System Report for CCL24-Eval Task 1: Construction of CFSP Model Based on Non-Finetuning Large Language Model

Fugeng Huang[a,1], **Zhongbin Guo**[b,1], **Wenting Li**[c,†], **Haibo Cheng**[d,†]
[a]School of Software and Microelectronics, Peking University, Beijing, 102600, China
[b]School of Computer and Technology, Beijing Institute of Technology, Beijing, 100081, China
[c]School of Information Engineering, Beijing Institute of Graphic Communication, 102600, China
[d]National Engineering Research Center for Software Engineering, Peking University, 100871, China
[a]2301210243@stu.pku.edu.cn
[b]1120220508@bit.edu.cn
[c,d]{wentingli, hbcheng}@pku.edu.cn

Abstract

Chinese Frame Semantic Parsing (CFSP) is an important task in the field of Chinese Natural Language Processing(NLP). Its goal is to extract the frame semantic structure from the sentence and realize the deep understanding of the events or situations involved in the sentence. This paper mainly studies the application of Large Language Model (LLM) for reasoning through Prompt Engineering without fine-tuning the model, and completes three subtasks of Chinese Framework Semantic Parsing tasks: frame identification, argument Identification and role identification. This paper proposes a Retrieval Augmented Generation (RAG) method for target words, and constructs more refined sample Few-Shot method. We achieved the second place on the B rankings in the open track in the "CCL2024-Eval The Second Chinese Frame Semantic Parsing" competition[*].

1 Introduction

Chinese Frame Semantic Parsing (CFSP) is a research method based on frame semantics (Charles J. Fillmore, 1982), which is based on the semantic representation and annotation of Chinese FrameNet (CFN) (You et al., 2007; You and Liu, 2005; Li et al., 2024), and achieves the purpose of semantic parsing by extracting the frame semantic structure of sentences (Gildea and Jurafsky, 2002). This method is of great significance for a series of downstream tasks such as reading comprehension (Guo et al., 2020a; Guo et al., 2020b; Wang et al., 2016), text summarization (Guan et al., 2021a; Guan et al., 2021b), and relationship extraction (Zhao et al., 2020).

1.1 Task Definition

Table 1 presents the 'Deciding' framework within the Chinese FrameNet, which illustrates the cognitive process of making decisions among various explicit or potential options. Framework elements refer to the participants in the semantic scene corresponding to the framework. For example, 'Cognizer' in the "Deciding" framework is one of the framework elements.

[†] Corresponding authors.
[*] https://tianchi.aliyun.com/competition/entrance/532179

©2024 China National Conference on Computational Linguistics
Published under Creative Commons Attribution 4.0 International License

Proceedings of the 23rd China National Conference on Computational Linguistics, pages 1-9, Taiyuan, China, July 25 - 28, 2024.
Volume3: Evaluations
(c) Technical Committee on Computational Linguistics, Chinese Information Processing Society of China

CCL 2024

Frame name	Deciding
Frame definition	The cognitive person makes a decision in a variety of explicit or potential choices, which may be an entity or a behavioral process.
Frame element	**Cognizer** The cognitive decides to do something.
	Decision The decision represents the entity or process determined by the cognitive.
	Possibilities The cognizer decides which one to choose from a range of possible options.
	Topic Guided by “about”, it indicates the matters involved in the decision, and sometimes the matters involved also indicate the general content of the decision.
	Time Indicates the relative position of the occurrence, progress or end of the action behavior or state in the time dimension, including both time points and time periods.

Table 1: An example of “decision” frame in Chinese FrameNet (CFN)

This evaluation divides Chinese Frame Semantic Parsing into three downstream tasks: Frame Identification (FI) (Su et al., 2021), Argument Identification (AI) and Role Identification (RI). The task of FI is the core task in the research of frame semantics. It requires finding an activated frame for the target word in a given sentence according to its context, which can help the computer identify the key information and semantic framework in the sentence, so as to better understand the meaning of the sentence (Hermann et al., 2014). The main purpose of AI task is to determine the position of argument (i.e. frame element) involved in each target word in the sentence, so as to help the system more accurately identify the semantic role of argument. The RI task aims to determine the semantic role of each argument in its own framework, which plays a vital role in information extraction, relationship extraction and machine translation. The following figure shows the specific work examples of the three downstream tasks of CFSP.

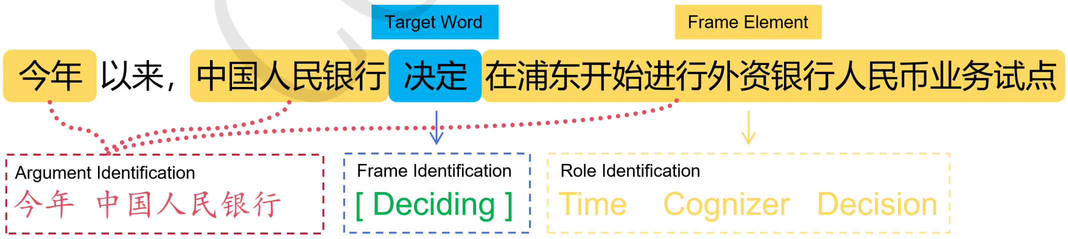


Figure 1: Schematic diagram of Chinese Frame Semantic Parsing task

With the continuous progress of technology, ChatGPT and other Large Language Model (LLM) continue to appear, researchers have also carried out a series of research on the CFSP task based on LLM. The baseline scheme given by the evaluation task shows that under the guidance of Chain of Thinking (CoT), the performance of the three subtasks of CFSP in Zero-Shot and Few-Shot scenarios is not ideal, and LLM cannot understand the text from the perspective of frame, argument and role (Li et al., 2023). This research is mainly committed to using LLM to effectively complete the three downstream tasks under the task of CFSP through the improved scheme.

CCL 2024

1.2 Contribution

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

1. We build a hierarchical index Retrieval Augmented Generation (RAG) system based on target words, use the target word information to filter out some options.
2. We use HanLP to segment sentences, and use the BM25 (Robertson et al., 1994) retrieval algorithm to index, which effectively improves the sample quality of LLM in the Few-Shot scenario.
3. We establish balanced Few-Shot sample categories to ensure that each target word category has a certain amount of data closest to the problem.
4. We change the input from the model to text, and then conduct post-processing to convert it back to the list, so as to avoid wasting attention on the mapping relationship in the learning process of the model.

2 Related Work

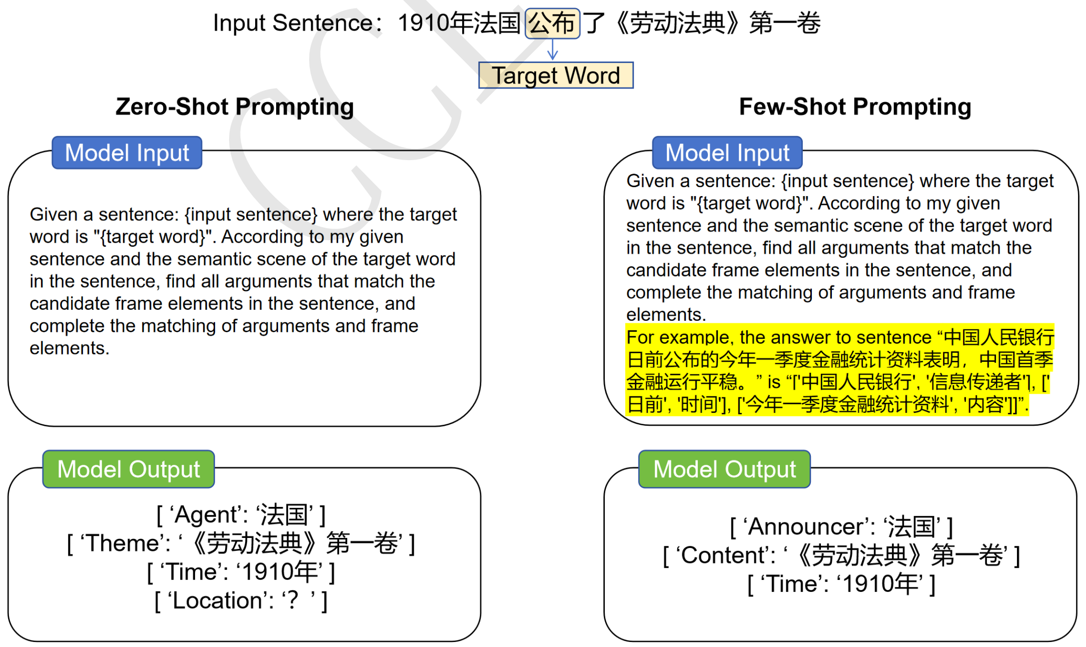
Due to the late appearance of LLM, the relatively novel architecture, and the rapid development and updating speed, there are few researches on the application of LLM in the Chinese Frame Semantic Parsing tasks. Yang (Yang et al., 2023) represents the first attempt at leveraging large pre-trained language models (LLM) for SPARQL generation to address Chinese knowledge graph question answering. (Li et al., 2023) tested the effect of ChatGPT on FI task in Zero-Shot and Few-Shot scenarios, and tried to design the Cot to carry out multiple rounds of dialogue, guiding ChatGPT to better complete the tasks of AI and RI, but the final performance was not ideal as well.

CCL 2024

3 Model

3.1 Baseline Evaluation

The baseline model given in this task mainly guides LLM by building the Chain-of-Thought (CoT) prompting method (Wei et al., 2023), and tests the results of ChatGPT in the Zero-Shot and Few-Shot scenarios. The specific operation and effect are shown in the following figure:



Given a sentence: {input sentence} where the target word is "{target word}". According to my given sentence and the semantic scene of the target word in the sentence, find all arguments that match the candidate frame elements in the sentence, and complete the matching of arguments and frame elements.

[‘Agent’: ‘法国’]
[‘Theme’: ‘《劳动法典》第一卷’]
[‘Time’: ‘1910年’]
[‘Location’: ‘？ ’]

Given a sentence: {input sentence} where the target word is "{target word}". According to my given sentence and the semantic scene of the target word in the sentence, find all arguments that match the candidate frame elements in the sentence, and complete the matching of arguments and frame elements.
For example, the answer to sentence “中国人民银行日前公布的今年一季度金融统计资料表明，中国首季金融运行平稳。” is “['中国人民银行', '信息传递者'], ['日前', '时间'], ['今年一季度金融统计资料', '内容']]”.

[‘Announcer’: ‘法国’]
[‘Content’: ‘《劳动法典》第一卷’]
[‘Time’: ‘1910年’]

Figure 2: Example of comparison between the effects of Few-Shot and Zero-Shot

https://www.hanlp.com/

It is not difficult to see that compared with Zero-Shot input, The output of the model in Few-Shot scenario is much better, reducing the output of useless information and capturing the corresponding arguments and roles more accurately. The research results of (Li et al., 2023) also show that the accuracy rate of LLM for FI task can only reach 37% in the scenario without samples, while it is significantly increased to 53% in the scenario with samples, which is enough to show the importance of giving samples of corresponding target words for Chinese Frame Semantic Parsing tasks.

3.2 Model Construction

Through the effect analysis of the baseline model, we found that the **target word** is undoubtedly the key to identify the semantic framework, which directly affects the performance of LLM on FI task. Therefore, we build a **hierarchical index RAG system** (Chen et al., 2024) based on target words, which uses keyword information to filter out a certain amount of options, reduces the length of tokens, and avoids the decline of LLM reasoning ability caused by long tokens.

At the same time, because the Few-Shot scenario greatly improves the performance of LLM, the sample quality provided to LLM is also an important link to determine the performance of the model. Therefore, we first use the HanLP tool to segment the sample sentences to make the sentence structure clearer. Then, in order to make the ability of the model to identify each frame similar, we constructed a **balanced Few-Shot sample category**. For each target word category, we matched the nearest pieces of data as a Few-Shot, ensuring that each category of the target word had the same number of data as samples, and at the same time using BM25 (Robertson et al., 1994) to ensure that the selected data were the closest to the problem.

As for the specific principle of BM25 retrieval algorithm, we first analyze the morpheme of the sentence to generate morpheme q_i. In this study, we directly regard the process of word segmentation through hanlp as morpheme analysis, and each word segmentation is regarded as morpheme q_i. Then, for each search statement d, the correlation score of each morpheme q_i and d is calculated. Finally, the correlation score of q_i relative to d is weighted and summed to obtain the correlation score of the sentence and d. Its general formula can be written as below:

$$Score(Q,d)=\sum_{i}^{n} IDF(q_i)\cdot R(q_i,d)$$

Where Q is the desired statement, $IDF(q_i)$ indicates the weight of morpheme q_i in the set of attributive sentences, which is called Inverse Document Frequency (IDF). That is, when many sentences contain q_i, The discrimination of q_i is not high, so the importance of using q_i to judge correlation is low, the lower the $IDF(q_i)$. The specific weight calculation method is as follows:

$$IDF(q_i)=\log\frac{N-n(q_i)+0.5}{n(q_i)+0.5}$$

Where $R(q_i,d)$indicates the correlation score between morpheme q_i and sentence d. In this study, BM25 algorithm in elasticsearch* is applied, and its correlation calculation method is as follows:

$$R(q_i,d)=\frac{f_i\cdot(k_1+1)}{f_i+K}$$

$$K=k_1\cdot(1-b+b\cdot\frac{dl}{avgdl})$$

k_1 and b are set to 1.2 and 0.75 by default, f_i represents the frequency of morpheme q_i in Sentence d, $avgdl$represents the average length of all sentences.

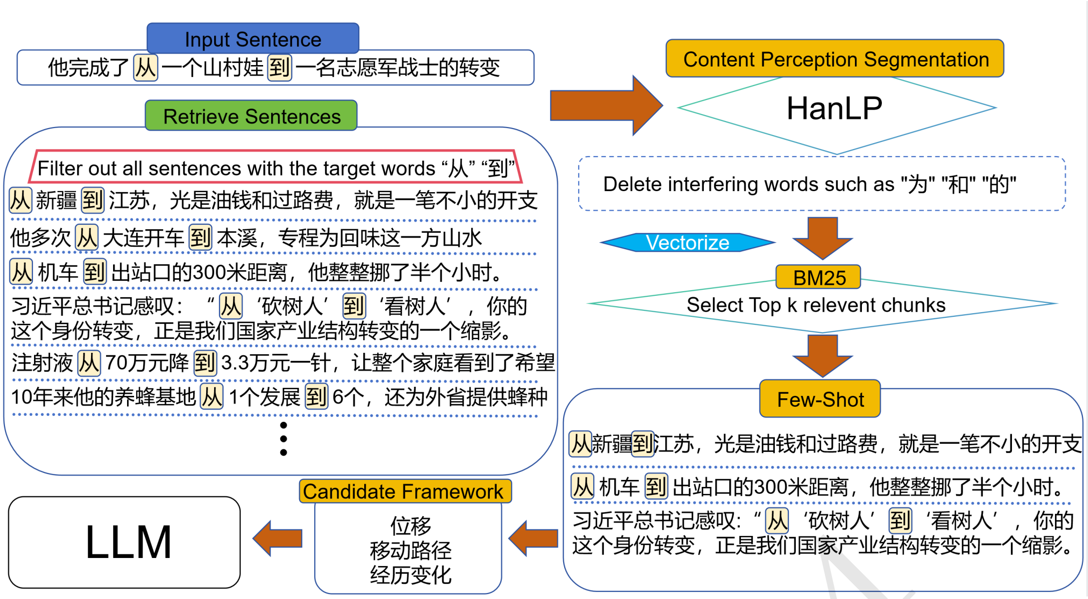
To sum up, the correlation score formula of BM25 algorithm in the study can be integrated as follows:

$$Score\left(Q,d\right)=\sum_{i}^{n} IDF\left(q_i\right)\cdot\frac{f_i\cdot(k_1+1)}{f_i+k_1\cdot(1-b+b\cdot\frac{dl}{avgdl})}$$

*https://www.elastic.co/cn/elasticsearch

CCL 2024

On this basis, the framework set corresponding to selected sentences is selected as a candidate frame for LLM to choose and judge. The overall RAG system construction process is as follows:


Input Sentence
他完成了从一个山村娃到一名志愿军战士的转变
Retrieve Sentences
Filter out all sentences with the target words “从” “到”
Content Perception Segmentation
HanLP
Delete interfering words such as "为" "和" "的"
Vectorize
BM25
Select Top k relevent chunks
Few-Shot
Candidate Framework
位移
移动路径
经历变化
LLM

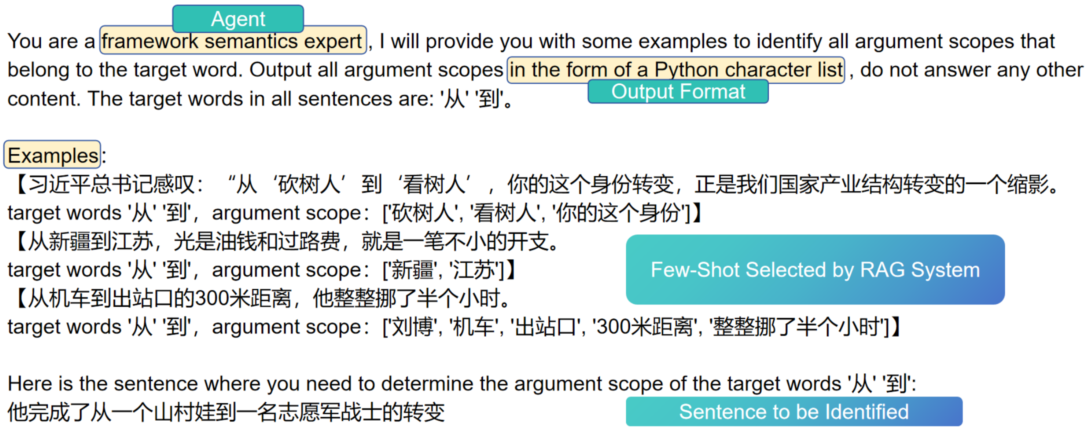
从新疆到江苏，光是油钱和过路费，就是一笔不小的开支
他多次从大连开车到本溪，专程为回味这一方山水
从机车到出站口的300米距离，他整整挪了半个小时。
习近平总书记感叹：“从‘砍树人’到‘看树人’，你的这个身份转变，正是我们国家产业结构转变的一个缩影。
注射液从70万元降到3.3万元一针，让整个家庭看到了希望
10年来他的养蜂基地从1个发展到6个，还为外省提供蜂种

Figure 3: Hierarchical indexing RAG system based on target words

CCL 2024

After the above operation, we noticed that about 13% of the data candidate frame information in the test set was empty, and the test found that the probability of the frame identified by LLM belonging to the given semantic frame was less than 4%. We made a special treatment for the target words of this part of Zero-Shot. By giving all the frame options to LLM for judgment, the probability of identifying the frame belonging to the given semantic frame increased to 94.6%.

For AI and FI tasks, we note that LLM is often not good at regular mapping, but is sensitive to semantics. It can effectively improve the performance of the model by handing over the mechanical work to pre-processing and post-processing. Therefore, based on the same construction of RAG system and high-quality Few-Shot samples, we change the input from the model to text, and then conduct post-processing to convert it back to the list, in order to reduce the computational load on the model, rather than waste its attention on the mapping relationship. We also incorporated the Agent features of LLM and limited the specific output format of LLM in prompts. The specific conversion process and prompt example are shown below:


Agent
Output Format
Few-Shot Selected by RAG System
Sentence to be Identified

You are a framework semantics expert, I will provide you with some examples to identify all argument scopes that belong to the target word. Output all argument scopes in the form of a Python character list, do not answer any other content. The target words in all sentences are: '从' '到'。

Examples:
【习近平总书记感叹：“从‘砍树人’到‘看树人’，你的这个身份转变，正是我们国家产业结构转变的一个缩影。
target words '从' '到'，argument scope：['砍树人', '看树人', '你的这个身份']】
【从新疆到江苏，光是油钱和过路费，就是一笔不小的开支。
target words '从' '到'，argument scope：['新疆', '江苏']】
【从机车到出站口的300米距离，他整整挪了半个小时。
target words '从' '到'，argument scope：['刘博', '机车', '出站口', '300米距离', '整整挪了半个小时']】

Here is the sentence where you need to determine the argument scope of the target words '从' '到':
他完成了从一个山村娃到一名志愿军战士的转变

Figure 4: Sample semantic enhancement conversion process example

4 Experiment

4.1 Experimental Setup

The following table shows the official data distribution of this evaluation (Yang et al., 2024). The experimental research in this paper is all based on the ChatGPT (gpt-3.5-turbo) model of Openai[†].

	Training set	Validation set	Test(B) set	Total
Sentences	10700(700)	2300(300)	4600(600)	17600(1600)
Frames	671(32)	354(24)	504(33)	695(86)
Frame Elements	947	649	796	987
Lexical Units	2359	670	572	3132

Table 2: CFN2.1 Dataset distribution

4.2 Evaluation Matrix

1. Frame Identification (FI)

   The only evaluation criterion for FI task is the accuracy:

$$FI_{acc} = \frac{correct}{total}$$

   Where $correct$ indicates the correct quantity predicted by the model, $total$ is the total number of frames to be identified.

2. Argument Identification (AI)

   AI task adopts precision (P), recall (R) and F1-Score ($F1$) were used as evaluation indexes.

$$AI_P = \frac{\text{InterSec}(\text{gold},\ \text{pred})}{\text{Len}(\text{pred})}$$

$$AI_R = \frac{\text{InterSec}(\text{gold},\ \text{pred})}{\text{Len}(\text{gold})}$$

$$AI_{F1} = \frac{2 \times AI_P \times AI_R}{AI_P + AI_R}$$

   Of which, $gold$ and $pred$ represent the real results and the predicted results respectively, $Intersec(*)$ means to calculate the number of tokens shared by both, $Len(*)$ indicates to calculate the number of tokens.

3. Role Identification (RI)

   RI task also use P, R and $F1$ as evaluation indexes.

$$RI_P = \frac{Count(gold \cap pred)}{Count(pred)}$$

$$RI_R = \frac{Count(gold \cap pred)}{Count(gold)}$$

$$RI_{F1} = \frac{2 \times RI_P \times RI_R}{RI_P + RI_R}$$

   Where $gold$ and $pred$ represent the real and predicted results respectively, $Count(*)$ indicates that the number of set elements is calculated.

[†] https://platform.openai.com/docs/models/gpt-3-5-turbo

CCL 2024

The final total evaluation score is the score obtained by weighting the three downstream tasks in the proportion of $(0.3, 0.3, 0.4)$, i.e

$$Score = 0.3 \times FI_{acc} + 0.3 \times AI_{F1} + 0.4 \times RI_{F1}$$

4.3 Evaluation Results and Analysis

The following table lists the scores of each evaluation index of the top three teams of the open track in this evaluation task:

Team ID	Score	AI_R	RI_R	AI_P	RI_P	FI_{acc}	AI_{F1}	RI_{F1}
Tangled	**48.77**	53.84	**38.66**	44.81	**43.97**	**58.62**	48.91	**41.14**
Our Team	40.12	**67.85**	19.52	52.18	14.53	52.54	**58.99**	16.66
UIR-MASTER	21.48	38.87	1.94	**66.64**	2.83	38.90	49.10	2.30

Table 3: CFN data distribution

It can be seen that our team's model has shown excellent results in AI task, and the effect in FI task is similar to that of the first team. Model's performance in RI tasks is poor, and the model needs to be further improved and optimized.

5 Other and Future Work

In terms of model optimization, we analyze and discuss the performance of LLM in Chinese Frame Semantic Parsing tasks. Combined with the relevant research of (Chen et al., 2024), we believe that we can try to use BAAI General Embedding (BGE) (Chen et al., 2024) for semantic embedding in subsequent research, and by using the complementary nature of different ranking algorithms, the Reciprocal Rank Fusion (RRF) (Cormack et al., 2009), which has excellent results in previous studies, can get better search results by mixing their ranking results, so as to obtain better data, and then improve the analytical ability of the model.

In addition, we used the Qwen1.5-7B model[‡] in the local validation phase to evaluate the first 1000 data sets in the test (B) set. Using the above-mentioned statement processing method, we have achieved an accuracy rate of 11.2%, and it is estimated that the accuracy rate of more than 45% can be obtained on the whole test (B) set. It can be seen that after reducing the difficulty of the task, LLM with smaller parameters can also perform well. In the future, we can try to use more open source LLM combined with fine-tuning operations in the corresponding fields to achieve better performance.

Finally, due to the limitation of funds and competition time, we only used the gpt-3.5-turbo model which released a year ago for experiments. We believe that the use of more advanced LLM (such as gpt-4o[§]) will bring considerable performance improvement in the future.

6 Conclusion

In this study, we have developed a hierarchical index Retrieval Augmented Generation (RAG) system based on target words to improve the efficiency and accuracy of Chinese Frame Semantic Parsing (CFSP). By leveraging target word information, our approach filters and indexes documents, which significantly enhances the performance of semantic parsing tasks. The key components of our system include data preparation, text tokenization using the HanLP tokenizer, index creation, data insertion into Elasticsearch (ES) indices, cluster configuration, and index population.

Our hierarchical indexing method ensures efficient and scalable retrieval, which is critical for handling large datasets in CFSP. The utilization of the BM25 retrieval algorithm further optimizes the quality of samples in Few-Shot scenarios, ensuring balanced and relevant data for the model. By refining the input and post-processing stages, we reduce the computational load on the model, allowing it to focus on semantic understanding rather than mapping relationships.

[‡] https://huggingface.co/Qwen/Qwen1.5-7B

[§] https://platform.openai.com/docs/models/gpt-4o

CCL 2024

The results of our implementation demonstrate a marked improvement in the retrieval and parsing processes, highlighting the effectiveness of our hierarchical index RAG system. This approach not only enhances the performance of existing models but also provides a scalable solution for future CFSP tasks. As we move forward, exploring the integration of more advanced language models and fine-tuning techniques will be essential to further optimize and refine our system.

Our work underscores the importance of efficient indexing and retrieval methods in natural language processing and sets the stage for future advancements in the field of Chinese Frame Semantic Parsing.

References

Ru Li, Yunxiao Zhao, Zhiqiang Wang, Xuefeng Su, Shaoru Guo, Yong Guan, Xiaoqi Han, Hongyan Zhao 2024. *A Comprehensive Overview of CFN From a Commonsense Perspective. Mach. Intell. Res*, 21, 239–256 (2024). https://doi.org/10.1007/s11633-023-1450-8.

Charles J. Fillmore. 1982. *Frame semantics[J]. Linguistics in the Morning Calm*, 1982:111-137.

You L, Liu T, Liu K. 2007. *Chinese FrameNet and OWL Representation[C]. Sixth International Conference on Advanced Language Processing and Web Information Technology (ALPIT 2007)*, IEEE, 2007: 140-145.

Liping You and Kaiying Liu. 2005. Building chinese framenet database. In *Natural Language Processing and Knowledge Engineering, 2005. IEEE NLP-KE' 05. Proceedings of 2005 IEEE International Conference on.*

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

Shaoru Guo, Ru Li*, Hongye Tan, Xiaoli Li, Yong Guan. 2020. *A Frame-based Sentence Representation for Machine Reading Comprehension[C]. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistic (ACL)*, 2020: 891-896.

Shaoru Guo, Yong Guan, Ru Li*, Xiaoli Li, Hongye Tan. 2020. *Incorporating Syntax and Frame Semantics in Neural Network for Machine Reading Comprehension[C]. Proceedings of the 28th International Conference on Computational Linguistics (COLING)*, 2020: 2635-2641.

Yong Guan, Shaoru Guo, Ru Li*, Xiaoli Li, and Hu Zhang. 2021. *Integrating Semantic Scenario and Word Relations for Abstractive Sentence Summarization[C]. Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2021: 2522-2529.

Yong Guan, Shaoru Guo, Ru Li*, Xiaoli Li, and Hongye Tan. 2021. *Frame Semantic-Enhanced Sentence Modeling for Sentence-level Extractive Text Summarization[C]. Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2021: 404-4052.

Hongyan Zhao, Ru Li*, Xiaoli Li, Hongye Tan. 2021. *CFSRE: Context-aware based on frame-semantics for distantly supervised relation extraction[J]. Knowledge-Based Systems*, 2020, 210: 106480.

Zhiqiang Wang, Ru Li, Jiye Liang, Xuhua Zhang, Juan Wu, Na Su. 2016. *Jiyu hanyu pianzhang kuangjia yuyi fenxi de yuedu lijie wenda yanjiu. Jisuanji xuebao*, 39(4):13.

Juncai Li, Zhichao Yan, Xuefeng Su, Boxiang Ma, Peiyuan Yang1, Ru Li. 2023. *Overview of CCL23-Eval Task 1:Chinese FrameNet Semantic Parsing. Proceedings of the 22nd Chinese National Conference on Computational Linguistics (Volume 3: Evaluations)*, pages 113–123, Harbin, China. Chinese Information Processing Society of China.

K. M. Hermann, D. Das, J. Weston, K. Ganchev. 2014. *Semantic frame identification with distributed word representations[C]. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2014: 1448-1458.

Shuangtao Yang, Mao Teng, Xiaozheng Dong, Fu Bo 2023. *Llm-based sparql generation with selected schema from large scale knowledge base[C]. China Conference on Knowledge Graph and Semantic Computing*. Singapore: Springer Nature Singapore, 2023: 304-316.

Stephen E. Robertson, Steve Walker, Susan Jones, Micheline Hancock-Beaulieu, and Mike Gatford. 1994. *Okapi at TREC-3[J]. Proceedings of the Third Text REtrieval Conference (TREC 1994)*, Gaithersburg, USA, November 1994.

CCL 2024

Jianlv Chen, Shitao Xiao, Peitian Zhang, Kun Luo, Defu Lian, Zheng Liu. 2024. *BGE M3-Embedding: Multi-Lingual, Multi-Functionality, Multi-Granularity Text Embeddings Through Self-Knowledge Distillation*. *arXiv preprint*, arXiv: 2402.03216.

G. V. Cormack, C. L. A. Clarke, S. Buettcher. 2009. *Reciprocal rank fusion outperforms condorcet and individual rank learning methods[C]*. *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, 2009: 758-759.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, Denny Zhou. 2023. *Chain-of-Thought Prompting Elicits Reasoning in Large Language Models*. *Advances in neural information processing systems*, 35, 24824-24837.

Xinyue Chen, Pengyu Gao, Jiangjiang Song, Xiaoyang Tan. 2024. *HiQA: A Hierarchical Contextual Augmentation RAG for Massive Documents QA*. *arXiv preprint*, arXiv: 2402.01767.

S. W. Kim, J. M. Gil. 2019. *Research paper classification systems based on TF-IDF and LDA schemes*. *Human-centric Computing and Information Sciences*, 9, 1-21.

Xuefeng Su, Ru Li, Xiaoli Li, Jeff Z. Pan, Hu Zhang, Qinghua Chai, and Xiaoqi Han. 2021. *A Knowledge-Guided Framework for Frame Identification*. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 5230–5240.

Yang Peiyuan and juncai li and Zhichao Yan and Xuefeng Su and Ru Li 2024. *Overview of CCL24-Eval Task1 Chinese Frame Semantic Parsing(CFSP) Evaluation Task*. *Submitted to The 23rd China National Conference on Computational Linguistics (Evaluation Workshop)*, 2024, https://openreview.net/forum?id=ObthsPZyah.

CCL 2024

(c) Technical Committee on Computational Linguistics, Chinese Information Processing Society of China