System Report for CCL24-Eval Task 1: Leveraging LLMs for Chinese Frame Semantic Parsing

Yahui Liu, Chen Gong, Min Zhang
School of Computer Science and Technology, Soochow University, Suzhou, China
yahuiliu.nlp@foxmail.com
{gongchen18, minzhang}@suda.edu.cn

Abstract

We participate in the open track of the Chinese frame semantic parsing (CFSP) task, i.e., CCL24-Eval Task 1, and our submission ranks first. FSP is an important task in Natural Language Processing, aiming to extract the frame semantic structures from sentences, which can be divided into three subtasks, e.g., Frame Identification (FI), Argument Identification (AI), and Role Identification (RI). In this paper, we use the LLM Gemini 1.0 to evaluate the three subtasks of CFSP, and present the techniques and strategies we employed to enhance subtasks performance. For FI, we leverage mapping and similarity strategies to minimize the candidate frames for each target word, which can reduce the complexity of the LLM in identifying the appropriate frame. For AI and RI subtasks, we utilize the results from small models as auxiliary information and apply data augmentation, self-training, and model ensemble techniques on these small models to further enhance the performance of subtasks.

1 Introduction

Chinese Frame Semantic Parsing (CFSP) is a fine-grained semantic parsing method based on Chinese FrameNet (CFN) (You and Liu, 2005; Li et al., 2023a), which is first proposed during CCL2003-Eval (Li et al., 2023a). It aims to extract frame-semantic information from a sentence (Wang et al., 2020). Previous works have shown that CFSP can help various downstream tasks, including text summarization (Guan et al., 2021a; Guan et al., 2021b), reading comprehension (Guo et al., 2020a; Guo et al., 2020b), relation extraction (Zhao et al., 2020), etc.

CFN, proposed by Shanxi University, is constructed based on the theory of Frame Semantics, with English FrameNet (Fillmore et al., 2003) as a reference and Chinese realistic corpora as the data foundation (Li et al., 2024). It is a structured representation of knowledge that establishes relationships between vocabulary and concepts through frames. In the CCL2023-Eval CFSP task, part of the CFN data is released for the first time. Compared to the previous evaluation, this evaluation adds information on constructional target words. Constructional target words refer to the words within a construction that are considered central or core. These words play a crucial role in activating the meaning of the entire construction. A construction is a fixed expression unit in a language that has specific form and meaning (Boas, 2021; Willich, 2022). It can be a single word, a phrase, or even a sentence. For example, in the phrase "爱买不买(love to buy or not)", the construction is "爱+V+不+V (love + V + not +V)", where "V" represents a verb. Table 1 provides an example of the frame "量变" (Change position on a scale). The concept expressed by this frame is the relative positional change of a certain attribute of an entity, and its lexemes (also called target word) are "提高(increase)", "从...到... (from...to)" and so on. The word "提高" is a general target word, while "从...到" is a constructional target word. The frame elements are used to capture the semantic information related to the frame in a sentence. Different frame elements are assigned different meanings, for example, "初值(Initial value)" means the starting point of the entity's attribute value change.

Frame Name	量变(Change position on a scale)	
Frame Definition	该框架表示实体在某个维度上（即某属性）的相对位置发生变化，其属性值从初值变至终值 (The framework represents the relative change of an entity along a certain dimension (i.e., attribute), with its attribute value transitioning from an initial value to a final value.)	
Frame Elements	**fe_name**	**fe_def**
	实体(Entity)	在某属性上具有一定量值的事物(Something with a certain quantity value on an attribute)
	属性(Attribute)	实体的有数量变化的属性(The entity's attributes with changing quantities)
	初值（Initial value)	实体的属性值变化的起点(The starting point of the entity's attribute value change)
	终值(Final value)	实体最后达到的量值(The final quantity value reached by the entity)
	初始状态(Initial state)	实体经历属性值变化之前的状态(The state of the entity before undergoing changes in attribute values)
	终状态(Final state)	实体经历属性值的变化之后所达到的状态(The state reached by the entity after undergoing changes in attribute values)
	变幅(Difference)	实体在某维度上变动的幅度(The extent of the entity's variation along a certain dimension)
	值区间(Value range)	属性值的变动范围(The range of variation in attribute values)

Table 1: An example about "量变"(Change position on a scale) frame in Chinese FrameNet.

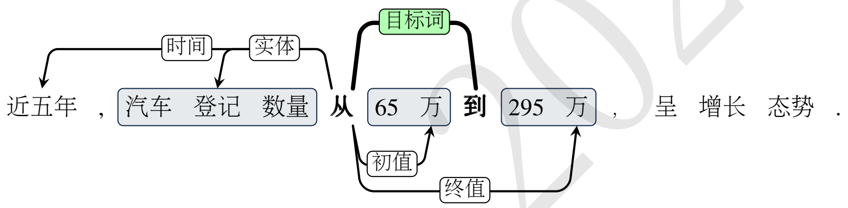

Figure 1: An example of CFSP.

This evaluation divides CFSP task into three subtasks: Frame Identification, Argument Identification and Role Identification. **Frame Identification (FI)** aims to find the corresponding frame for the target word in all frames according to its meaning in the sentence. Frame elements, also referred to as arguments, can be a single word or a span in the sentence. The goal of **Argument Identification (AI)** is to locate all arguments within the sentence and determine their boundaries. The purpose of **Role Identification (RI)** is to assign appropriate semantic role labels to the identified arguments. Taking Figure 1 as an example, the constructional target word in the sentence "近五年，汽车登记数量从65万到295万，呈增长态势(Over the past five years, the number of registered cars has increased from 0.65 million to 2.95 million, showing a growth trend)" is "从...到", which corresponds to the frame "量变(Change position on a scale)" in table 1. In this sentence, there are multiple arguments related to this frame, i.e., "近五年", "汽车登记数量", "65万", "295万". "时间", "实体", "初值", and "终值" are their corresponding semantic role labels, respectively.

In this CFSP evaluation task, there are two tracks: open and closed. The open track allows the use of LLM such as ChatGPT for inference. In the open track, we leverage the LLM Gemini (Team et al., 2023) released by Google to evaluate three subtasks of CFSP. For FI subtask, we provide a small number of relevant frames for each target instead of all frames for LLM to select from. For AI and RI subtasks, we enhance the performance of Gemini on CFSP based on the results of frame identification and argument identification from small models, respectively. All of our datasets and codes are available at `https://github.com/yahui19960717/CCL2024-CFSP-LLM.git`.

2 Related works

CCL2023-Eval (Li et al., 2023a) propose CFSP task based on CFN for the first time. The existing works of CFSP mainly use neural network-based methods (Li et al., 2023b; Huang et al., 2023; Liu et al., 2023; Guan et al., 2023), where Li (2023b) propose a method based on rotational positional encoding (Su et al., 2021) to calculate the attention signal between entities, which has a significant effect on FI subtask. Huang (2023) use a variety of optimization strategies during training to improve the robustness of the model, for example, Exponential Moving Average (EMA) and Fast Gradient Method (Miyato et al., 2016). For FI subtask, Liu (2023) transform CFSP into a word-based graph parsing task, identifying the frame of the target word together with the corresponding arguments in an end-to-end framework. For AI and RI subtasks, they transform CFSP into a tree parsing task to enhance the ability to identify argument boundaries and roles by modeling the internal structure of arguments. Guan (2023) attempt various pre-trained models for different sub-tasks and explore multiple approaches to solving each task from the perspectives of feature engineering, model structure, and other tricks.

In the era of LLM, the performances of LLM on most NLP tasks are still well below the supervised baselines (Sun et al., 2023). Li (2023a) test the ability of the LLM on different subtasks of CFSP using ChatGPT (gpt-3.5-turbo-16k) (Brown et al., 2020). They construct different prompts for the three subtasks on part of the test set and guide the model to generate more reliable results by designing the chain-of-thought. Finally, they find that the performance of ChatGPT on the three sub-tasks of CFSP is not ideal.

For this, we propose different strategies to enhance LLM performance in CFSP task. Given the restriction on the token number in an LLM prompt, for FI subtask, we provide relevant frames in the LLM prompt instead of all frames as candidate frames for each target word. The AI subtask is based on the FI subtask, and RI is a further refinement of the argument. Considering that FI, AI and RI are three interrelated subtasks and the performance of the small model is better than LLM in these subtasks, we utilize the results of small models to parse AI and RI subtasks.

3 Methods

Prompting is a recently-mainstream for LLM and LLM-friendly evaluation method (Zhou et al., 2023). The performance of a task is highly related to the information contained in the prompt. For this, we design appropriate prompts for the three subtasks.

For FI subtask, we need to provide candidate frames for each target word in the prompt. To mitigate the issue of excessively long prompts, we obtain a small number of candidate frames for each word based on the mapping relationships between target words and frames in the given training and dev data. After providing the target word and its triggered frame in the prompt, we can parse the AI task using LLM. Similarly, after providing the target word and its corresponding arguments in the sentence in the prompt, we can parse the RI task. Due to the relatively low performance of the LLM in the AI and FI subtasks, we use small models to obtain FI and AI results, which are then used to assist AI and RI subtasks, respectively. To enhance the performance of the small models, we employ data augmentation techniques (including copying and generating data) for the FI task and self-training methods for the AI task. In both tasks, we adopt model fusion techniques (i.e., multiple model voting) to further improve the model performance.

3.1 Frame Identification

The FI subtask is to identify the frames triggered by the given target word in the sentence from a total of 715 provided frames. If we treat all frames as candidates in the prompt for each target word, the performance of FI subtask will be low (Gemini 1.0 obtains an accuracy of only 15.5% on FI for testA). We suspect that this is probably because it exceeds the maximum number of tokens limit imposed by Gemini in a prompt. The maximum number of tokens allowed in a prompt for Gemini 1.0 is 2048. However, just the length of the Chinese name of all the frames connected by commas is 2940, which is too long for a single prompt.

Stratigies of FI subtask. For this, we reduce the number of candidate frames for each target word. Specifically, we identify all corresponding frames for each target word through the mapping between words and frames in the dataset (there are a total of 4,093 target words with corresponding candidate frames in the testB). For target words without an existing mapping (507 target words in the testB), we compute the cosine similarity between the word vectors of the target words and each frame, selecting the six most similar frames as candidates. We chose six because the average number of frames per target word in the dataset is six. Some frame names are long and not tokenized, such as the frame "事件发生时间变化(Change event time)". In such cases, we use Jieba for tokenization. The word vectors we use are from FastText [0], but some target words or tokenized frame names might not be in FastText. In these instances, we sum the word vectors of each character from the tokenization to obtain the final word vector representation.

FI prompt. After obtaining the candidate frames for each target word, we designed appropriate task formats to accommodate the generative characteristics of LLM. Figure 2 shows an example of a prompt we provided. In the prompt, we present the LLM with a sample, and the model selects the most suitable frame based on the sample, the input text, the target word, and the candidate frames for the target word.

给定句子和其目标词,根据目标词在句子中的语义，从框架集合中选择最符合目标词触发的框架。请以字符串的形式输出。
示例:
输入:
text: "双方还就各自的责任、权利以及资金的运作程序等达成了协议。"
target:"达成"
框架集合："['观点一致', '成就']"
输出:
"观点一致"
给定输入:
text:"很快，他完成了从一个山村娃到一名志愿军战士的转变"
target:"从...到"
框架集合:"['事件时量发生变化', '时量场景', '量变', '经历变化', '移动路径', '位移']"
请输出其合适的目标词。

Figure 2: An example prompt for the FI subtask.

3.2 Argument Identification

The AI subtask refers to identifying the arguments relevant to specific semantic frames triggered by the target word in a sentence. Therefore, it is necessary to first obtain the right frames for target words. However, the FI performance of Gemini is relatively poor, and errors in FI propagate to the AI subtask. We utilize the FI results of the small model as the basis for the AI subtask.

Stratigies of AI subtask. For the small model, we adopted the models proposed by Li (2023b) (A sequence labeling model based on rotation position encoding) and Liu (2023) (A word-based graph parsing model). In the training set, we observe a severe imbalance in the distribution of frames, as illustrated in Figure 3. 76 frames had no corresponding examples, while 14 frames had over 100 corresponding examples. To improve the performance of the small model, we utilize the LLM to generate examples for frames without any examples. For frames with fewer than 15 examples, we augment the existing examples by duplicating them to reach a total of 15 examples. Figure 4 illustrates the prompt we used for sentence generation with the large model. Specifically, we train the models using the original training

[0]https://fasttext.cc/docs/en/crawlvectors.html

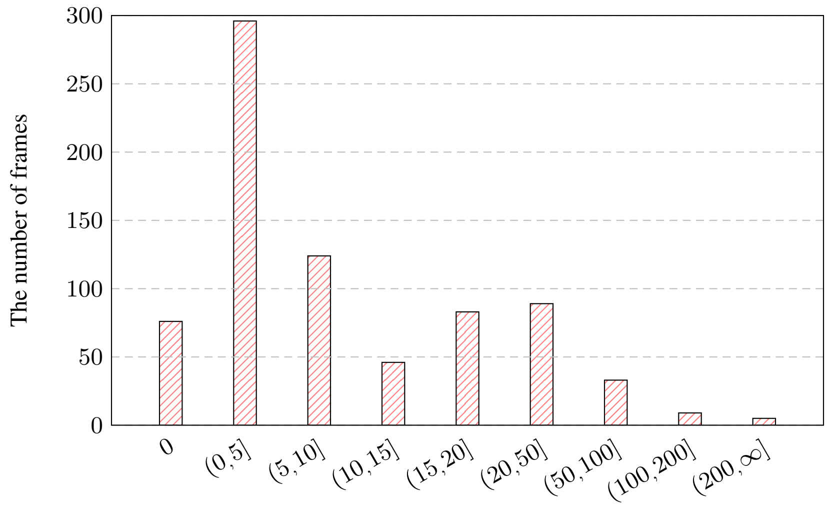

Figure 3: The distribution of the number of example sentences corresponding to each frame.

data	train	dev	testB	frames
CFN	10,700	2,300	4,600	715

Table 2: Data statistic of CFN.

data along with the constructed data. Each model is trained three times, and the final results are obtained through voting.

AI prompt. Finally, the performance of the FI task on the small model is 72%. After obtaining the FI results, we design an appropriate prompt to use Gemini to parse the AI task. Figure 5 provides an example of our prompt.

3.3 Role Identification

Stratigies of RI subtask. The RI subtask aims to assign semantic labels to the arguments identified in the sentence that are related to the frame. Therefore, we first need to identify all the arguments in the sentence. We use the small model to obtain more accurate AI results for RI subtask. Like the AI subtask, we also use the models proposed by Li (2023b) and Liu (2023). To enhance the performance of the small model, we employ the self-training and ensemble techniques. Specifically, we first use a trained model to predict the data from CoNLL09 and CPB1.0. Then, we used the predicted data to train a new model. This model is then fine-tuned on the training data. The final argument result is selected by voting, and the performance of the AI task on the small model is 84%.

RI prompt. After obtaining the AI results, we design an appropriate prompt to use Gemini for parsing the RI task. Figure 6 provides an example of our prompt.

4 Experiments

4.1 Data and Settings.

Table 2 shows the data statistic of CFN. For the AI subtask, we generated 1,509 of data using Gemini 1.0 and constructed 4,688 of data using the copying method. For the RI subtask, we provided a total of 124,270 of data from CTB1.0 and CoNLL09 for self-training.

For the small model, we use the model settings of Li (2023b) and Liu (2023). For the FI subtask, after voting by 6 models, we selected the results with votes greater than or equal to 2 as the FI results. Similarly, we consider results with 2 or more votes as the final AI results. As for the RI subtask, we find that results with 3 or more votes are more reliable.

使用所提供的框架和框架信息，生成与之相关的20个句子，并标识出框架中的目标词(谓词），以json格式输出。
框架信息如下：
框架名称:盗窃
框架定义:以非法占有为目的，秘密窃取数额较大的公私财物或者多次盗窃公私财物的行为。
示例：
输入：
框架名称：到达
框架定义：指转移体朝目的地方向的移动。目的地可直接表达出来，或从上下文中得到理解，动词本身隐含目标之义。
输出2个句子：
{’text’: [’运动会’, ’闭幕’, ’后’, ’，’, ’他们’, ’将’, ’在’, ’北京’, ’继续’, ’逗留’, ’两’, ’天’, ’，’, ’同’, ’中国’, ’有关’, ’方面’, ’开展’, ’交流’, ’活动’, ’并’, ’参观’, ’游览’, ’，’, ’于’, ’１３日’, ’返回’, ’日本’, ’。’], ’target’: [’返回’], ’text’: [’迎接’, ’澳门’, ’回归’, ’系列’, ’图书’, ’出版’], ’target’: [’回归’]}

Figure 4: An example prompt for the AI subtask.

Organizer	Score
SUDA	48.77
DVTC	40.12
UIR-ISC	21.48

Table 3: Reproduction results of the top three teams.

The LLM settings are configured as follows for all subtasks. The primary model utilized is Gemini 1.0, specified by the parameter model = “gemini-pro”. For the generation process, the following configuration is applied: “temperature” is set to 0, “top_p” to 0.95, and “top_k” to 0. Other settings remained at their default values.

4.2 Evaluation Metrics.

We use the evaluation metrics provided by the official. The FI subtask employs accuracy as the metric, while the AI and RI subtasks adopt precision, recall, and F1 scores as measures. The formulas are as follows:

$$\text{ACC} = \frac{\text{the right number of frames}}{\text{Total number of frames}} \tag{1}$$

$$P = \frac{(gold \cap pred)}{Count(pred)} \tag{2}$$

$$R = \frac{Count(gold \cap pred)}{Count(gold)} \tag{3}$$

$$F1 = \frac{2 * P * R}{P + R} \tag{4}$$

where gold and pred denote the correct and predicted results, respectively. For the AI subtask, $Count(*)$ represents the number of tokens in results, whereas it signifies the number of elements in the Argument Role Identification task.

作为汉语框架语义学家，你的任务是抽取给定句子中与目标词相关的所有框架论元成分，并根据框架的论元定义为它们分配相应的论元标签。请注意，论元成分是指句子中与目标词直接相关的部分，而论元标签描述了这些成分的具体角色。
输出格式：每个输出元素应该是一个列表，包含两个字符串元素：
一个是论元成分,一个是论元标签，即[[论元成分1，论元标签1],[论元成分2，论元标签2],...]。
只需要以Python数组的形式输出，不要添加其他内容。
文本：长寿区生态环境局局长刘刚介绍，该区乡镇污水收集处理率由原来的不足30%提升到现在的80%。
目标词：“由...到”
框架名称:量变
论元标签：
{'实体', '属性', '初值', '终值', '初状态', '变幅', '值区间', '环境条件', '倚变因素', '时量', '倚变起点', '倚变终点', '方式', '路径', '处所', '速度', '时间', '终状态', '角色事件时间', '动作时量', '结果时量', '频率', '次数', '特定次数', '角色位置', '亚区', '接受者', '受益人', '受损者', '伴随实体', '身份', '形容', '事件评价', '背景事件', '并行事件', '相关变量', '原因', '目的', '工具', '材料', '根据', '方法', '物量', '角色范围', '程度', '媒介', '主题', '视角', '结果', '领域', '现象', '解释'}
示例：
输入：
文本：['运动会', '闭幕', '后', '，', '他们', '将', '在', '北京', '继续', '逗留', '两', '天', '，', '同', '中国', '有关', '方面', '开展', '交流', '活动', '并', '参观', '游览', '，', '于', '１３日', '返回', '日本', '。']
目标词：['返回']
框架名称:到达
论元标签：
{'终点', '转移体', '伴随体', '形容', '目的地状态', '方式', '方法', '移动模式', '路径', '起点', '时间', '方向', '角色事件时间', '动作时量', '结果时量', '频率', '次数', '特定次数', '处所', '角色位置', '亚区', '接受者', '受益人', '受损者', '伴随实体', '身份', '事件评价', '环境条件', '背景事件', '并行事件', '相关变量', '原因', '目的', '工具', '材料', '根据', '物量', '角色范围', '程度', '媒介', '主题', '视角', '结果', '领域', '现象', '解释'}
输出:
“[['他们', '转移体'], ['于１３日', '时间'], ['日本', '终点']]”

Figure 5: An example prompt for the AI subtask.

作为汉语框架语义学家，你的任务是根据给定的句子的目标词、框架以及目标词在句子中的相关论元成分，从候选的论元标签中为它们分配相应的论元标签。请注意，论元成分是指句子中与目标词直接相关的部分，而论元标签描述了这些成分的具体角色。
输出格式：每个输出元素应该是一个列表，包含两个字符串元素：
一个是论元成分,一个是论元标签，即[[论元成分1，论元标签1],[论元成分2，论元标签2],...]。
只需要以Python数组的形式输出，不要添加其他内容，如果没有相应的标签就返回"。
文本：长寿区生态环境局局长刘刚介绍，该区乡镇污水收集处理率由原来的不足30%提升到现在的80%。
目标词："由...到"
框架名称:量变
论元标签：
{'实体', '属性', '初值', '终值', '初状态', '变幅', '值区间', '环境条件', '倚变因素', '时量', '倚变起点', '倚变终点', '方式', '路径', '处所', '速度', '时间', '终状态', '角色事件时间', '动作时量', '结果时量', '频率', '次数', '特定次数', '角色位置', '亚区', '接受者', '受益人', '受损者', '伴随实体', '身份', '形容', '事件评价', '背景事件', '并行事件', '相关变量', '原因', '目的', '工具', '材料', '根据', '方法', '物量', '角色范围', '程度', '媒介', '主题', '视角', '结果', '领域', '现象', '解释'} 论元成分:
{'现在的80%'}
示例：
输入：
文本：['运动会', '闭幕', '后', '，', '他们', '将', '在', '北京', '继续', '逗留', '两', '天', '，', '同', '中国', '有关', '方面', '开展', '交流', '活动', '并', '参观', '游览', '，', '于', '１３日', '返回', '日本', '。']
目标词：['返回']
框架名称:到达
论元标签：
{'终点', '转移体', '伴随体', '形容', '目的地状态', '方式', '方法', '移动模式', '路径', '起点', '时间', '方向', '角色事件时间', '动作时量', '结果时量', '频率', '次数', '特定次数', '处所', '角色位置', '亚区', '接受者', '受益人', '受损者', '伴随实体', '身份', '事件评价', '环境条件', '背景事件', '并行事件', '相关变量', '原因', '目的', '工具', '材料', '根据', '物量', '角色范围', '程度', '媒介', '主题', '视角', '结果', '领域', '现象', '解释'}
论元成分：
{'他们', '北京', '于13日', '日本'} 输出:
"[['他们', '转移体'], ['北京',''], ['于１３日', '时间'], ['日本', '终点']]"

Figure 6: An example prompt for the RI subtask.

	FI	AI			RI			总分
	ACC	P	R	F1	P	R	F1	
SUDA (Gemini-w/ small model)	**58.62**	44.84	**53.91**	48.96	**44.09**	**38.75**	**41.24**	**48.77**
SUDA (Gemini-w/o small model)	58.62	**72.39**	40.76	**52.15**	21.49	14.05	16.99	40.03
baseline (ChatGPT)	53.00	60.98	22.52	32.90	6.38	7.59	6.93	28.54

Table 4: Comparison of results between ours and ChatGPT on the three sub-tasks of CFSP.

4.3 Results and Analysis.

Table 3 shows the final reproducibility scores of the top three teams on the testB dataset in the open track. It can be seen that we ranked first, with a lead of 8.77 points over the second place and 27.29 points over the third place.

Previous studies have also evaluated CFSP on LLM in CCL2003-Eval (Li et al., 2023a). They extracted a portion of samples from the test set, constructed different prompt information, and used ChatGPT (gpt-3.5-turbo-16k) to complete the corresponding subtasks. For the FI subtask, they tested the results of ChatGPT in ZeroShot and FewShot scenarios. For the AI and RI subtasks, they guided ChatGPT through multiple rounds of dialogue using a chain-of-thought prompting method to generate more reliable results. As shown in Figure 4, we compared our results with theirs, where their FI results are from the few-shot scenario. It can be seen that our score is 20% higher than theirs, proving the effectiveness of our techniques. However, we can observe that our improvement in the FI task is relatively small, possibly due to the incompleteness of the frame found for each target word using the mapping method. In addition, we find that the improvement in the RI sub-task is significant, indicating that the accuracy of argument identification is crucial for the RI task.

In addition, for AI and RI subtasks, we conducted experiments on Gemini without relying on small models, with results as shown in the major second row. We can find that the performance of RI is lower compared to when small models are used, which is as expected. This is because small models perform well in FI and AI tasks, where FI can provide labels for RI to use, and RI also needs to utilize the results from AI during evaluation. However, it is interesting that in the AI task, the F1 score is higher than when relying on the small model. Intuitively, using small models seems preferable because predicate argument identification occurs after determining frames. We speculate that the reason for the good performance in the AI subtask might be that AI has a weak dependence on results of FI and FI information in the prompt could potentially interfere with AI task generation. Additionally, we can observe that the AI results without using small models excel primarily in accuracy, reaching 72.39%.

5 Conclusion

In this evaluation task, we construct appropriate prompts in Gemini for the three subtasks of CFSP for parsing. For the FI task, we use the mapping between the target words and frames to obtain candidate frames for each target word. For the AI and RI subtasks, we leverage the results from the small model to enhance the prediction performance of the LLM. Ultimately, we achieve first place in the open track of the CFSP.

However, our work still has some notable limitations. For instance, in the FI subtask, the limited candidate frames we provided may constrain the choices of the LLM. On the closed track, the small model has achieved 72.82%, 86.97%, and 60.27% respectively for the three subtasks, indicating ample room for improvement for the large model in CFSP task. In the future, we can enhance the performance of the three subtasks by providing more examples or utilizing information provided by the framework, such as frame definitions and argument definitions. The output of the large model is diverse, and we can further enhance its performance by generating multiple prediction results and using voting mechanisms.

Proceedings of the 23rd China National Conference on Computational Linguistics, pages 21-31, Taiyuan, China, July 25 - 28, 2024.
Volume3: Evaluations

Acknowledgements

We thank the reviewers for their constructive feedback which helped to improve the paper. This work was supported by National Natural Science Foundation of China (Grant No.62306202), the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (Grant No.23KJB520034), and a Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions.

References

Hans C Boas. 2021. Construction grammar and frame semantics. In *The Routledge handbook of cognitive linguistics*, pages 43–77.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.

Charles J Fillmore, Christopher R Johnson, and Miriam RL Petruck. 2003. Background to framenet. *International journal of lexicography*, 16(3):235–250.

Yong Guan, Shaoru Guo, Ru Li, Xiaoli Li, and Hongye Tan. 2021a. Frame semantic-enhanced sentence modeling for sentence-level extractive text summarization. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 4045–4052.

Yong Guan, Shaoru Guo, Ru Li, Xiaoli Li, and Hu Zhang. 2021b. Integrating semantic scenario and word relations for abstractive sentence summarization. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 2522–2529.

Yingxuan Guan, Xunyuan Liu, Lu Zhang, Zexian Xie, and Binyang Li. 2023. System report for CCL23-eval task 3: UIR-ISC pre-trained language medel for Chinese frame semantic parsing. In *Proceedings of the 22nd Chinese National Conference on Computational Linguistics (Volume 3: Evaluations)*, pages 124–138.

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 *Proceedings of the 28th International Conference on Computational Linguistics*, pages 2635–2641.

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 *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 891–896.

Shutan Huang, Qiuyan Shao, and Wei Li. 2023. System report for CCL23-eval task 3: Chinese frame semantic parsing based on multi task pipeline strategy. In *Proceedings of the 22nd Chinese National Conference on Computational Linguistics (Volume 3: Evaluations)*, pages 105–112.

Juncai Li, Zhichao Yan, Xuefeng Su, Boxiang Ma, Peiyuan Yang1, and Ru Li. 2023a. Overview of CCL23-eval task 1:Chinese FrameNet semantic parsing. In *Proceedings of the 22nd Chinese National Conference on Computational Linguistics (Volume 3: Evaluations)*, pages 113–123.

Zuoheng Li, Xuanzhi Guo, Dengjian Qiao, and Fan Wu. 2023b. System report for CCL23-eval task 3: Application of entity classification model based on rotary position embedding in chiness frame semantic parsing. In *Proceedings of the 22nd Chinese National Conference on Computational Linguistics (Volume 3: Evaluations)*, pages 94–104.

Ru Li, Yunxiao Zhao, Zhiqiang Wang, Xuefeng Su, Shaoru Guo, Yong Guan, Xiaoqi Han, and Hongyan Zhao. 2024. A comprehensive overview of cfn from a commonsense perspective. *Machine Intelligence Research*, 21:239–256.

Yahui Liu, Zhenghua Li, and Min Zhang. 2023. System report for CCL23-eval task3: SUDA CFSP system. In *Proceedings of the 22nd Chinese National Conference on Computational Linguistics (Volume 3: Evaluations)*, pages 84–93.

Takeru Miyato, Andrew M Dai, and Ian Goodfellow. 2016. Adversarial training methods for semi-supervised text classification. *arXiv preprint arXiv:1605.07725*.

Xuefeng Su, Ru Li, Xiaoli Li, Jeff Z Pan, Hu Zhang, Qinghua Chai, and Xiaoqi Han. 2021. A knowledge-guided framework for frame identification. In *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.

Xiaofei Sun, Linfeng Dong, Xiaoya Li, Zhen Wan, Shuhe Wang, Tianwei Zhang, Jiwei Li, Fei Cheng, Lingjuan Lyu, Fei Wu, et al. 2023. Pushing the limits of ChatGPT on NLP tasks. *arXiv preprint arXiv:2306.09719*.

Gemini Team, Rohan Anil, Sebastian Borgeaud, Yonghui Wu, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M Dai, Anja Hauth, et al. 2023. Gemini: A family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*.

Xiaohui Wang, Ru Li, Zhiqiang Wang, Qinghua Chai, and Xiaoqi Han. 2020. Syntax-aware Chinese frame semantic role labeling based on self-attention. In *Proceedings of the 19th Chinese National Conference on Computational Linguistics*, pages 616–623.

Alexander Willich. 2022. Introducing construction semantics (cxs): A frame-semantic extension of construction grammar and constructicography. *Linguistics Vanguard*, 8:139–149.

Liping You and Kaiying Liu. 2005. Building Chinese Framenet database. In *2005 international conference on natural language processing and knowledge engineering*, pages 301–306.

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

Houquan Zhou, Yang Hou, Zhenghua Li, Xuebin Wang, Zhefeng Wang, Xinyu Duan, and Min Zhang. 2023. How well do large language models understand syntax? An evaluation by asking natural language questions. *arXiv preprint arXiv:2311.08287*.