Research on Fine-grained S&T Entity Identification with Contextual Semantics in Think-Tank Text

Mengge Sun[1,2], Yanpeng Wang[1,2] and Yang Zhao[1,2]

[1]National Science Library, Chinese Academy of science Beijing 100190

[2]Department of Information Resources Management, School of Economics and Management, University of Chinese Academy of Sciences Beijing 100190

Abstract

Automatically extracting fine-grained S&T problems from think-tank reports written by numerous experts, has become one of the effective ways to perceive the global trend of S&T development. We transform the automatic identification task for fine-grained S&T problems into a multi category S&T entity extraction task with contextual semantics. To address the shortage of high-quality data sets and fully exploit the potential of LLMs, we take LLMs as annotators and puts them into an active learning loop to determine which samples to annotate efficiently. During the cyclic data annotation process, we simultaneously trained the target's entity extraction model "RoBERTa-BiLSTM-CRF". Finally, the model achieved an F1 value of 86.02% in our task. The effectiveness and reliability of the model were verified by comparing it with the benchmark model through experiments. This study to some extent solves the problem of manually annotating dataset dependencies, while providing high-quality data support and effective model methods for mining and analyzing fine-grained S&T problems.

Keywords

S&T entity with contextual semantics, LLM annotators, active learning, RoBERTa-BiLSTM-CRF,

1. Introduction

In order to fully seize the historical opportunities of each technological revolution, governments around the world have been focusing on key scientific and technological(S&T) development direction, by launching a series of policies, strategies, plans, and specific actions files to encourage industry, academia, and research institutions to undertake research and development challenges. These files, interpreted by professionals, are collectively referred to as think-tank files. The ultimate goal is to achieve original innovation breakthroughs and control over key technologies, in hopes of securing a technological first-mover advantage.

In this context, this paper strives to objectively perceive the key technological development directions currently being focused on by various countries. Therefore, this paper emphasizes defining the concept of "fine-grained technological problem" and investigating how to use deep learning methods to automatically construct a "problems bank" of fine-grained technological innovation. This resource is intended for further exploration by intelligence analysts and subsequent evaluation by decision-makers.

the 5th Extraction and Evaluation of Knowledge Entities from Scientific Documents (EEKE2024), April 22 - 26, 2024, Changchun, China

*Corresponding author.

wangyanpeng@mail.las.ac.cn (Y. Wang)

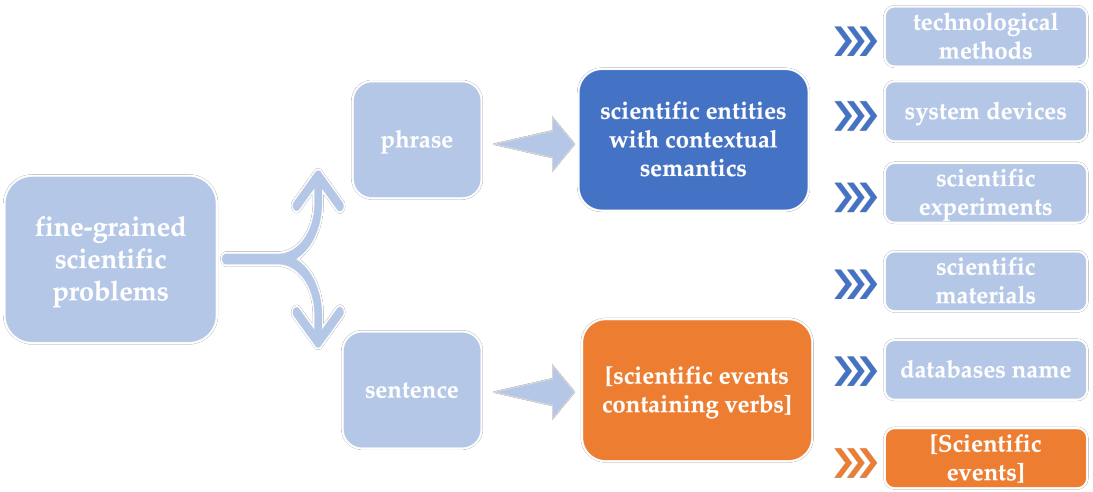

Figure 1: Conceptual analysis of fine-grained S&T problem

This paper defines ”fine-grained S&T problems” as ”research directions or problems with limited conditions such as application scenarios, technological solutions, and technological routes”, and further analogizes them as ”S&T entities with contextual semantics”. S&T entities can be further subdivided into multiple finer grained categories. Depending on the type of scientific solution sought, they can be distinguished into: identification and judgment about the research object and the inherent mechanisms and laws of research, as shown in Fig.1. Correspondingly, the research objects include ”technological methods”, ”system devices”, ”scientific experiments”, ”scientific materials”, and ”databases name”. Examples include ”cell-based cancer immunotherapy and gene therapy”, ”ferrosilicon alloy latent heat photovoltaic cells”, ”deep underground neutrino experiments” and ”two-dimensional materials for future heterogeneous electronic devices”. And the underlying mechanisms of things, such as “the principle of evolution controlled from top to bottom”.

In addition, some technological problem also manifest in the form of ”S&T events containing verbs”, such as ”natural language processing algorithms will be used to study the principle of viral gene mutations”, which is not within the scope of this study.

2. Related works

The think tank is composed of multidisciplinary experts in a country and gathers national intellectual resources, which is an important force to influence government decision-making and promote social development. Usually, think tank reports tend to focus on major issues of great concern to the national government or the public, which are generated through tracking, integrating and compiling online news, research reports, journal papers, newspaper articles, policies, regulations and speeches, and other open information. To a certain extent, think tank reports represent indicators and weather vane of national policies and scientific research, and

have high intelligence values. Therefore, the automatic extraction of scientific and technological entities mentioned in think tank reports can further clarify policy and public concerns efficiently and objectively.

The automatic identification of fine-grained S&T problem is essentially a task of information extraction. This typically involves constructing annotated datasets and designing effective identification methods. Therefore, this article mainly conducts related research from two perspectives: dataset annotation and information extraction.

Most of the S&T problem representations extracted by researchers in the past have focused on the extraction of scientific research problem sentences, adopting several methods such as manual annotation, rule-based matching, machine learning-based, hybrid model-based, and deep learning-based methods. H. Chu and Q. Ke [1] used manual annotation to analyze the distribution of methods (mainly data collection methods) in different academic journals. However, those expert annotation methods are relatively highly accurate, but costly and time-consuming. S. Gupta and C D. Manning [2] designed matching rules for identifying research problem , including using the word "applied" for rule matching, and then using the Bootstrapping method to find new rule templates based on the newly matched vocabulary. K. Heffernan and S. Teufel [3] treated scientific method identification as a classification task, using classification algorithms such as support vector machines, Naive Bayes, and logistic regression, and introduced features such as N-gram, sentiment polarity, part of speech, whether it is a negative word, discourse information (such as the likelihood of method words in the introduction and future work), and part of speech into the algorithm to enhance its performance. Semeval 2018 Task7 [4] also conducted extraction of various types of entities in academic papers. In this task, many teams used convolutional neural networks and Long Short-Term Memory networks to achieve performance superior to traditional machine learning methods (such as SVM), which also proved the usefulness of deep learning models. In terms of deep learning methods, Xuesi Li et al. [5] designed a sentence classification model based on the BERT-CNN architecture, and automatically identified research issue sentences in scientific papers with an F1 value of 94.8%. Z. Zhong and D. Chen [6] compared the performance of BERT and SciBERT, two pre-trained language models, in the extraction of relations in academic papers, and found that SciBERT performed better than BERT.

Since 2020, large language models (LLMs) have exhibited remarkable few-shot performance in information extraction tasks, with only a few demonstrations and well-designed prompts. However, with rapid advancements comes vast potential risks in adopting LLMs for widespread downstream production applications. One of the main concerns is about data privacy and security. Under the prevalent "Language-Model-as-a-Service" (Sun et al. 2022) setting, users are required to feed their own data, potentially including sensitive or private information, to third-party LLM vendors to access the service, which increases the risk of data leakage.To exploit the abundant unlabeled corpus, an alternative is to employ LLMs as annotators, which generate labels in a zero-shot or few-shot manner.

Table 1
Sample data sources and S&T entities

Title	Contents
The National Science Foundation has launched a joint initiative to build an Integrated Data and Knowledge Infrastructure	The initiative will cost 20 million dollars to build **a prototype of an integrated data and knowledge infrastructure** called the Open Knowledge Network. The Open Knowledge Network is a set of publicly accessible, interconnected data repositories and related knowledge maps that will provide data-driven, AI-based solutions to a wide range of societal challenges.
Ten hot spots in US science and technology policy in 2023	Looking beyond 2035, it is recommended that the United States prepare to build **an electron-positron "Higgs factory", a high-energy muon or Hadron collider, and a next-generation gravitational-wave observatory**.
NASA awarded TransAstra an 850,000 dollars contract to produce a capture bag that can capture debris in orbit	NASA awarded TransAstra an 850,000 dollars contract to produce **a capture bag that can capture debris in orbit** and demonstrate on the ground how the device can open and close.

3. Methods

3.1. Data

The selected data source is high-quality strategic dynamic briefing data monitored and compiled by various departments of the Chinese Academy of Sciences and the State Council, which is available on the agency's website[1] . The data source includes: (1) the trends of top scientific journals, showcasing the latest scientific research achievements in disciplines such as physics, Earth, and biology; (2) the latest strategic deployments of various countries in the field of S&T, representing the direction of national S&T development.

These information contents can to some extent represent the will of the country and scientists [7]. To complement each other's perspectives, we obtained high-quality expert compiled data from various research institutes, such as the S&T Frontier Bulletin, comprehensively reflecting the overall trend of global S&T development. Data examples are shown in Table1. Finally, we crawled all the information from the three sites from 2018 to 2023, totaling 42,984 reports, with an average of about 12 sentences per report.

[1]http://www.casisd.cn/zkcg/ydkb/kjqykb/
https://news.sciencenet.cn/AInews/newlist.aspx?
http://www.globaltechmap.com/document/index

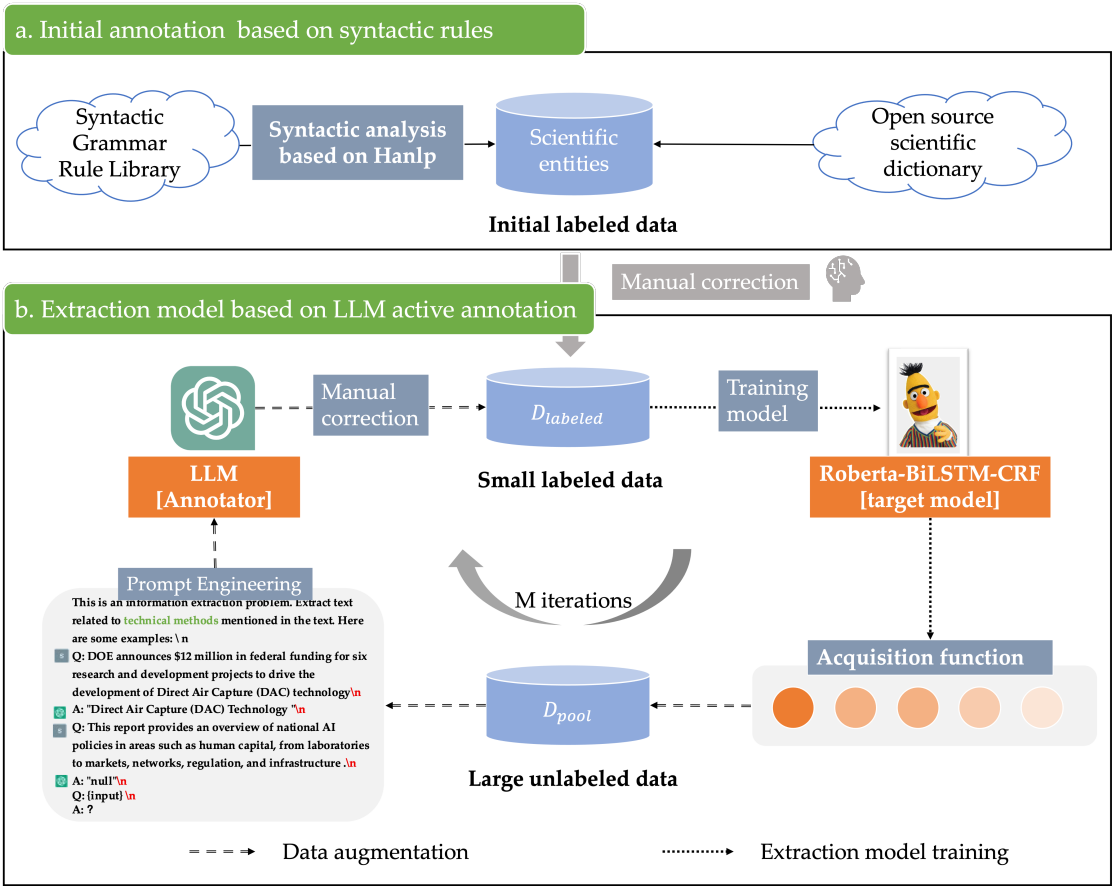

Figure 2: Main research framework

3.2. Main Framework

Based on the above data sets, the research work of this paper mainly includes three parts: initial annotation based on syntatic rules, active annotation based on LLM, and train extraction model during active learning process. As shown in figure 2.

a) In the part of initial annotation based on syntactic rules, we mainly uses a rule-based extraction method as a cold start, combined with manual correction, to obtain a small amount of high-quality contextual S&T entities databases.

b) In the extraction part of based on LLM active annotation, the main goal is to gradually fine screen a small-scale annotated data from a large amount of unlabeled data, while using a large language model as the annotation model. At the same time, a S&T entity extraction model called "Roberta-BiLSTM-CRF" is trained. In general, the entire process is carried out in m iterations, each of which mainly consists of two parts:

1) using the prompt engineering technology of LLM to assign labels to some unlabeled data, and then using manual correction to remove noisy data;

2) training the target model based on the labeled data obtained, and select the data to be

Table 2
Examples of lexical and syntactic rules

Guide words	Syntax rules
exploit	improve.*by.*?
promote	push.*by.*?
propel	strengthen.*by.*?
focus on	understand.*by.*?
support	prepare.*by.*?
set up	based on.*?
realize	.*? for.*
facilitate	use.*? to develop.*

annotated in the next iteration using the acquisition function mechanism. Among them, the target model uses the Chinese RoBERTa WWM model as the embedding model, and the BiLSTM model and CRF model as the label sequence prediction layer to obtain the label sequence of S&T entities and complete the automatic extraction of fine-grained S&T problem.

Finally, evaluate the model results based on soft matching strategy.

3.3. Initial annotation based on syntactic rules

The initial annotation process based on syntactic rules is as follows: after dividing the think-tank text into sentences, the candidate scientific entities with contextual semantics are first extracted by constructing a lexical and syntactic rule library, with the principle of maximizing the high recall rate. Then, by constructing a scientific dictionary, the "scientific degree" index of each phrase mentioned above is calculated as the final annotation result, greatly saving manpower in annotation work. It is specifically divided into three parts:

(1) Build annotation rules

After the process of "model output results - manual review - rule modification", the syntax and grammar rule library was repeatedly summarized and modified. As of now, there are a total of 162 lexical and syntactic rules, as shown in Table 2. Then, combined with the dependency syntax analysis function of a pretrained HanLP model, candidate scientific entity phrases with contextual semantics are obtained.

(2) Build a S&T dictionary The extracted phrases in think-tank texts may not necessarily be S&T problem phrases. Determine whether the phrase is a S&T problem phrase based on whether it contains S&T terms. The dictionary mainly comes from the open source S&T dictionary: including 17 fields of S&T terms - ship engineering, geographic surveying and mapping, electronic engineering, steel and metallurgy, industrial design, aerospace, chemical engineering, environmental science, mechanical engineering, computer industry, mining exploration, mathematical science, hydraulic engineering, astronomical science, communication engineering, and physical science, with a total of 179,146 items.

(3) Identify S&T problems by calculating the "scientific degree " index

This article defines the "scientific degree" index to represente the degree of overlap between candidate S&T problems and S&T dictionaries. Among them, w_i represents the number of words

in the i_{th} candidate phrase in the S&T dictionary, and n_i represents the number of words in the i_{th} candidate phrase. The " scientific degree" indicator t_i is defined as: $t_i = w_i/n_i$. After multiple tests, we sets a threshold of 70% for "scientific degree" to determine that the candidate phrase is a S&T problem. Eventually, through cross-review by two graduate students, we finally obtained a supervised dataset.

3.4. Extraction model based on Roberta-BiLSTM-CRF

The S&T problem databases annotated under the above two strategies were manually validated to obtain a training dataset. Then, the BIO annotation standard is used to annotate it and input it into the deep learning model.

Training corpus generation

The data annotation section in this study uses the BIO annotation standard. At the same time, location information and category information are connected by "-", such as the device name "high energy hadron collider" is labeled with the device name EQU, and the corresponding tag sequence is ['B-EQU', 'I-EQU', 'I-EQU', 'I-EQU']. The original sequence and tag sequence together form the learning corpus. Afterwards, the learning corpus is divided into a training set, a testing set, and a validation set in an 8:1:1 ratio.

Extraction model structure

The model structure is mainly divided into two parts: the embedding feature encoding layer and the label sequence prediction layer.

Embedding feature encoding layer based on Roberta

RoBERTa-WWM [8] is an improved model based on BERT, which has the same framework as BERT. However, due to the use of improved training strategies, the accuracy on numerous datasets has been improved by 5% to 20% compared to BERT. Improvements have mainly been made in two aspects:

(1) Adopting the training strategy of WWM (whole word masking). When generating training samples, unlike the original WordPiece based Mask method, which divides a complete word into several words, the WWM strategy randomly masks the entire word by phrase, which is more in line with the organization of Chinese.

(2) Adopting dynamic masking method. The BERT model will randomly masks a certain number of tokens during the data preprocessing stage; However, RoBERTa-WWM uses a dynamic masking approach, which generates a new masking sequence for each input to the model. During this training process, the model gradually adapts to different masking strategies and learns different language representations, which is beneficial for improving word embedding performance.

Label sequence prediction layer based on BiLSTM and CRF

The BiLSTM model consists of forward LSTM and backward LSTM, which can capture long-distance dependencies between input character sequences, but lacks consideration for dependencies between label sequences. However, Conditional random fields can perform the best joint prediction on the label sequence corresponding to characters, capturing the relationship between labels and effectively solving the problem of label bias. Therefore, there is complementarity between the BiLSTM model and the CRF model. That is, after completing the feature learning of think-tank texts based on the BiLSTM model, the hidden state sequence

needs to be used as the input of the CRF model to ultimately obtain the best labeled prediction sequence.

3.5. Data augmentation strategy based on LLM

Next, this article leverages the semantic understanding ability of GPT3.5 in semantic information extraction, annotates S&T problems, and expands the supervised training dataset of the extraction model. It mainly consists of two parts: first, selecting the optimal prompt and annotating data using prompt engineering; The second is to remove noisy data (data with inaccurate annotations) through robust learning.

Active data acquisition

Active learning (AL) seeks to reduce labeling efforts by strategically choosing which examples to annotate. We consider the standard pool-based setting, assuming that a large pool of unlabeled data D_{pool} is available. AL loop starts with a seed labeled set $D_{labeled}$. **At each iteration, we train a model** M **on** $D_{labeled}$ **and then use acquisition function** $f(\cdot, M)$ **to acquire a batch** B **consisting of** b **examples from** D_{pool}. We then query the LLM annotator to label B. The labeled batch is then removed from the pool D_{pool} and added to labeled set $D_{labeled}$, and will serve as training data for the next iteration. The process is repeated for m times.

Active acquisition strategies generally maximize either uncertainty or diversity. On one hand, uncertainty-based methods leverage model predictions to select hard examples. On the other hand, diversity-based methods exploit the *heterogeneity of sampled data.*

Maximum Entropy. Entropy is one of the most widely used estimations of uncertainty (Settles 2009). Data for which the model M has the highest entropy are sampled for annotation according to

$$\underset{x \in D_{pool}}{\operatorname{argmax}} - \sum_{y \in Y} P_M(y|x) \log P_M(y|x)$$

.

Least Confidence. [9] propose to sort examples with the probability assigned by M to predicted class $\hat{y}$, which samples

$$\underset{x \in D_{pool}}{\operatorname{argmax}} \left(1 - P_M(\hat{y}|x)\right)$$

.

K-Means. Diversity sampling intends to select batches of data that is heterogeneous in the feature space. We apply k-means clustering to the L2-normalized embeddings of M, and sample the nearest neighbors of the k cluster centers.

Optimizing LLM as better annotator According to literature research, it has been found that the current GPT series models are highly sensitive to different PROMPT expressions. When different annotators use different PROMPT expressions, there is a significant difference in the response results of GPT. The robustness of the model on NLP tasks is relativaly weak [10]. Previous studies show that the design of task-specific prompts varies between near state-of-the-art and random guesses [11]. Therefore, finding the best prompts for given tasks and given data points is very critical.

This paper adopts the Chain of thought (CoT) prompts strategy, which gradually generates label sequences that meet expectations by setting some conditions in each model. Guided by

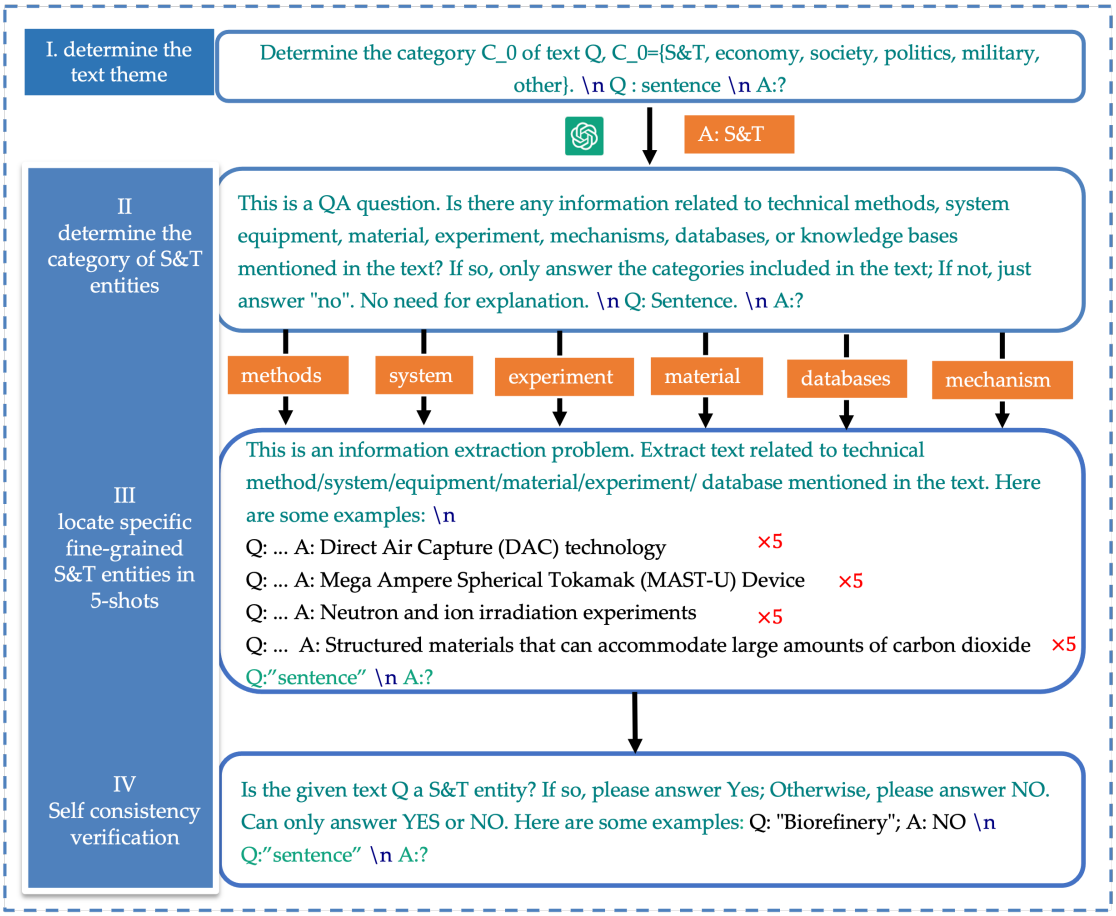

Figure 3: Flowchart of GPT annotation technology issues under CoT prompt strategy

the CoT approach, this article transforms this task into a multi round Q&A question, enabling the GPT model to gradually locate the fine-grained categories of S&T entities contained in the text through conversation, and finally annotate them. Specifically, this chapter focuses on the construction process of PROMOPT for different categories of S&T problems, as shown in Figure 3.

Manual correction with noisy labels LLM annotators inevitably produce noisy labels, especially with domain-specific data. To stay robust against training label bias, we use manual proofreading to correct the LLM annotated data.

3.6. Verification of annotation effect

To cater to human preferences, the answers generated by the GPT series models tend to identify longer spans of S&T issues than annotated ones. Therefore, the previous hard matching strategy was not suitable for evaluating the output results of GPT series models. Therefore, this paper proposes a **soft matching strategy**, and the algorithm process is as follows. The algorithm

indicates that as long as there is an inclusion relationship between the generated technology problem (span) and the labeled technology problem (span) and the similarity threshold is reached, the result is considered correct, as shown in table 3.

Table 3
Algorithm : Soft-Matching Strategy

steps	descriptions
0	**Input**: the sentence s, the list of annotated spans L_A in sentence s, a predicted span p, the similarity threshold γ
1	**Output**: Return True if just the offsets of two spans are different and similarity is greater than γ, otherwise return False
2	Similarity ← []
3	for t in L_A:
4	score ← GetSimilarity(t, p)
5	Similarity.append(score)
6	score, max_index ← max(Similarity)
7	t ← L_A [max_index]
8	if p contains t or t contains p:
9	if score > γ:
10	return True
11	return False

The S&T problems that have been matched through the "soft matching" method, continue to undergo manual testing and repeatedly train to obtain the optimal PROMPT module library.

4. Results and discussion

4.1. Analysis of data annotation results

Initial supervised data based on the rule annotation In the data annotation based on statistical rules, a total of 7540 sentences of 352 S&T think-tank texts (include frontier and strategy reports) released in 2018 were selected as the test set. It was found that the extraction effect of the model was: accuracy: 0.36; recall rate: 0.82; F1 value: 0.50. That is to say, the majority of S&T phrases annotated by statistical rule-based annotation methods are not within the category of S&T, and their level of S&T cannot be accurately judged. Therefore, an AI model that can deeply understand and analyze semantics is particularly needed for annotation and extraction.

Further analysis of the annotation effect reveals that this method has two main shortcomings: firstly, it heavily relies on text types, and the extraction effect in S&T frontiers texts is significantly better than that in strategic texts. The second is that the rules cannot be exhausted. Finally, this article obtained 2680 fine-grained S&T entities datasets through manual proofreading of annotated data, in order to test and optimize the annotation effect of the LLM.

Analysis of LLM annotation results

Firstly, we randomly selected 20 texts from the annotated dataset as the test set to determine the number of examples in the Few-shot strategy, as shown in Table 4. Here, we choose GPT3.5

Table 4
GPT3.5 model extraction test results under the Few prompt strategy

Strategy	1-shot	2-shot	3-shot	4-shot	5-shot
Precision	30.6	29.1	40.1	59.1	76.3

Table 5
Performance of the Large Model at Each Prompt Stage in final iteration

ID	Precision	Recall Rate
Stage I: Determine the theme to which the text belongs	100.0	–
Stage II: Determine the category of S&T entities contained in the technology text	90.87	–
Stage III: Extract and locate specific fine-grained S&T entities expressions	71.20	88.41
Stage IV: Confirm its output result with the GPT model again	92.01	–

as the representative of LLM. The test shows that the more relevant and semantically similar the given examples are to the test text, the better the annotation effect of GPT3.5. In the 1-shot scenario where an example is given, the performance of the given example is sensitive and unstable to GPT3.5; Overall, 5-shot prompt performs better because combining multiple random examples can reduce the impact of noise.

After determining the number of given examples in the Few shot strategy, we conducted multiple tests to select the most effective example for each stage of the PROMPT. The performance of each prompt stage is shown in Table 5.

(1) In terms of category judgment, GPT's performance is almost perfect. That is to say,for classification tasks with more popular query semantics and more obvious semantic differences, the GPT model which be trained by 175 billion Internet open source data has better performance.

(2) In terms of information extraction, GPT has lower accuracy and higher recall. There are two main manifestations of low accuracy: firstly, the extracted S&T entities are mainly in the form of nouns phrases, which are not comprehensive, such as "natural language processing algorithms will be used to study the principle of virus gene mutation"; second, the extracted S&T entities still have different granularities, too coarse or too fine. For example, "multiple dashes are likely a combination of a bunch of technologies," which often have too coarse granularity.

Finally, after multiple rounds of annotation and manual proofreading, a total of 352 frontier and strategic S&T reports, totaling 7540 sentences, were released in 2018; 123 strategic reports released in 2019, with a total of 2510 sentences; 4395 frontiers reports released in 2020, totaling 9695 sentences; Finally, a total of 19745 sentences formed a supervised training dataset.

4.2. Analysis of Model extraction effect

Dataset we chose 2680 fine-grained S&T entities datasets as seed labeled set $D_{labeled}$ from initially annotated dataset, use the whole 19745 sentences as D_{pool} and randomly acquired 100 samples per batch for 10 iterations, which generate 9,921 annoted samples in total.

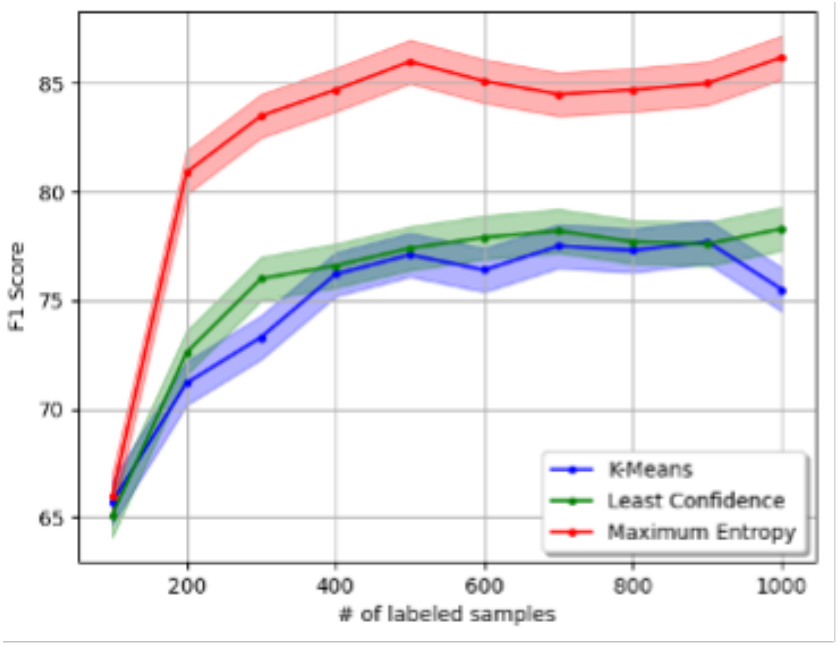

Figure 4: the model performance with different active acquisition strategies

Baselines We compare RoBERTa-BiLSTM-CRF with the following baselines: (1) In-context learning (i.e. PROMPTING). The PROMPTING enables LLM to conduct few-shot inference without fine-tuning. (2) SUPERVISED(i.e. BERT-BiLSTM-CRF). The surpervised model is trained on whole clean-labeled data $D_{labeled}$.

Accelerating with Active Learning The last layer in the above extraction model is the CRF model, whose output result is the probability score of the BIO label corresponding to each character. Here, we use this probability score as the confidence score and input it into two uncertainty based active learning strategies. Figure 4 shows the performance of the extraction model under three active learning strategies.

In figure 4, uncertainty-based methods, i.e. maximal entropy and least confidence, perform significantly better than the K-Means(where $k = 26$), with faster convergence and higher F1 scores at the end of iterations. This suggests that it may require more training data for K-Means to learn informative representations, and such a diversity-based method may fail in low-resource environments. In summary, maximal entropy active learning strategies enable extraction model to be more efficient and more capable.

Implementation The backbone of the benchmark model, RoBERTa-WWM (ext-large version), that we utilize is a 12-layer stacked bidirectional Transformer. The Adam optimizer is employed for gradient optimization and parameter updates during the training phase. Notably, due to the limitation of the maximum input length of the BERT series model to 512, we adopt the ”sliding window” approach to process the data. This method samples and constructs multiple sub-samples in the form of sliding windows to prevent semantic loss caused by direct truncation. The remaining model parameters are set as follows:

Table 6
Model comparison experiment results

METHOD	Precision	Recall Rate	F1 Value
PROMPTING	67.72	76.72	67.72
BERT-BiLSTM-CRF	70.54	75.66	73.00
Chinese-Roberta-WWM-BiLSTM-CRF(last iteration)	**82.20**	**90.23**	**86.02**

- Hidden_dim: 128
- Learning_rate: 3×10^{-5}
- Batch_size: 32
- Epoch: 64
- Dropout: 0.5
- Embedding_len: 768

In conclusion, we choose the Chinese-Roberta-WWM model as the feature encoding layers, and combine them with the "BiLSTM+CRF" model as the comprehensive architecture. The results of the S&T entities extraction tasks are shown in Table 6.

5. Conclusions

To some extent, the think tank report serves as the knowledge intermediary between policy orientation and S&T research. Automatic extraction of contextual technology entities with contextual semantics from think tank reports can more efficiently capture key research development directions. In this paper, GPT is used as the teacher model and Roberta-Bilstm-CRF is used as the student model. Through active learning method, the training data generated by GPT is fine-tuned to the local extraction model, forming a set of feasible fine-grained S&T entity recognition framework.

The biggest limitation of our study is that it mainly focuses on the discussion of the effectiveness of the method, and the standard of high accuracy has not been reached in practical engineering applications, and the model effect will continue to be optimized in the future.

References

[1] H. Chu, Q. Ke, Research methods: What's in the name?, Library & Information Science Research 39 (2017) 284–294.

[2] S. Gupta, C. D. Manning, Analyzing the dynamics of research by extracting key aspects of scientific papers, in: Proceedings of 5th international joint conference on natural language processing, 2011, pp. 1–9.

[3] K. Heffernan, S. Teufel, Identifying problems and solutions in scientific text, Scientometrics 116 (2018) 1367–1382.

[4] D. Buscaldi, A.-K. Schumann, B. Qasemizadeh, H. Zargayouna, T. Charnois, Semeval-2018 task 7: Semantic relation extraction and classification in scientific papers, in: International Workshop on Semantic Evaluation (SemEval-2018), 2017, pp. 679–688.
[5] Y. L. Y. W. Xuesi Li, Zhixiong Zhang, Research on problem sentence recognition methods in scientific literature research, Library and Information Service 67 (2023) 132–140.
[6] Z. Zhong, D. Chen, A frustratingly easy approach for entity and relation extraction, arXiv preprint arXiv:2010.12812 (2020).
[7] X. C. Y. L. X. L. Yanpeng Wang, Xuezhao Wang, Analysis of key technologies and initiatives of the fourth industrial revolution based on science and technology policy and frontier dynamics, Journal of the China Society for Scientific and Technical Information 41 (2022) 29–37.
[8] Y. Cui, W. Che, T. Liu, B. Qin, Z. Yang, Pre-training with whole word masking for chinese bert, IEEE/ACM Transactions on Audio, Speech, and Language Processing 29 (2021) 3504–3514.
[9] A. Culotta, A. McCallum, Reducing labeling effort for structured prediction tasks, in: AAAI, volume 5, 2005, pp. 746–751.
[10] J. Gao, H. Zhao, C. Yu, R. Xu, Exploring the feasibility of chatgpt for event extraction, arXiv preprint arXiv:2303.03836 (2023).
[11] T. Gao, A. Fisch, D. Chen, Making pre-trained language models better few-shot learners, arXiv preprint arXiv:2012.15723 (2020).