System Report for CCL25-Eval Task 2
Solving Frame Semantic Parsing with LLMs

Du Jingtao

Lianyungang Daily newspaper

`1006391414@qq.com`

Abstract

Frame Semantic Parsing (FSP) is a critical task in natural language processing (NLP) that involves identifying semantic frames, argument spans, and their corresponding roles within a sentence. This paper presents a novel approach to Chinese Frame Semantic Parsing by fine-tuning the Qwen3 large language model to simultaneously address three sub-tasks: Frame Identification, Argument Identification, and Role Identification. We propose a unified prompt-based framework with iterative refinements, including direct argument output for span identification and a majority-voting mechanism for frame prediction. Our experiments demonstrate significant improvements in argument and role identification through modified output formats, while frame identification benefits from ensemble voting. However, integrating Chain-of-Thought (CoT) reasoning with model-generated explanations yielded suboptimal results, suggesting limitations in the auxiliary model's performance. This work highlights the potential of fine-tuned large language models for complex semantic parsing tasks and identifies avenues for further optimization.

1 Introduction

Frame Semantic Parsing (FSP) is a pivotal task in NLP that aims to extract structured semantic representations from text by identifying frames, argument spans, and their semantic roles(Guan et al., 2021a; Guan et al., 2021b). In Chinese, FSP is particularly challenging due to the language's lack of explicit morphological markers and complex syntactic structures. The Chinese Frame Semantic Parsing task, in this competition, comprises three sub-tasks: (1) Frame Identification, which identifies the semantic frame evoked by a target word; (2) Argument Identification, which determines the spans of arguments associated with the target word; and (3) Role Identification, which assigns semantic role labels to identified arguments. Recent advancements in large language models (LLMs) have shown promise in tackling multi-task NLP problems. In this work, we leverage the Qwen3 model, fine-tuned for Chinese FSP, to address all three sub-tasks simultaneously. Our approach evolves through iterative prompt engineering and output format modifications, achieving notable performance improvements. Specifically,

Proceedings of the 24th China National Conference on Computational Linguistics, pages 70-75, Jinan, China, August 11-14, 2025.

(c) Technical Committee on Computational Linguistics, Chinese Information Processing Society of China   70

we enhance Argument Identification by directly outputting argument spans as text, improve Frame Identification using a majority-voting mechanism, and explore Chain-of-Thought (CoT) reasoning to incorporate frame and role definitions. While the first two modifications yield significant gains, CoT integration underperforms, highlighting the need for further investigation. This paper details our methodology, evaluates its performance, and discusses implications for future research.

2 Related Work

Frame Semantic Parsing (FSP) is grounded in FrameNet(Fillmore et al., 2003), a resource that defines semantic frames and their associated roles. In the context of Chinese, early approaches to FSP predominantly relied on rule-based systems and statistical models. These methods, while foundational, often struggled with scalability and adaptability to the nuances of the Chinese language. The advent of neural architectures and Large Language Models (LLMs) marked a significant shift, introducing models capable of capturing complex semantic relationships(Yan et al., 2024).

The integration of LLMs into Chinese FSP has further propelled the field forward. Studies have demonstrated the efficacy of LLMs like Gemini 1.0 in CFSP tasks, employing techniques such as mapping and similarity strategies to streamline frame identification, and leveraging data augmentation and model ensemble methods to enhance argument and role identification . Additionally, approaches utilizing non-fine-tuned LLMs with prompt engineering have shown promise, achieving competitive results in CFSP evaluations(Yahui et al., 2024).

Building upon this trajectory, our work leverages Qwen3(Yang et al., 2025), a multilingual LLM supporting 119 languages and dialects, including Chinese . We fine-tune Qwen3-8B and Qwen3-14B for the CFSP task, introducing novel prompt designs that effectively guide the model in frame and role identification. Furthermore, we implement ensemble strategies that combine outputs from multiple model instances, enhancing the robustness and accuracy of semantic parsing.

3 Methodology

We propose a unified approach to Chinese FSP using a fine-tuned Qwen3 model, addressing Frame Identification, Argument Identification, and Role Identification in a single pipeline. Below, we describe the task setup, initial prompt design, and iterative improvements.

3.1 Task Description

The Chinese FSP task involves three sub-tasks:

1. Frame Identification: Given a sentence, its ID, and a target word, identify the semantic frame evoked by the target word (e.g., "揭发" → "举报")

2. Argument Identification: Identify the spans of arguments associated with the target word in the sentence (e.g., "当地群众" in "当地群众纷纷起来联名向官府揭发刘方济利用传教为非作歹的丑恶罪行。")

3. Role Identification: Assign semantic role labels to identified arguments based on the frame (e.g., "当地群众" → "告发者")

3.2 Initial Prompt Design

Our initial prompt instructed the Qwen3 model to perform all three sub-tasks simultaneously:

你是汉语框架语义解析大师。

子任务 1: 框架识别（Frame Identification），识别句子中给定目标词激活的框架。

子任务 2: 论元范围识别（Argument Identification），识别句子中给定目标词所支配论元的边界范围。

子任务 3: 论元角色识别（Role Identification），预测子任务 2 所识别论元的语义角色标签。

原句：餐饮业是天津市在海外投资的重点之一。

分词后的句子：餐饮业　是　天津市　在　海外　投资　的　重点　之一　。

目标词 1：是，目标词词性:v。

请帮我预测给定目标词激活的框架、给定目标词所支配论元的边界范围和所识别论元的语义角色标签

---

the output of the above prompt is:

属于某类

0　　2

4　　14

3.3 Modified Output: Direct Argument Output

To address the poor performance in Argument Identification, we modified the argument spans as text (e.g., "餐饮业" instead of $[0, 2]$) rather than numerical indices. This change leveraged Qwen3's strength in text generation, significantly improving the accuracy of span and role identification. The prompt maintained the same and we only change the output format:

属于某类

餐饮业

天津市在海外投资的重点

3.4 Majority Voting for Frame Identification

Frame Identification performance lagged behind state-of-the-art results. To address this, we implemented a majority-voting mechanism, generating 16 candidate frame predictions per sentence and selecting the most frequent frame. This ensemble approach improved frame prediction accuracy by leveraging the model's diverse outputs.

3.5 Chain-of-Thought (CoT) Integration

To further enhance performance, we experimented with CoT reasoning by prompting Qwen3 14B model to generate short reasons for 3 subtasks. The CoT prompt is:

> 你是汉语框架语义解析大师。
> 子任务 1: 框架识别（Frame Identification），识别句子中给定目标词激活的框架。
> 子任务 2: 论元范围识别（Argument Identification），识别句子中给定目标词所支配论元的边界范围。
> 子任务 3: 论元角色识别（Role Identification），预测子任务 2 所识别论元的语义角色标签。
> 原句：餐饮业是天津市在海外投资的重点之一。
> 分词后的句子：餐饮业　天津市　在　海外　投资　的　重点　之一　。
> 目标词 1: 是, 目标词词性:v。
> 请帮我预测给定目标词激活的框架、给定目标词所支配论元的边界范围和所识别论元的语义角色标签
> 结合框架定义: 该框架表示某个实体属于某个范畴，（通常暗含）指某个认知者认为的归类。以及语义角色定义: 实体: 指某个具体范畴的实体实例。范畴: 该框架元素表示某个实体所属的一种概括性的类型或者类别。
> 来推断出答案为什么是:
> 属于某类
> 餐饮业
> 天津市在海外投资的重点
> 推理原因三个任务各一句话概括。

However, this approach underperformed, potentially due to only one sentence for each subtask or misalignment with models fine-tuning.

4 Experiments

4.1 Dataset and Evaluation

We evaluated our approach on the dataset from the Third Chinese Frame Semantic Parsing Evaluation. Performance was measured using standard metrics: accuracy for Frame Identification and F1 scores for Argument and Role Identification.

4.2 Results

- Initial Prompt: The baseline model struggled with numerical span outputs, resulting in low F1 scores for Argument Identification (F1 0.42) and Role Identification (F1 0.03) in leaderboard A

- Direct Argument Output: Switching to text-based span outputs significantly improved Argument Identification (F1 0.79) and Role Identification (F1 0.51) in leaderboard A, as the model better captured textual boundaries.

- Majority Voting: The voting mechanism boosted Frame Identification accuracy from 0.55 to 0.72 in Leaderboard B with Qwen3-8B, as we changed to Qwen3-14B in last submit, the score even boosted to 0.74 in Leaderboard B.

- CoT Integration: CoT reasoning slightly degraded performance across all sub-tasks (Frame Accuracy 0.73, Argument F1 0.83, Role F1 0.54 in leaderboard B).

4.3 Discussion

The direct argument output modification highlights the importance of aligning LLM output formats with their strengths in text generation. Numerical indices posed challenges for precise boundary detection, whereas text spans leveraged Qwen3's contextual understanding. The majority-voting mechanism effectively mitigated frame prediction errors, suggesting that ensemble techniques can enhance LLM robustness in semantic parsing. However, the CoT approach' s failure underscores the dependency on high-quality reasoning models and the need for better integration strategies. Potential issues include Qwen' s limited reasoning capacity or mismatches between generated explanations and true knowledge.

5 Conclusion and Future Work

This paper presents a unified approach to Chinese Frame Semantic Parsing using a fine-tuned Qwen3 model. Our iterative refinements—direct argument output and majority voting —significantly improved performance across sub-tasks, demonstrating the potential of LLMs in complex semantic parsing. However, CoT integration requires further exploration, particularly with higher-quality reasoning models or improved prompt designs. Future work will focus on optimizing CoT strategies, exploring alternative ensemble methods, and extending the approach to other languages and semantic parsing tasks.

References

Charles J. Fillmore, Christopher R. Johnson, and Miriam R.L. Petruck. 2003. Background to framenet. *International Journal of Lexicography*, 16(3):235–250, 09.

Proceedings of the 24th China National Conference on Computational Linguistics, pages 70–75, Jinan, China, August 11-14, 2025.

(c) Technical Committee on Computational Linguistics, Chinese Information Processing Society of China   74

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*, pages 4045–4052, Online and Punta Cana, Dominican Republic, November. 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*, pages 2522–2529, Online and Punta Cana, Dominican Republic, November. Association for Computational Linguistics.

Liu Yahui, Gong Chen, and Zhang Min. 2024. Leveraging LLMs for Chinese frame semantic parsing. In Hongfei Lin, Hongye Tan, and Bin Li, editors, *Proceedings of the 23rd Chinese National Conference on Computational Linguistics (Volume 3: Evaluations)*, pages 21–31, Taiyuan, China, July. Chinese Information Processing Society of China.

Zhichao Yan, Jiapu Wang, Jiaoyan Chen, Xiaoli Li, Ru Li, and Jeff Z Pan. 2024. Atomic fact decomposition helps attributed question answering. *arXiv preprint arXiv:2410.16708*.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, Chujie Zheng, Dayiheng Liu, Fan Zhou, Fei Huang, Feng Hu, Hao Ge, Haoran Wei, Huan Lin, Jialong Tang, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiaxi Yang, Jing Zhou, Jingren Zhou, Junyang Lin, Kai Dang, Keqin Bao, Kexin Yang, Le Yu, Lianghao Deng, Mei Li, Mingfeng Xue, Mingze Li, Pei Zhang, Peng Wang, Qin Zhu, Rui Men, Ruize Gao, Shixuan Liu, Shuang Luo, Tianhao Li, Tianyi Tang, Wenbiao Yin, Xingzhang Ren, Xinyu Wang, Xinyu Zhang, Xuancheng Ren, Yang Fan, Yang Su, Yichang Zhang, Yinger Zhang, Yu Wan, Yuqiong Liu, Zekun Wang, Zeyu Cui, Zhenru Zhang, Zhipeng Zhou, and Zihan Qiu. 2025. Qwen3 technical report.