Research on Paper Semantic Novelty Measurement Based on Large Language Model*

Xinpeng Qiu[1,†] and Jing Li[1,*,†]

[1] Sun Yat-sen University, No. 132, Outer Ring East Road, University Town, Guangzhou, Guangdong Province, China

Abstract

This paper proposes a semantic novelty measurement model for scientific papers using a large language model to generate question and method words semi-supervisedly. LoRA and prompt words enhance keyword generation accuracy and structural measurement. The model achieves 66.0% recall, 63.6% precision, and 65.9% sum, improving with more training samples. At 3000 samples, the training set is cost-effective. The proposed method, leveraging fine-tuned large language models, is effective and robust.

Keywords

paper evaluation, semantic novelty, large language model, natural language generation

1. Fine-tuning design of large language model

According to the demand of "extracting paper keywords", this paper adopts the LoRA framework of intrinsic rank adapter to fine-tune and train the large model and the weight matrix of the pre-trained model: $W_0 = R^{d\times k}$, Its update is represented by low-rank decomposition.

$$W_0 + \Delta W = W_0 + BA \tag{1}$$

Where, W_0 represents the weight matrix of the pre-training model, ΔWrepresents the parameter update during fine-tuning, B is a trainable matrix, which is all 0 matrix at initialization. A is also a trainable matrix, r << min(d, k), B$\in R^{d\times r}$, A$\in R^{r\times k}$, In the training process, W_0gradient update is no longer carried out, and A and B are trainable parameters.

In this paper, prompt engineering template is used to fine-tune the large language model. The main function of prompt is to accurately identify and generate keywords related to the original text from the paper abstract. The composition of the template is shown in the formula below.

$$Prompt = Instruct + Example + Input + Output \tag{2}$$

Instruct represents the description of the keyword generation task, Example represents the instance, Input represents the input text, and Output represents the output result requirement. In this paper, the prompt engineering design template is designed to clearly and specifically describe the task objectives and task contents in "Instruct", and the specific paper abstract and expected keyword generation results are given in "Example".

* Joint Workshop of the 2th Innovation Measurement for Scientific Communication (IMSC) in the Era of Big Data (IMSC2024), Dec 20th, 2024, Hong Kong, China and Online
Corresponding author.
†These authors contributed equally.
qiuxp7@mail2.sysu.edu.cn (Xinpeng Qiu); lijing359@mail.sysu.edu.cn (Jing Li)
0009-0002-2974-9608 (Xinpeng Qiu); 0000-0002-0004-907X (Jing Li)

2. Paper novelty score calculation

After early data acquisition, processing and fine-tuning of LLaMA3 large language model, this study summarizes the keywords of each sample paper generated, takes the publication year of the sample paper as the reference point, and uses the fine-tuning large language model LLaMA3 to calculate the semantic similarity. Specifically, We compared the occurrence frequency of keywords in other research literature in the same research field earlier than the publication time of the current sample papers. This comparison process uses the fine-tuned trained large language model LLaMA3 to ensure the accuracy and validity of the comparison. We record the frequency of these keywords and mark them as specific reference data. Let's call it. Substituting into the calculation formula, the score obtained is the semantic novelty score of the sample papers and the paper novelty measure is shown below.

$$Nov_n = \frac{\sum_{a=1}^{|Q|} \frac{1}{ln[n(Q_k)+1]+1}}{|Q|} \tag{3}$$

Where, Nov_n represents the novelty score of the paper to be tested, $n(Q_k)$ Represents the frequency of occurrence of keywords in the paper to be tested compared with that in the paper before publication.

3. Data collection and parameter setting

This paper uses the Web of Science core set as a data source, selecting scientific and technological papers from top-level Computer Science disciplines published in 2018-2019. The search criteria retrieved 15,348 papers. To evaluate semantic novelty, it obtained a complete database of the field, including paper titles, abstracts, citation frequencies, and JCR partitions. The study divided the sample dataset into a training set (6,100 papers from 2018, with subsets of 500, 1500, and 3000 samples) and a test set (9,300 papers from 2019) to build and evaluate a fine-tuning model.The GPU used in the experimental environment of this paper is NVIDIA A800 SXM4, Python version 3.10, and Pytorch version 2.0.1. The parameter Settings of the large model training in this paper are shown in Table 1.

Table 1
Parameter setting

Parameter	Number
Learning rate	2e-5
per_device_train_batch_size	4
per_device_eval_batch_size	4
num_train_epochs	3
evaluation_strategy	'steps'

4. Empirical analysis

4.1. Analysis with general language model

The current representative general large language models Gemma2, Phi3, GPT-4 and LLaMA3 are selected for comparison with the fine-tuning model in this paper, and the results are shown in Table 2.

Table 2
Comparison of keyword generation effect between fine-tuned language model and general language model

Method	Recall	Precision	F_1
Phi3	34.0%	46.0%	40.8%
Gemma2	30.0%	48.0%	36.9%
LLaMA3	32.6%	54.0%	40.7%
GPT-4	54.2%	68.0%	60.3%
Ours	**66.0%**	**81.0%**	**72.7%**

Conclusion: The fine-tuned LLaMA3 can generate better paper keywords.

4.2. Ablation experiment

Table 3
Comparison of keyword generation effect

Method	Recall	Precision	F_1
LLaMA3	32.6%	54.0%	40.7%
Ours/LT	58.0%	63.5%	60.6%
Ours/PT	63.0%	75.0%	68.5%
Ours	**66.0%**	**81.0%**	**72.7%**

Conclusion: LoRA fine-tuning and hint word fine-tuning can significantly improve the generation effect of the model in this paper.

4.3. Analysis of validity of paper novelty results

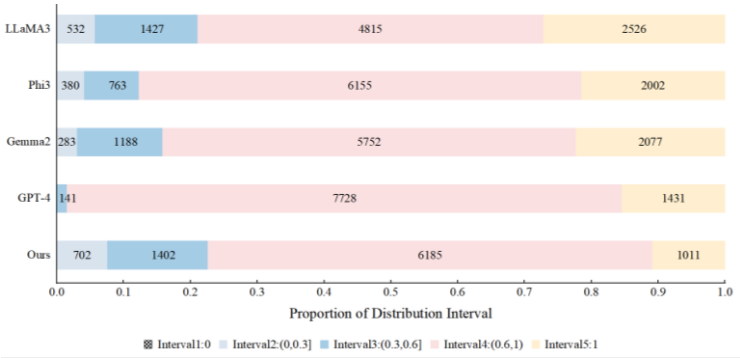

Figure 1: Interval plot of novelty score distribution based on different model measures

Conclusion: The fine-tuning model proposed in this paper can significantly improve the discrimination and accuracy of novelty score measurement in the application of text keyword generation, and can achieve the effect of model improvement.

Besides, in this study, the Delphi method is used to rank the 10 papers with high and 10 papers with low semantic novelty out of order and score them to the experts.

5. Discussion

Through the above empirical analysis, it is proved that the fine-tuned LLaMA3 can effectively improve the keyword generation effect of the paper, and further optimize the semantic novelty measurement of the paper.

Acknowledgements

This research is supported by grants from the National Social Science Foundation of China (22BTQ097).

Declaration on Generative AI

During the preparation of this work, the author(s) used LLaMA3 in order to: The text data is processed and the abstract keywords are extracted. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

[1] Chuanjun Suo, Miao Yu, Yanxin Pai and Juntao Rong. 2024. Theoretical framework of data-driven academic evaluation. Libr. Inf. Serv. 68, 1 (Jan, 2024), 5-12. DOI: https://doi.org/10.13266/j.issn.0252-3116.2024.01.001.

[2] Zara Nasar, Syed Waqar Jaffry and Muhammad Kamran Malik. 2018. Information extraction from scientific articles: A survey. Scientometrics 117, 3 (Sept, 2018), 1931-1990. DOI: https://doi.org/10.1007/s11192-018-2921-5.

[3] Jiajia Qian, Zhuoran Luo and Wei Lu. 2021. Novelty measurement and innovation type identification of scientific literature based on question-method combination. Libr. Inf. Serv. 65, 14 (Jul, 2021), 82-89. DOI: https://doi.org/10.13266/j.issn.0252-3116.2021.14.010.

[4] Hong Huang, Chong Chen and Jingying Zhang. 2022. Review on identifying the semantics of scientific literature content. J. China Soc. Sci. Tech. Inf. 41, 9 (Jan, 2022), 991-1002. DOI: https://doi.org/10.13266/j.issn.0252-3116.2024.01.001.

[5] Peter D. Turney. 2000. Learning algorithms for keyphrase extraction. Inf. Retr. 2, 4 (May, 2000), 303-336. DOI: https://doi.org/10.1023/A:1009976227802.