Revisiting Classical Chinese Event Extraction with Ancient Literature Information

Anonymous ACL submission

Abstract

The research on classical Chinese event extraction trends to directly graft the complex modeling from English or modern Chinese works, neglecting the utilization of the unique characteristic of this language. We argue that, compared with grafting the sophisticated methods from other languages, focusing on classical Chinese's inimitable source of **Ancient Literature** could provide us with extra and comprehensive semantics in event extraction. Motivated by this, we propose a Literary Vision-Language Model (VLM) for classical Chinese event extraction, integrating with literature annotations, historical background and character glyph to capture the inner- and outer-context information from the sequence. Extensive experiments build a new state-of-the-art performance in the GuwenEE, CHED datasets, which underscores the effectiveness of our proposed VLM, and more importantly, these unique features can be obtained precisely at nearly zero cost.

1 Introduction

Classical Chinese, as a written form of the Chinese, had been widely used in recording thousands years of history, culture and science, being a important carrier of digital humanities. Many studies have endeavored to the fundamental tasks such as PoS tagging (Tian and Guo, 2022) and segmentation (Huang et al., 2010), as well as downstream tasks like poetry generation (Zhang and Lapata, 2014) and translation (Wang et al., 2023a) in classical Chinese. However, only a few works focus on event extraction, it aims to extract events from the literature, each of which consists of four types of elements: a *trigger* and multiple *arguments* are the raw spans in the input text, and they will be classified into a corresponding *event type* or *role type*. For instance, an event of war in Figure 1: a *war* event triggered by "引师" (command), the *Person* argument is "高祖" (emperor

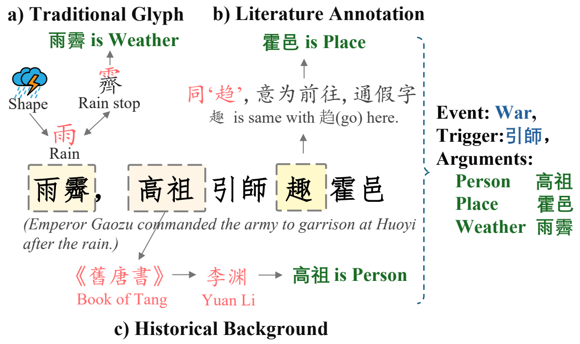

Figure 1: Example of how ancient literature help classical Chinese event extraction.

gaozu), *Weather* argument is "雨霁" (stop of rain), *Place* argument is "霍邑" (huoyi). Event extraction could have a significant contribution for analyzing valuable historical literature, facilitating the preservation and inheritance of historical and cultural heritage.

However, the majority of previous studies tend to directly implement techniques from English (Ji et al., 2021; Tang et al., 2024; Lin et al., 2020; Liu et al., 2021) or modern Chinese (Congcong et al., 2023; Wang et al., 2023b) on classical Chinese datasets. Despite their effectiveness, previous works' modelings are not unique to classical Chinese: either phonetic symbols (Huang et al., 2010), syntax structures (Wang et al., 2023b) or punctuation (Wang and Li, 2024) are also inclusive to English or modern Chinese, making their systems sub-optimal for classical Chinese.

In this study, we shift our attention to the long-existing yet often neglected characteristic of classical Chinese data: They are coming from **Ancient Literature**. Behind which enriched resources can be utilized to enhance extraction of events such as historical background, annotation (注解) and the glyph they are written with. We first illustrate the value of annotation from literature in Figure 1 b), where the "趣"(interesting) could be misunderstood but easily corrected into “趋” (go) with the precise annotation and the following “霍邑” obviously should be a *Place*. The contribution of literature background is shown with the case “高祖”(gaozu) in Figure 1 c), which is not a common word in modern Chinese but could be interrupted as “唐高祖李渊”(Yuan Li) if we know the sample comes from “旧唐书” (book of Tang) and “高祖” should be a *Person*. The final is the glyph of Chinese they are written in the literature. To illustrate, the shape of the “雨” (rain) in “雨霁”(stop of rain) directly evolved from the silhouette of shape of rain and exist as a shared radical in “霁” as shown in Figure 1 a). This intuitive glyphic representation facilitates the straightforward extraction and classification of “雨霁” as a *Weather*.

However, it is challenging to incorporate ancient literature into event extraction. The difficulties are threefold: The first is, which aspects of information should we grasp from the ancient literature as it is unclear how literature information impacts the extraction. Secondly, how can we precisely obtain the literature information as the samples could be distributed from a hundred books? Finally, even with current powerful large language models (LLMs), modeling the obtained information is complicate due to their varied forms, which can be continuous paragraphs (background), discrete dictionaries (annotation), or even visual images (glyph), not only across text forms, but even across modalities.

In this study, we tackle the above challenges with proposed Literary Vision-Language Model for classical Chinese event extraction. We first show the three aspects of background, annotation and glyph from ancient literature are the unique features that could greatly contribute to the event extraction in Section 3, where we further introduce the innovative methodologies for precisely gasping the three aspects from ancient literature in the corresponding subsections. We finally modeling the literature information with fusion instruction and visual bridging in the proposed Literary Vision-Language Model (VLM) as shown in Figure 5, which can cover both the continuous and discrete texts along with the visual glyphic image, deciphering the interplay across different aspects.

The detailed evaluation shows that our proposed model significantly advances the SOTA performance on several benchmarks, validating the value of the literature information in downstream classical Chinese tasks. To the best of our knowledge, we are the first to use the unique source of ancient literature in classical Chinese works.

2 Related Works

2.1 Event Extraction

Event extraction works have indeed leveraged modeling methods from diverse perspectives, from the sequence tagging (Chen et al., 2015; Wang et al., 2019b; Sha et al., 2016; Lin et al., 2020; Fan and He, 2023), structure and graph (Cui et al., 2020a; Yang et al., 2023; Lin et al., 2020; Liu et al., 2023) to multi-modalities(Li et al., 2023b,a; Nguyen et al., 2023; Majumder et al., 2020; Zhang et al., 2024a). Recent trends have shifted towards harnessing the power of LLMs to generate the target sequence (Lu et al., 2021; Liu et al., 2023; Yang et al., 2023; Zhang et al., 2024b).

Besides, the works on Chinese event extraction tend to tailor their methods specifically to the unique characteristics of the Chinese language (Chen and Ji, 2009; Li and Zhou, 2012; Li et al., 2012; Ding et al., 2019), For instance, Xu et al. (2020) addressed the issue of overlapping roles, while Shen et al. (2020) introduced hierarchical event features. Separately, Lin et al. (2018) approached with a hybrid representation for each character.

2.2 Classical Chinese Event Extraction

The majority of research on classical Chinese invariably implement English techniques on classical Chinese datasets directly (Wang et al., 2023b; Ji et al., 2021; Congcong et al., 2023; Tang et al., 2024; Wang et al., 2023c). Among them, Congcong et al. (2023) implements the sequence tagging with pre-trained Guwen-BERT while Tang et al. (2024) showcase the weak performance of several LLMs on classical Chinese task. Some works further design the model based on the features of Chinese (Wang et al., 2023b). For instance, Huang et al. (2010) adopt the phonetic symbols, Wang et al. (2023b) parse the text for syntax structures and Wang and Li (2024) employ the punctuation. However, these modelings could also be applied on modern Chinese, are not exclusive and optimal to classical Chinese.

Different from previous studies, our method stands out as the first to argue for the proprietary modeling for classical Chinese tasks. We introduce an innovative approach that manipulates

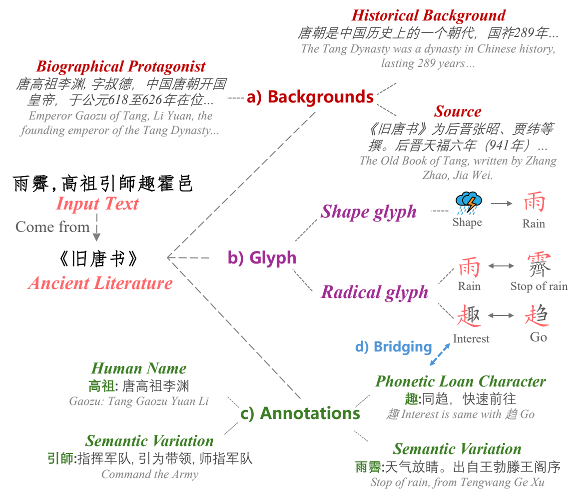

Figure 2: Illustration of ancient literature information.

ancient literature, the unique origin of classical Chinese, with the literary vision-language models specifically accommodating it.

3 Value and Extraction of Ancient Literature Information

In this secton, We showcase how the ancient literature can help improve the task from the following three aspects along with the extraction of them.

Backgrounds

Backgrounds is a feature that provides essential macro background knowledge for a specific sample as shown in Figure 2. Specifically, the backgrounds information contains 3 types:

- **Literature source** has been used for providing the most direct and structured knowledge for a sample (Kasanishi et al., 2023; Naik et al., 2022). We however is the first to stand for it's works in classical Chinese works, to illustrate：

  "遂听信计，部署诸将所击"
  (Took Xin's advice, attacked with the generals)

  where "信" could be easily interpreted into "采纳/听从" (take) with a wrong segmentation of "听信" (take), but could be corrected into "韩信" (Xin Han, exists as the *Person* argument) if we know the sample comes from "史记 · 淮阴侯列传" and the protagonist of it is "韩信".

- **Historical background** can paint a somehow much larger image for us and give us extra information in a more indirect way such as the historical background "唐朝" (Tang Dynasty) of the given example in Figure 2. By which we can idenify the "高祖" (Gaozu) as "唐高祖李渊" (Tang Gaozu Yuan Li) otherwise it could be misunderstood as "汉高祖刘邦" (Han Gaozu Bang Liu) or "魏高祖曹丕" (Wei Gaozu Pi Cao).

- **Biographical protagonist** contains the identity of the protagonist, has been proofed (Ruppenhofer et al., 2020; Palmero Aprosio and Tonelli, 2015) that can help us effectively differentiate between people with common appellations as most of the ancient literature are wrote in biographical formation (纪传体). For instance:

  "而上又上泰山，有秘祠其颠"
  (The emperor climbed the mountain Tai again, held a secret rituals at the peak.)

  the first "上" (emperor, exists as *Person*) could be garbled with the second "上" (climb). However, if been awared of the protagonist of the sample is "汉武帝" (emperor wu of han), it will be easier for us to distinguish the two "上".

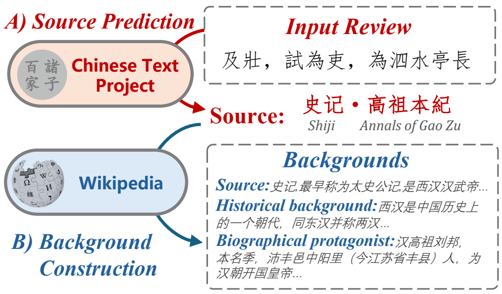

Figure 3: Process of backgrounds construction.

As the value of background has been illustrated, we subsequently proceed to obtain it. Background intrinsically is a cluster of information that form as a tree rooted at the specific literature source, with various aspects derived from it. We thus grasp the background information with proposed pipeline starting at extracting the source of the sample as illustrated in the Figure 3. The unprocessed sample is initially searched in *Chinese Text Project* (中国哲学书电子化计划)[1] to identify the specific ancient literary source accurately. Once this source is identified, it is used to gather additional relevant background details from Wikipedia. These details are then converted into actionable instructions in the following section.

Annotation

Annotation (注解) refers to the contemporary Chinese interpretations crafted by expert historians

[1] https://ctext.org

for ancient literary works. Leveraging annotations can enable us to connect classical Chinese task with modern Chinese. There are 3 types included:

- **Semantic variation** (古今异义) is an significant challenge for downstream tasks (Perrone et al., 2019). We thus annotates the words whose semantic meaning has been changed when compared with the time of writing the literature. For instance:

  "予除右丞相兼枢密使"
  (I was appointed as the Right Chancellor and Minister of the Imperial Secretariat.)

  the meaning of the word "除" (appoint, exists as the event trigger) is different from its meaning of *remove* in modern Chinese.

- **Phonetic loan character** (通假字) represents characters which have been formed from "borrowed" characters of homophonous (or near-homophonous) morphemes (Jiang and Ren, 2023) such as the "趣" and "趋" in Figure 2.

- **Human name** is more intuitive when compared with the previous two, explaining the specific part of the sample as the human name to prevent misunderstanding since the sample is picked and cut out of context. For instance:

  "尚悲感，发病恍惚"
  (Shang felt sorrowful, with dizziness.)

  "尚" could be wrongly interpreted as *still* but it actually refers to "夏侯尚" (Shang Xiahou) as a *Person* argument.

For extraction, the annotations are more discrete when compared with the backgrounds, their semantic meanings are paired up with the corresponding words. We thus extract the glyph follow the manner of dictionary. Initially, the input sentence undergoes segmentation with the pre-trained classical Chinese tokenizer Jiayan[2]. Subsequently, the tokenized tokens are retrieved from the Ancient Chinese Dictionary (古代汉语词典) to procure their modern Chinese interpretations.

Glyph

As a hieroglyphic language, glyph of characters that the literature are written in could contains enriched semantic information, which could be illustrated from two aspects. The first is the shape

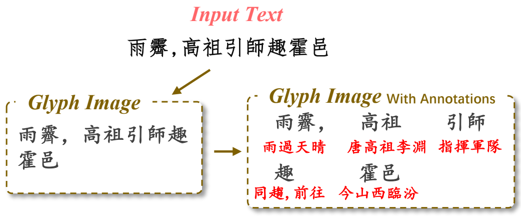

Figure 4: Construction of glyphic image.

glyph, such as the shape of the character "山","雨" and "门" in Figure 2, directly evolved from the actual silhouette of the mountain, rain and door respectively. This intuitive glyphic representation facilitates the straightforward semantic understanding of them. Furthermore, there are also exist connections in radical glyph plays a critical role in communicating the semantic essence of characters such as the shared radical of "雨" (rain) between "雷"(thunder) and "霆" (lightning), the "讠" (speech) between "讲" (speak) and "辩" (argue).

However, obtaining glyph is different from the previous two information, it comes from visual modality. Previous works' splitting into radicals or characters involves the affordable retraining of LLMs. We thus adopt the sentence-level glyph image (Bao et al., 2024) to capture the glyph information. Specifically, given a sentence, each character is transformed to an image of size $p \times q$ with a specific font Song (宋体) as shown in Figure 4. Then, a sentence containing N characters is constructed in a image where the characters are concatenated sequentially follow the writing order of from left to right and starting a new line below.

Considering there are also exist connections in glyphic level for annotations such as the shared radical "辶" in "趣" and "趋" in Figure 2 d), we further build connections between classical input and modern annotation in the glyphic image. Specifically，we follow the layout of explanatory book (注解本), placing each annotation underneath its corresponding word as an active visual bridging between classical and modern Chinese.

4 Classical Chinese Event Extraction via Literary Vision-Language Model

Even if the information is obtained, how to modeling them is another challenge due to their various forms. In this study, we propose Literary Vision-Language Model as shown in Figure 5. Specifically, we propose a fusion instruction to combine all the above information, which contains mixed

[2] https://github.com/jiaeyan/Jiayan

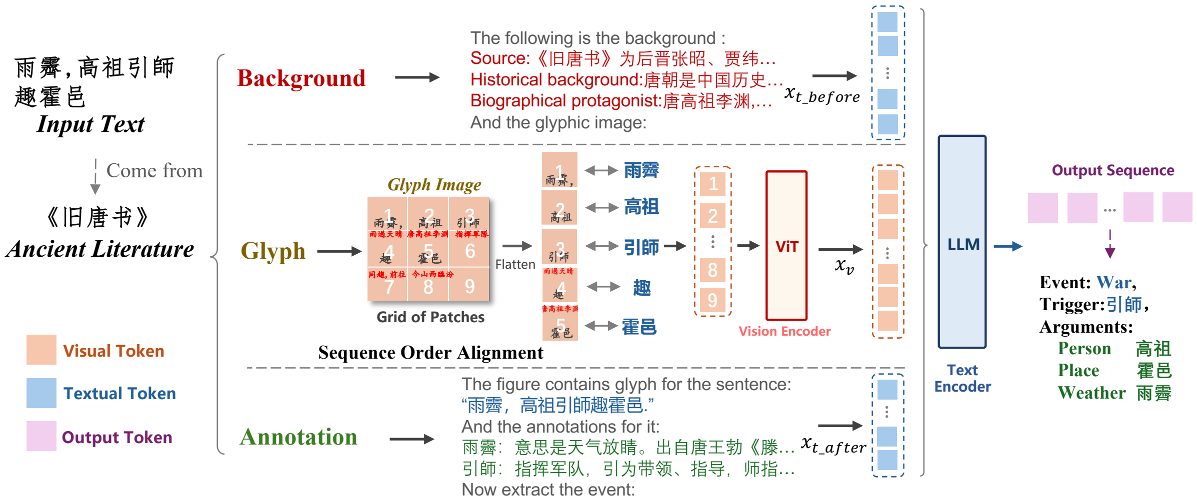

Figure 5: The illustration of our Literary Vision-Language Model.

visual and textual contents. We will first introduce the modeling of the visual glyph and then cover the rest with fusion instruction.

4.1 Modeling of Character Glyph

Vision Transformer (ViT) is employed as our image encoder for extraction glyphic information, which is excels at distilling high-level visual features from raw images. We adopt the Sequence Order Alignment (Bao et al., 2024) method to align with the textual writing order of raw sentence.

Specifically, the input image is divided into patches, and embedded into visual tokens with position encoding. The grids are then flatten into a sequence that align with the textual tokens in the inputted review as shown in Figure 5. The patch 1 will be placed at the start of the flattened sequence, followed by the adjacent patch. Once reaching the end of a line, it will continue in a zigzag pattern, moving to the rightmost patch in the line below. With the alignment method, the visual tokens are in the same order with the textual tokens, which then augmented with positional encodings to obtain the final image representations x_v.

4.2 Fusion Instruction

We specifically design the instructions for modeling the rest background and annotation, which also responsible for guiding the model to fuse the textual contents with visual representations x_v obtained. The fusion instruction are designed as shown in Figure 5, which include a guiding instruction x_{t_before} and x_{t_after} at both before and after the visual tokens, the instructions x_{t_before} before it covers the extracted background information while the x_{t_after} after covers the annotations along with the specific sentence for extracting.

We subsequently fuse the obtained the vision encoder's output as visual tokens x_v and the tokenized text as textual tokens x_{t_before} and x_{t_after} for LLM's processing. These tokens are merged to create the input x:

$$x = [x_{t\_before},\ x_v,\ x_{t\_after}] \tag{1}$$

Given the fused sequence $x = x_1, ..., x_{|x|}$ as input, The decoder predicts token-by-token. At the i-th step of generation, the decoder predicts the i-th token y_i and decoder state h_i^d as:

$$y_i, h_i^d = ([h_1^d, ..., h_{i-1}^d], y_{i-1}) \tag{2}$$

The conditional probability of the whole output sequence $p(y|x)$ is progressively combined by the probability of each step $p(y_i|y_{<i}, x)$:

$$p(y|x) = \prod_{i=1}^{|y|} p(y_i|y_{<i}, x) \tag{3}$$

where $y_{<i} = y_1...y_{i-1}$, and $p(y_i|y_{<i}, x)$ are the probabilities over target vocabulary V.

The objective functions is to maximize the output target sequence X_T probability given the review sentence X_O. Therefore, we optimize the negative log-likelihood loss function:

$$\mathcal{L} = \frac{-1}{|\tau|} \sum_{(X_O, X_T) \in \tau} \log p(X_T|X_O; \theta) \tag{4}$$

where θ is the model parameters, and (X_O, X_T) is a (*sentence, target*) pair in training set τ, then

$$\log p(X_T|X_O; \theta) = \sum_{i=1}^{n} \log p(x_T^i|x_T^1, x_T^2, ...x_T^{i-1}, X_O; \theta) \tag{5}$$

where $p(x_T^i|x_T^1, x_T^2, ...x_T^{i-1}, X_O; \theta)$ is calculated by decoder.

Method	GuwenEE						CHED					
	Tri-I			Tri-C			Tri-I			Tri-C		
	P.	R.	F1.	P.	R.	F1.	P.	R.	F1.	P.	R.	F1.
Chinese-Bert-CRF	0.427	0.471	0.447	0.304	0.342	0.321	0.736	0.699	0.717	0.701	0.665	0.682
Guwen-Bert-CRF	0.476	0.507	0.490	0.368	0.399	0.382	0.752	0.713	0.731	0.733	0.709	0.719
DMBert	0.498	0.529	0.512	0.391	0.415	0.402	0.771	0.724	0.746	0.740	0.715	0.727
Mengzi-T5-base	0.551	0.512	0.531	0.432	0.403	0.418	0.783	0.735	0.758	0.751	0.726	0.738
LLaMA-3-8B	0.526	0.520	0.523	0.411	0.409	0.410	0.769	0.707	0.736	0.742	0.693	0.716
LLaMA-3-Chinese-8B	0.537	0.534	0.535	0.433	0.421	0.426	0.773	0.721	0.745	0.751	0.699	0.724
ChatGLM-3-6B	0.502	0.505	0.503	0.391	0.301	0.341	0.739	0.677	0.709	0.685	0.636	0.659
InternLM-2-7B	0.524	0.533	0.527	0.417	0.406	0.411	0.767	0.731	0.748	0.715	0.734	0.724
Qwen-2-7B	0.531	0.523	0.526	0.428	0.422	0.425	0.781	0.753	0.766	0.763	0.729	0.746
NPN	0.539	0.533	0.534	0.409	0.473	0.438	0.843	0.731	0.782	0.776	0.676	0.721
TLNN	0.551	0.536	0.543	0.436	0.447	0.441	0.822	0.763	0.791	0.772	0.701	0.734
ONEIE	**0.572**	0.531	0.550	0.457	0.441	0.448	0.826	0.792	0.808	0.761	0.739	0.749
Ours	0.568	**0.580**	**0.574**	**0.461**	**0.470**	**0.466**	**0.811**	**0.831**	**0.822**	**0.751**	**0.770**	**0.761**

Table 1: Comparison with baselines in Event Detection.

5 Experiment

5.1 Dataset and Experiment Setting

In this study, we use GuwenEE[3] for Event Extraction and CHED2023 (Congcong et al., 2023) for Event Detection. We employ the pre-trained weight InternLM-XComposer2-VL (Dong et al., 2024) for our Vision-Language Model and LoRA fine-tune the LLM parameters. The image size is set to 490 × 490. The glyph is interpreted with traditional Chinese and Song (宋体). Our experiments are carried out with Nvidia RTX A6000.

We use the same criteria as (Zhang et al., 2019; Wadden et al., 2019) for evaluation. A **Trigger** is correctly identified (Tri-I) if its offsets match a ground truth trigger. It is correctly classified (Tri-C) if its event type also matches. An **Argument** is correctly identified (Arg-I) if its offsets and event type match a ground truth argument. It is correctly classified (Arg-C) if its role label also matches.

5.2 Main Results

Our baselines could be divided into three parts, including common methods with Chinese language models, such as sequence tagging: Chinese-Bert-CRF (Cui et al., 2020b), Guwen-Bert-CRF (Datahammer, 2020), DMBert (Wang et al., 2019a), and generative models: Mengzi-T5 (Zhang et al., 2021), LLaMA-3-8B (AI@Meta, 2024) LLaMA-3-Chinese-8B (Cui et al., 2023) ChatGLM-3-6B (Zeng et al., 2023) and Qwen-2-7B (qwe, 2024). We also have feature-enriched models, include NPN (Lin et al., 2018), TLNN (Ding et al., 2019), ONEIE (Lin et al., 2020).

[3]https://github.com/Lyn4ever29/GuwenEE

Method	Arg-I			Arg-C		
	P.	R.	F1.	P.	R.	F1.
Mengzi-T5-base	0.377	0.368	0.372	0.354	0.361	0.357
LLaMA-3-8B	0.369	0.363	0.365	0.337	0.342	0.339
ChatGLM-3-6B	0.356	0.355	0.355	0.323	0.321	0.322
InternLM-2-7B	0.362	0.359	0.360	0.341	0.336	0.338
Qwen-2-7B	0.394	0.388	0.391	0.363	0.371	0.366
ONEIE	0.383	0.396	0.389	0.367	0.374	0.370
Ours	**0.398**	**0.413**	**0.406**	**0.376**	**0.388**	**0.382**

Table 2: Results in Argument Extraction in GuwenEE.

As shown in Table 1 and Table 2, we first find that, among the common methods, the methods with language models that specifically pretrained on Chinese or classical Chinese can outperform the others(Guwen-Bert versus Chinese-Bert, LLaMA-3-Chinese versus LLaMA-3), revealing the advantage of having a language model pretrained on target language in event extraction. In addition, the methods integrate specific designs for Chinese surpass the common methods above, showing us the value of language adaptation and localization in designing model.

Moreover, our proposed model exhibits significant improvements over all prior studies ($p < 0.05$), demonstrating the efficacy of literature information when applied with large language models for classical Chinese event extraction. To the best of our knowledge, this is the first attempt to leverage literature information in event extraction.

5.3 Contribution of Literature Information

After analyzing the overall performance, a natural question arises: *How much does the literature information contribute to it*? To investigate this, we gradually incorporate various literature information into VLM, starting from the backgrounds up to the visual glyphs. We use "Basic" in Table 3 to refer to the removing of all of the literature information, relying solely on the raw sentence.

As depicted in Table 3, when using only raw sentences, the performance is notably low, which is excepted since the VLM is not pre-trained on classical Chinese. Significantly improved performance is observed when the background information is included, we attribute this as it provides the macro background knowledge of the sample. Furthermore, the annotation also contributes positively to event extraction, inside which the annotation provides a precise way of communicating the classical and modern Chinese, the semantic variation also stands out as it can detailed explain the changes of semantics. Besides, the glyphic image and visual bridging are designed to actively guide the model to integrate glyphic information, proving that grasping the glyphic feature though visual modality by sentence-level image for capturing more shared glyphes is practical.

Additionally, our proposed model, which combines all the of the above features, achieves the best performance and showcases the value of literature information in event extraction.

6 Analysis and Discussion

6.1 LLMs' Knowledge in Ancient Literature

We subsequently offer another analysis for our results from the perspective of analyzing LLM's knowledge in ancient literature. Specifically, we adopt the perplexity metrics (Wan et al., 2024) to evaluate different models' knowledge on proposed aspects of ancient literature. The perplexities will be calculated over the spans of background, annotation and raw sample in GuwenEE's test set, lower values represent more knowledge of the model in this aspect. A perplexity of the rest span will also be calculated to show the level of the model's knowledge in non-classical-Chinese language for comparison, marked as "Plain Text".

As shown in Figure 6, current mainstream LLMs have a significant higher perplexity on the classical Chinese aspects when compared with plain texts, validating their lack of corresponding knowledge and our motivation of making up for it. On the other hand, after fine-tuning, our model keeps the values of literature aspects closing to the plain text on the test set, showing the ancient literature knowledge have been effectively injected and facilitating the successful extraction.

Method	GuwenEE		CHED
	Tri-C	Arg-C	Tri-C
Basic	0.411	0.338	0.724
Background			
+Literature source	0.427	0.348	0.731
+Historical background	0.434	0.356	0.736
+Biographical protagonist	0.423	0.340	0.729
+*All*	**0.438**	**0.368**	**0.746**
Annotation			
+Semantic variation	0.430	0.353	0.735
+Phonetic loan character	0.426	0.351	0.728
+Human Name	0.424	0.339	0.733
+*All*	**0.433**	**0.359**	**0.739**
Glyph			
+Sentence image	0.422	0.358	0.741
+Sentence image, Visual bridging	**0.439**	**0.366**	**0.744**
Ours	**0.466**	**0.382**	**0.761**

Table 3: The contribution of the literature information.

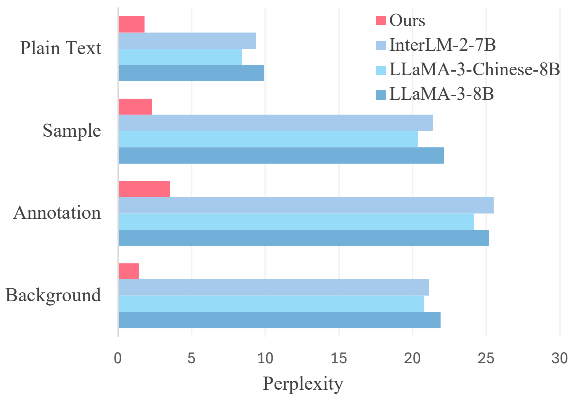

Figure 6: LLMs' perplexities on proposed aspects.

6.2 Impact of Literature Information Source

As we pay efforts to obtain the precise literature information from Wikipedia and 古代汉语词典, a intuitive question is *Can the powerful large language models replace them?* We thus investigate this question by replacing all our literature information sources with LLMs. Specifically, for all the information shown in Figure 2, we design prompts (in Appendix A) to fetch the information from LLM at one stop, and solely rely on the output information to proceed the entire workflow.

As shown in Table 5, the human-made literature sources have shown their value again: our way of obtain the precise literature information beats all of the LLM sources, giving us a conclusion that current level of LLMs still can not replace the precise human-made literature as they are suffering

Input	Subtask	w/o Literature	w Literature	Remark
丞相守道曰：汉文字恐初亦未必能如此	Argument Identification	丞相 ✗ Prime minister	丞相守道✓ Prime minister Shoudao	完颜守道 (Shoudao Wanyan) is a person's in Jing Dynasty.
天子诏帝朝会乘舆升殿	Trigger Identification	升殿 ✗ Went to palace	诏✓ Convene	The radical "讠"(speech) in "诏"(edict) indicating a action.
已而信约，与赵、魏合从畔秦	Trigger Identification	合从 ✗ Join	畔✓ Betray	"畔"(bank) is the same as "叛"(betray).
帝次山阳，闻败，卷甲与数十人造江上征问	Trigger Classification	次: 人生-死亡 ✗ Life-Death	次: 军事-备战-驻扎✓ Military-War-Garrison	"次"(times) could represent "驻扎"(garrison).

Table 4: Case study

Method		GuwenEE		CHED
		Tri-C	Arg-C	Tri-C
Basic		0.411	0.338	0.724
Source	InternLM-2-7B	0.423	0.353	0.734
	Qwen-2-7B	0.431	0.361	0.748
	LLaMA-3-8B	0.436	0.359	0.744
	Literature (Ours)	**0.466**	**0.382**	**0.761**

Table 5: Comparison with various information sources.

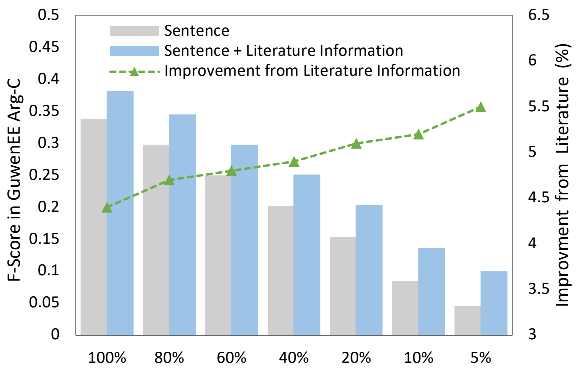

Figure 7: Improvement of data efficiency.

from the hallucination and the lack of pre-trained knowledge in various specific minor domains, the reliable human-made literature still can be a valuable ally for us in the era of large language models.

6.3 Analysis of Data Efficiency

Compared with the raw sentence, one of the advantages of literature information is that there are large amount of shared knowledge such as the historic background (e.g., two samples come from the same dynasty), making it easier to build semantic connection across samples with a small size of training data. We thus investigate how the literature improves the data efficiency of our model by comparing with using raw sentence under limited training data in Figure 7.

From the figure, we find that the more training data, the higher performance our proposed model can reach. Moreover, the advantage of the performance brought by the literature information increases under limited data size, showing the superiority of literature information in low resource situation where a pool of shared knowledge can be easily build compared with raw sentences.

7 Cases Study

We launch case studies to make a more intuitive comparison between the VLMs with and without literature information in Table 4.

We demonstrate that historic background can help effectively identify the argument target in the first examples. The "守道" is a person in Jing(金) Dynasty and the VLM without literature information ignores it, resulting in wrong identification.

Furthermore, we illustrate that the glyph can better convey semantics in the second example, where the "讠"(speech) in "诏"(edict) indicating "诏" as an action and the VLM can easily identify the right target with it.

The third and fourth example showcase the value of annotation: the "畔" was interpreted as a bank of river while the "次" was interpreted as a death event without modern Chinese annotation while ours can correctly comprehend them as "叛乱"(betray) and "驻扎"(garrison) respectively.

8 Conclusion

In this study, we explore how to improve the classical Chinese event extraction from the inimitable characteristics of it. By leveraging the long-existing yet often overlooked origin of literature of classical Chinese, our proposed LVLM achieves SOTA performance in several benchmarks without the need for complex and costly annotation.

Furthermore, our results validate that for the works focusing on the minor languages, solely relaying on the LLMs' knowledge whose major training corpus is in English may not be an optimal option. Instead, the reliable human-made features still can be a precise and valuable ally for us.

Limitations

The limitations of our work can be stated from two perspectives. Firstly, besides the literature information, there is another feature whose effect on downstream tasks is not yet known: pronunciation. In future research, further exploration of the impact of pronunciation, especially the variation of pronunciation between ancient and modern Chinese could provide valuable insights.

Secondly, our focus has been primarily on a single language. While we have achieved promising results in this language, it is important to acknowledge that the generalizability of our approach is limited since other languages may not have the source of ancient literature.

References

2024. Qwen2 technical report.

AI@Meta. 2024. Llama 3 model card.

Xiaoyi Bao, Jinghang Gu, Zhongqing Wang, Minjie Qiang, and Chu-Ren Huang. 2024. Employing glyphic information for Chinese event extraction with vision-language model. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 1068–1080, Miami, Florida, USA. Association for Computational Linguistics.

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, Beijing, China. Association for Computational Linguistics.

Zheng Chen and Heng Ji. 2009. Language specific issue and feature exploration in Chinese event extraction. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Short Papers*, pages 209–212, Boulder, Colorado. Association for Computational Linguistics.

Wei Congcong, Feng Zhenbing, Huang Shutan, Li Wei, and Shao Yanqiu. 2023. CHED: A cross-historical dataset with a logical event schema for classical Chinese event detection. In *Proceedings of the 22nd Chinese National Conference on Computational Linguistics*, pages 875–888, Harbin, China. Chinese Information Processing Society of China.

Shiyao Cui, Bowen Yu, Tingwen Liu, Zhenyu Zhang, Xuebin Wang, and Jinqiao Shi. 2020a. Edge-enhanced graph convolution networks for event detection with syntactic relation. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2329–2339, Online. Association for Computational Linguistics.

Yiming Cui, Wanxiang Che, Ting Liu, Bing Qin, Shijin Wang, and Guoping Hu. 2020b. Revisiting pre-trained models for Chinese natural language processing. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 657–668, Online. Association for Computational Linguistics.

Yiming Cui, Ziqing Yang, and Xin Yao. 2023. Efficient and effective text encoding for chinese llama and alpaca. *arXiv preprint arXiv:2304.08177*.

Datahammer. 2020. Guwen-bert-base.

Ning Ding, Ziran Li, Zhiyuan Liu, Haitao Zheng, and Zibo Lin. 2019. Event detection with trigger-aware lattice neural network. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 347–356, Hong Kong, China. Association for Computational Linguistics.

Xiaoyi Dong, Pan Zhang, Yuhang Zang, Yuhang Cao, Bin Wang, Linke Ouyang, Xilin Wei, Songyang Zhang, Haodong Duan, Maosong Cao, Wenwei Zhang, Yining Li, Hang Yan, Yang Gao, Xinyue Zhang, Wei Li, Jingwen Li, Kai Chen, Conghui He, Xingcheng Zhang, Yu Qiao, Dahua Lin, and Jiaqi Wang. 2024. Internlm-xcomposer2: Mastering free-form text-image composition and comprehension in vision-language large model.

Zhiyuan Fan and Shizhu He. 2023. Efficient data learning for open information extraction with pre-trained language models. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 13056–13063, Singapore. Association for Computational Linguistics.

Hen-Hsen Huang, Chuen-Tsai Sun, and Hsin-Hsi Chen. 2010. Classical Chinese sentence segmentation. In *CIPS-SIGHAN Joint Conference on Chinese Language Processing*.

Zijing Ji, Yuxin Shen, Yining Sun, Tian Yu, and Xin Wang. 2021. C-clue: A benchmark of classical chinese based on a crowdsourcing system for knowledge graph construction. In *Knowledge Graph and Semantic Computing: Knowledge Graph Empowers New Infrastructure Construction*, pages 295–301, Singapore. Springer Singapore.

Lihong Jiang and Mingshan Ren. 2023. *A Brief Introduction to Form and Meaning of Chinese Characters & Dictionaries*, pages 9–48. Springer Nature Singapore, Singapore.

Tetsu Kasanishi, Masaru Isonuma, Junichiro Mori, and Ichiro Sakata. 2023. SciReviewGen: A large-scale dataset for automatic literature review generation. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 6695–6715, Toronto, Canada. Association for Computational Linguistics.

Jiaqi Li, Chuanyi Zhang, Miaozeng Du, Dehai Min, Yongrui Chen, and Guilin Qi. 2023a. Three stream based multi-level event contrastive learning for text-video event extraction. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 1666–1676, Singapore. Association for Computational Linguistics.

Peifeng Li and Guodong Zhou. 2012. Employing morphological structures and sememes for Chinese event extraction. In *Proceedings of COLING 2012*, pages 1619–1634, Mumbai, India. The COLING 2012 Organizing Committee.

Peifeng Li, Guodong Zhou, Qiaoming Zhu, and Libin Hou. 2012. Employing compositional semantics and discourse consistency in Chinese event extraction. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 1006–1016, Jeju Island, Korea. Association for Computational Linguistics.

Ruiqi Li, Patrik Haslum, and Leyang Cui. 2023b. EDeR: Towards understanding dependency relations between events. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 14969–14983, Singapore. Association for Computational Linguistics.

Hongyu Lin, Yaojie Lu, Xianpei Han, and Le Sun. 2018. Nugget proposal networks for Chinese event detection. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1565–1574, Melbourne, Australia. Association for Computational Linguistics.

Ying Lin, Heng Ji, Fei Huang, and Lingfei Wu. 2020. A joint neural model for information extraction with global features. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7999–8009, Online. Association for Computational Linguistics.

Jian Liu, Dianbo Sui, Kang Liu, Haoyan Liu, and Zhe Zhao. 2023. Learning with partial annotations for event detection. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 508–523, Toronto, Canada. Association for Computational Linguistics.

Jiangwei Liu, Jingshu Zhang, Xiaohong Huang, and Liangyu Min. 2021. Syntactic-gcn bert based chinese event extraction.

Yaojie Lu, Hongyu Lin, Jin Xu, Xianpei Han, Jialong Tang, Annan Li, Le Sun, Meng Liao, and Shaoyi Chen. 2021. Text2Event: Controllable sequence-to-structure generation for end-to-end event extraction. 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 2795–2806, Online. Association for Computational Linguistics.

Bodhisattwa Prasad Majumder, Navneet Potti, Sandeep Tata, James Bradley Wendt, Qi Zhao, and Marc Najork. 2020. Representation learning for information extraction from form-like documents. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6495–6504, Online. Association for Computational Linguistics.

Aakanksha Naik, Sravanthi Parasa, Sergey Feldman, Lucy Lu Wang, and Tom Hope. 2022. Literature-augmented clinical outcome prediction. In *Findings of the Association for Computational Linguistics: NAACL 2022*, pages 438–453, Seattle, United States. Association for Computational Linguistics.

Khanh Duy Nguyen, Zixuan Zhang, Reece Suchocki, Sha Li, Martha Palmer, Susan Windisch Brown, Jiawei Han, and Heng Ji. 2023. RESIN-EDITOR: A schema-guided hierarchical event graph visualizer and editor. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 365–372, Singapore. Association for Computational Linguistics.

Alessio Palmero Aprosio and Sara Tonelli. 2015. Recognizing biographical sections in Wikipedia. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 811–816, Lisbon, Portugal. Association for Computational Linguistics.

Valerio Perrone, Marco Palma, Simon Hengchen, Alessandro Vatri, Jim Q. Smith, and Barbara McGillivray. 2019. GASC: Genre-aware semantic change for Ancient Greek. In *Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change*, pages 56–66, Florence, Italy. Association for Computational Linguistics.

Josef Ruppenhofer, Ines Rehbein, and Carolina Flinz. 2020. Fine-grained named entity annotations for German biographic interviews. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 4605–4614, Marseille, France. European Language Resources Association.

Lei Sha, Jing Liu, Chin-Yew Lin, Sujian Li, Baobao Chang, and Zhifang Sui. 2016. RBPB: Regularization-based pattern balancing method for event extraction. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1224–1234, Berlin, Germany. Association for Computational Linguistics.

Shirong Shen, Guilin Qi, Zhen Li, Sheng Bi, and Lusheng Wang. 2020. Hierarchical Chinese legal event extraction via pedal attention mechanism. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 100–113, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Xuemei Tang, Qi Su, Jun Wang, and Zekun Deng. 2024. CHisIEC: An information extraction corpus for Ancient Chinese history. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 3192–3202, Torino, Italia. ELRA and ICCL.

Yanzhi Tian and Yuhang Guo. 2022. Ancient Chinese word segmentation and part-of-speech tagging using data augmentation. In *Proceedings of the Second Workshop on Language Technologies for Historical and Ancient Languages*, pages 146–149, Marseille, France. European Language Resources Association.

David Wadden, Ulme Wennberg, Yi Luan, and Hannaneh Hajishirzi. 2019. Entity, relation, and event extraction with contextualized span representations. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5784–5789, Hong Kong, China. Association for Computational Linguistics.

Fanqi Wan, Xinting Huang, Deng Cai, Xiaojun Quan, Wei Bi, and Shuming Shi. 2024. Knowledge fusion of large language models. In *The Twelfth International Conference on Learning Representations*.

Hao Wang, Hirofumi Shimizu, and Daisuke Kawahara. 2023a. Kanbun-LM: Reading and translating classical Chinese in Japanese methods by language models. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 8589–8601, Toronto, Canada. Association for Computational Linguistics.

Ping Wang, Shitou Zhang, Zuchao Li, and Jingrui Hou. 2023b. Enhancing Ancient Chinese understanding with derived noisy syntax trees. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 4: Student Research Workshop)*, pages 83–92, Toronto, Canada. Association for Computational Linguistics.

Xiaozhi Wang, Xu Han, Zhiyuan Liu, Maosong Sun, and Peng Li. 2019a. Adversarial training for weakly supervised event detection. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 998–1008, Minneapolis, Minnesota. Association for Computational Linguistics.

Xiaozhi Wang, Ziqi Wang, Xu Han, Zhiyuan Liu, Juanzi Li, Peng Li, Maosong Sun, Jie Zhou, and Xiang Ren. 2019b. HMEAE: Hierarchical modular event argument extraction. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5777–5783, Hong Kong, China. Association for Computational Linguistics.

Xuebin Wang and Zhenghua Li. 2024. Two sequence labeling approaches to sentence segmentation and punctuation prediction for classic Chinese texts. In *Proceedings of the Third Workshop on Language Technologies for Historical and Ancient Languages (LT4HALA) @ LREC-COLING-2024*, pages 237–241, Torino, Italia. ELRA and ICCL.

Yuxuan Wang, Jack Wang, Dongyan Zhao, and Zilong Zheng. 2023c. Rethinking dictionaries and glyphs for Chinese language pre-training. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 1089–1101, Toronto, Canada. Association for Computational Linguistics.

Nuo Xu, Haihua Xie, and Dongyan Zhao. 2020. A novel joint framework for multiple Chinese events extraction. In *Proceedings of the 19th Chinese National Conference on Computational Linguistics*, pages 950–961, Haikou, China. Chinese Information Processing Society of China.

Yuqing Yang, Qipeng Guo, Xiangkun Hu, Yue Zhang, Xipeng Qiu, and Zheng Zhang. 2023. An AMR-based link prediction approach for document-level event argument extraction. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 12876–12889, Toronto, Canada. Association for Computational Linguistics.

Aohan Zeng, Xiao Liu, Zhengxiao Du, Zihan Wang, Hanyu Lai, Ming Ding, Zhuoyi Yang, Yifan Xu, Wendi Zheng, Xiao Xia, Weng Lam Tam, Zixuan Ma, Yufei Xue, Jidong Zhai, Wenguang Chen, Peng Zhang, Yuxiao Dong, and Jie Tang. 2023. Glm-130b: An open bilingual pre-trained model.

Meishan Zhang, Hao Fei, Bin Wang, Shengqiong Wu, Yixin Cao, Fei Li, and Min Zhang. 2024a. Recognizing everything from all modalities at once: Grounded multimodal universal information extraction. In *Findings of the Association for Computational Linguistics ACL 2024*, pages 14498–14511, Bangkok, Thailand and virtual meeting. Association for Computational Linguistics.

Tongtao Zhang, Heng Ji, and Avirup Sil. 2019. Joint Entity and Event Extraction with Generative Adversarial Imitation Learning. *Data Intelligence*, 1(2):99–120.

Weiyan Zhang, Wanpeng Lu, Jiacheng Wang, Yating Wang, Lihan Chen, Haiyun Jiang, Jingping Liu, and Tong Ruan. 2024b. Unexpected phenomenon: LLMs' spurious associations in information extraction. In *Findings of the Association for Computational Linguistics ACL 2024*, pages 9176–9190, Bangkok, Thailand and virtual meeting. Association for Computational Linguistics.

Prompt

```
对于给出的文言文文本:
For the given literary text.

<sample>
<sample>

给出其详尽的文献来源<source>，历史背景<historical background>，传记主信息<biographical protagonist>与注释信息<annotation 1> <annotation 2>，并严格按照以下格式:
Give the source <source>, historical background <historical background>, biographical protagonist <biographical protagonist> , and annotations <annotation 1> <annotation 2>, strictly in the following format:

###
文献来源:  <source>
历史背景:  <historical background >
传记主信息: <biographical protagonist>
注释信息1: < annotation 1>
注释信息2: < annotation 2>
…
###
###
Source: <source>
Historical background: <historical background>
biographical protagonist: <biographical protagonist>
Annotation 1: < annotation 1>
Annotation 2: < annotation 2>
...
###

现在直接按照上述格式给出<sample>的信息。
Now give the information for <sample> directly in the above format.
```

Figure 8: Illustration of the prompt.

Xingxing Zhang and Mirella Lapata. 2014. Chinese poetry generation with recurrent neural networks. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 670–680, Doha, Qatar. Association for Computational Linguistics.

Zhuosheng Zhang, Hanqing Zhang, Keming Chen, Yuhang Guo, Jingyun Hua, Yulong Wang, and Ming Zhou. 2021. Mengzi: Towards lightweight yet ingenious pre-trained models for chinese.

A Prompts for Querying LLMs

In Figure 8, we give the prompts for querying with LLMs about the literature information to check if the LLMs can replace the precise information obtaining method we proposed, which is investigated in Section 6.2.