linguistics.

Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. 2017. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988.

Ying Lin, Heng Ji, Fei Huang, and Lingfei Wu. 2020. A joint neural model for information extraction with global features. In Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel R. Tetreault, editors, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, pages 7999–8009. Association for Computational Linguistics.

Shulin Liu, Yubo Chen, Shizhu He, Kang Liu, and Jun Zhao. 2016. Leveraging framenet to improve automatic event detection. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2134–2143.

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 124-138, Harbin, China, August 3 - 5, 2023.

Hector Llorens, Estela Saquete, and Borja Navarro. 2010. Timeml events recognition and classification: learning crf models with semantic roles. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pages 725–733.

Qing Lyu, Hongming Zhang, Elior Sulem, and Dan Roth. 2021. Zero-shot event extraction via transfer learning: Challenges and insights. In Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli, editors, *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 2: Short Papers), Virtual Event, August 1-6, 2021*, pages 322–332. Association for Computational Linguistics.

Thien Huu Nguyen and Ralph Grishman. 2015. Event detection and domain adaptation with convolutional neural networks. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 365–371.

Thien Huu Nguyen, Kyunghyun Cho, and Ralph Grishman. 2016. Joint event extraction via recurrent neural networks. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies*, pages 300–309.

Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. In *North American Chapter of the Association for Computational Linguistics*.

Ellen Riloff et al. 1993. Automatically constructing a dictionary for information extraction tasks. In *AAAI*, volume 1, pages 2–1. Citeseer.

Connor Shorten, Taghi M Khoshgoftaar, and Borko Furht. 2021. Text data augmentation for deep learning. *Journal of big Data*, 8:1–34.

Michalina Strzyz, David Vilares, and Carlos Gómez-Rodríguez. 2019. Viable dependency parsing as sequence labeling. In Jill Burstein, Christy Doran, and Thamar Solorio, editors, *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)*, pages 717–723. Association for Computational Linguistics.

Elior Sulem, Jamaal Hay, and Dan Roth. 2022. Yes, no or IDK: the challenge of unanswerable yes/no questions. In Marine Carpuat, Marie-Catherine de Marneffe, and Iván Vladimir Meza Ruíz, editors, *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL 2022, Seattle, WA, United States, July 10-15, 2022*, pages 1075–1085. Association for Computational Linguistics.

Yu Sun, Shuohuan Wang, Shikun Feng, Siyu Ding, Chao Pang, Junyuan Shang, Jiaxiang Liu, Xuyi Chen, Yanbin Zhao, Yuxiang Lu, et al. 2021. Ernie 3.0: Large-scale knowledge enhanced pre-training for language understanding and generation. *arXiv preprint arXiv:2107.02137*.

Amir Pouran Ben Veyseh, Franck Dernoncourt, Quan Hung Tran, Varun Manjunatha, Lidan Wang, Rajiv Jain, Doo Soon Kim, Walter Chang, and Thien Huu Nguyen. 2021. Inducing rich interaction structures between words for document-level event argument extraction. In Kamal Karlapalem, Hong Cheng, Naren Ramakrishnan, R. K. Agrawal, P. Krishna Reddy, Jaideep Srivastava, and Tanmoy Chakraborty, editors, *Advances in Knowledge Discovery and Data Mining - 25th Pacific-Asia Conference, PAKDD 2021, Virtual Event, May 11-14, 2021, Proceedings, Part II*, volume 12713 of *Lecture Notes in Computer Science*, pages 703–715. Springer.

David Wadden, Ulme Wennberg, Yi Luan, and Hannaneh Hajishirzi. 2019. Entity, relation, and event extraction with contextualized span representations. In Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan, editors, *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pages 5783–5788. Association for Computational Linguistics.

Jason Wei and Kai Zou. 2019. Eda: Easy data augmentation techniques for boosting performance on text classification tasks. *arXiv preprint arXiv:1901.11196*.

Kaiwen Wei, Xian Sun, Zequn Zhang, Jingyuan Zhang, Zhi Guo, and Li Jin. 2021. Trigger is not sufficient: Exploiting frame-aware knowledge for implicit event argument extraction. In Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli, editors, *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 1: Long Papers), Virtual Event, August 1-6, 2021*, pages 4672–4682. Association for Computational Linguistics.

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 124-138, Harbin, China, August 3 - 5, 2023.

Computational Linguistics

Hongyan Zhao, Ru Li, Xiaoli Li, and Hongye Tan. 2020. CFSRE: context-aware based on frame-semantics for distantly supervised relation extraction. *Knowl. Based Syst.*, 210:106480.