Is Prompt the Future?
A Survey of Evolution of Relation Extraction Approach Using Deep Learning and Big Data

Zhen Zhu, College of Information and Control, Keyi College of Zhejiang Sci-tech University, Shaoxing, China & AI Lab, NetEase Inc., Hangzhou, China
Liting Wang, College of Information and Control, Keyi College of Zhejiang Sci-Tech University, China
Dongmei Gu, Information and Educational Technology Center, Zhejiang Unyi College, Zhejiang Sci-Tech University, China
Hong Wu, College of Information and Control, Keyi College of Zhejiang Sci-tech University, Shaoxing, China*
Behrooz Janfada, School of Computer Engineering, Iran University of Science and Technology, Iran
Behrouz Minaei-Bidgoli, School of Computer Engineering, Iran University of Science and Technology, Iran

ABSTRACT
A vast amount of unstructured data is being generated in the age of big data. Relation extraction (RE) is the critical way to improve the utility of the data by extracting structured data, which has seen a great evolution in recent years. This paper first introduces five paradigms of RE, namely the rule-based paradigm, the machine learning paradigm, the deep learning model-based paradigm, and the two types of current mainstream methods with pretrained language models. Based on the RE scenario, a comprehensive introduction is made for the currently popular paradigm with prompt learning, which is investigated regarding four aspects. The main contributions of this paper are as follows. Since big models are too large to be easily trained, prompt learning has become a promising research direction for RE, our work is, therefore, a systematic introduction to this paradigm for RE and compared with traditional paradigms. In addition, this paper summarizes the current problems faced by RE tasks and proposes valuable research directions with prompt learning.

KEYWORDS
Big Data, Deep Learning, Pretrained Model, Prompt Learning, Relation Extraction

INTRODUCTION
With the Internet growing hugely popular in the age of big data, a vast amount of data is being generated every second. The data existing in various forms like social networks, video websites, news, advertisements, etc., are released by publishers on the network with specific purposes. It is of great value to extract effective information from these various data (Niklaus et al., 2018) to provide potential utility to human society, such as domain knowledge graph construction, public opinion monitoring, and problem analysis and diagnosis (Wang et al., 2014). Changing these unstructured data to structured data and extracting key information automatically (Simoes et al., 2009) is the information extraction
The information extraction task contains three main important tasks: word separation, named entity recognition (NER), and relation extraction (RE; Muslea et al., 1999). The topic of this paper is developing with relation extraction in mind.

Relation extraction aims to extract entity relation facts in a specific form (e.g., in the form of a triple). For example, “OpenAI was founded in the late 2015, and is headquartered in San Francisco, California.” It is possible to extract that OpenAI is located in San Francisco, California (OpenAI headquarters, located in San Francisco, California). Relation extraction is a fundamental task for many downstream tasks, such as machine translation (MT; Bordes et al., 2013), Q&A systems (Bordes et al., 2014; Li et al., 2015), search engines, etc. (Xiong et al., 2017; Schlichtkrull et al., 2018). There are various methods for entity relation extraction, including the rule-based template approach (Califf & Mooney, 1997), machine learning approaches (Kambhatla, 2004), deep learning model-based approaches (Zeng et al., 2014; Zhang et al., 2015), deep learning network model-based approaches with pretrained models, relation extraction approaches based on large-scale pretrained models for relation extraction methods (Zeng et al., 2017) that brought great changes in recent years.

Despite the fact that the relation extraction task has evolved for many years and has made significant progress, it remains to be fully addressed and is an area calling for continuous research and breakthroughs. The existing problems include insufficient labeled data and the intractable noise problem induced by distant-supervised approaches. Poor extraction performance also arises in massive relation categories under open domains. These problems lead to limited practical applications of relation extraction in industry. However, with the successful application of large models in several tasks in the natural language domain this year, this paper argues that a breakthrough in relation extraction is forthcoming in just a few years. A large model’s strong generalization ability, along with new paradigms such as prompt learning, will hopefully overcome the limitations such as insufficient labeled training data and too many open relation classifications, which can ultimately solve various practical problems in production and encourage wide use of relation extraction in industry. Since prompt learning with big models become a promising research direction for RE, our work is, therefore, a systematic introduction to this paradigm for RE and compared with traditional paradigms. In addition, our paper summarizes the current problems faced by RE tasks and proposes valuable research directions with prompt learning.

EVoLUTIoN oF E NTITy RELATIoN EXTRACTIoN APPR oACH

Relation extraction techniques were first proposed in the 1990s, and 30 years of evolution have dramatically changed the paradigm of relation extraction. Early relation extraction mainly relied on template matching methods, and the construction of manual rule templates played a crucial role in the results. By the 2000s, traditional machine learning methods began to be popularized, and feature-based models such as SVM and maximum entropy were used extensively (Lafferty et al., 2001; Och et al., 2004). For researchers and domain experts, it was a mainstream practice to use domain knowledge to extract features from the original dataset and then perform relation prediction with a traditional machine learning model of classification or clustering. Later there appeared neural network models (Bengio et al., 2013; Chung et al., 2014), such as CNNs and RNNs, which led to a dramatic change in the paradigm of relation extraction. Around 2012, neural network models were introduced to the work on relation extraction. At that time, the extraction work saw its focus shifting to the model architecture. A set of network architectures that enable learning to extract relational features were designed to improve the performance of extraction. However, by 2018 when large models started to prevail, the approach to relation extraction changed again, with many researchers shifting their focus to pretrained models combined with fine-tuning approaches (Dong et al., 2019; Lewis et al., 2020). In this paradigm, a generic model with a fixed architecture and large size is pretrained (Peters et al., 2018) in the first place. Additional training data was then introduced for fine-tuning for the specific task of relation extraction. Thus, relation extraction was finally achieved. In 2022, fine-tuning...
became difficult as larger models with higher performance became available, and many common lab-level machines could no longer support fine-tuning efforts. Then a new paradigm of pretrained models combined with prompt learning was proposed (Brown et al., 2020), bringing a second change to relation extraction, as shown in Figure 1. Because of the limitations of personal knowledge and available space, we focus only on some representative approaches under each paradigm, and there is still some very good relation extraction research not covered here. The major relation extraction approaches are summarized in Table 1.

Overall, deep learning-based pretrained language models have brought major breakthroughs to the task of relation extraction. With the advent of new models and methods, the performance of relation extraction is constantly improving. However, the existing methods have yet to address the following problems: insufficient labeled data, difficulties in open-domain extraction, low learning efficiency for resource-sparse scenarios, and difficulties in cross-sentence extraction. The paradigm of using prompt learning with big pretrained language models is a promising direction to overcome the above problems.

**PROMPT LEARNING OF RELATION EXTRACTION**

On November 30, 2022, OpenAI released a chat-bot known as ChatGPT (Chat Generative Pretrained Transformer) in the United States, which can talk to people very smoothly and perform some complex tasks that were previously impossible, and its powerful performance in application scenarios such as automatic programming and text generation has brought generative AI to unprecedented heights (Knox & Stone, 2011; Knox & Stone, 2009). Then, GPT-4, released on March 14, 2023 (Freedman & Nappier, 2023; Bubeck et al., 2023), reached a new height in reading comprehension (Katz et al., 2023). The aforementioned progress has created a considerable stir in the academic community. Big models such as GPT, BART, and BERT achieved the best results in several natural language domains. Therefore, new paradigms incorporating big language models have become very important research directions in the field of relation extraction.

**Model Selection in Prompt Learning**

**Evolution of Pretrained Language Models**

Pretrained language models are not a new concept, as pretraining methods have been studied as early as 2010 when they were first applied to the field of computer vision (Erhan et al., 2010). Whereas NLP requires an enormous annotated corpus, the original pretraining models for NLP were used to learn features, and the current pretrained models are positioned very differently (Collobert & Weston, 2008). Pretraining models and fine-tuning paradigms only became popular in the field of
Table 1. Summary of the main relation extraction approaches in the last three decades

<table>
<thead>
<tr>
<th>Paradigm</th>
<th>Category</th>
<th>Models</th>
<th>Year</th>
<th>Evaluation results (f-score)</th>
<th>Features and description</th>
<th>Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relation extraction approach based on the traditional rule-based parsing templates</td>
<td>Rule-based templates</td>
<td>Artificially constructed templates (Califf &amp; Mooney, 1997)</td>
<td>1997</td>
<td>-</td>
<td>Trigger word, rule-based templates</td>
<td>Online texts</td>
</tr>
<tr>
<td></td>
<td>Rule-based templates</td>
<td>Artificially constructed templates (Fundel et al., 2006)</td>
<td>2006</td>
<td>-</td>
<td>Word-level, sentence-level, and position features to construct rule templates</td>
<td>MUC</td>
</tr>
<tr>
<td>Statistical machine learning relation extraction approach</td>
<td>Supervised</td>
<td>Maximum entropy models (Kambhatla, 2004)</td>
<td>2004</td>
<td>52.8</td>
<td>Feature functions, using word-level features, sentence-level features, and external semantic relations, etc.</td>
<td>ACE02</td>
</tr>
<tr>
<td></td>
<td>Supervised</td>
<td>SVM (Zhou et al., 2005)</td>
<td>2005</td>
<td>68.0</td>
<td>Lexical, semantic, syntactic; external lexicon</td>
<td>ACE03</td>
</tr>
<tr>
<td></td>
<td>Supervised</td>
<td>SVM (Mooney, &amp; Bunescu, 2005)</td>
<td>2005</td>
<td>47.7</td>
<td>String kernel function</td>
<td>ACE</td>
</tr>
<tr>
<td></td>
<td>Supervised</td>
<td>SVM (Culotta, &amp; Sorensen, 2004)</td>
<td>2005</td>
<td>38.0</td>
<td>Syntax tree kernel function</td>
<td>ACE</td>
</tr>
<tr>
<td></td>
<td>Supervised</td>
<td>SVM (Zhang et al., 2008)</td>
<td>2008</td>
<td>72.1</td>
<td>Tree kernel overlay WorNet</td>
<td>ACE03</td>
</tr>
<tr>
<td></td>
<td>Distant-supervised</td>
<td>Automatic template extraction (Mintz et al., 2009)</td>
<td>2009</td>
<td>67.6 (precision)</td>
<td>Knowledge migration with knowledge base</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Distant-supervised</td>
<td>Multiple instances, logistic regression (Riedel et al., 2010)</td>
<td>2010</td>
<td>87.0 (precision)</td>
<td>Knowledge migration with knowledge base</td>
<td>New York Times corpus (NYT)</td>
</tr>
<tr>
<td></td>
<td>Unsupervised</td>
<td>K-means (Chen et al., 2005)</td>
<td>2005</td>
<td>41.3–50.6 (precision)</td>
<td>Word level information and surrounding word features</td>
<td>ACE</td>
</tr>
<tr>
<td>Neural network model-based relation extraction approach</td>
<td>Supervised</td>
<td>RNN (recurrent neural network; Socher et al., 2012)</td>
<td>2012</td>
<td>82.4</td>
<td>Syntactic features and lexical features</td>
<td>SemEval 2010</td>
</tr>
<tr>
<td></td>
<td>Supervised</td>
<td>CNN (Zeng et al., 2014)</td>
<td>2014</td>
<td>82.7</td>
<td>Syntactic features, lexical features, and WordNet</td>
<td>SemEval 2010</td>
</tr>
<tr>
<td></td>
<td>Supervised</td>
<td>Piece-wise CNN (Zeng et al., 2015)</td>
<td>2015</td>
<td>78.3</td>
<td>Add location information</td>
<td>NYT corpus</td>
</tr>
<tr>
<td></td>
<td>Supervised</td>
<td>CNN (Wang et al., 2016)</td>
<td>2016</td>
<td>88.0</td>
<td>Multi-layer attention mechanism</td>
<td>SemEval-2010 Task 8</td>
</tr>
<tr>
<td></td>
<td>Supervised</td>
<td>Bi-LSTM (Zhou et al., 2016)</td>
<td>2016</td>
<td>84.0</td>
<td>Attention mechanism, word-level, and sentence-level features</td>
<td>SemEval-2010; Task 8</td>
</tr>
<tr>
<td></td>
<td>Supervised</td>
<td>CNN (Lin et al., 2016)</td>
<td>2016</td>
<td>76.2</td>
<td>Multi-instance learning, attention mechanism</td>
<td>NYT corpus</td>
</tr>
<tr>
<td></td>
<td>Supervised</td>
<td>CNN (Ru et al., 2018)</td>
<td>2018</td>
<td>86.2 (precision)</td>
<td>Reinforcement learning, semantic jaccard</td>
<td>Wikipedia</td>
</tr>
<tr>
<td></td>
<td>Joint modeling</td>
<td>LSTMs (Zheng et al., 2017)</td>
<td>2017</td>
<td>49.5</td>
<td>New entity relation tags</td>
<td>NYT corpus</td>
</tr>
<tr>
<td></td>
<td>Joint modeling</td>
<td>Bi-LSTM-RNN (Li et al., 2017)</td>
<td>2017</td>
<td>71.4</td>
<td>The first layer handles named entity recognition; the second level deals with relation extraction</td>
<td>ADE task</td>
</tr>
<tr>
<td></td>
<td>Graphical neural network model</td>
<td>GCN (Zhang, Qi, &amp; Manning., 2018)</td>
<td>2018</td>
<td>84.4</td>
<td>Dependent syntactic analysis, graph convolution</td>
<td>SemEval 2010 Task 8</td>
</tr>
<tr>
<td></td>
<td>Capsule neural network model</td>
<td>Bi-LSTM + Primary Capsule (Zhang, Deng, et al., 2018)</td>
<td>2018</td>
<td>78.8</td>
<td>Sentence features, dynamic routing</td>
<td>NYT corpus</td>
</tr>
<tr>
<td></td>
<td>Neural snowball model</td>
<td>CNN, BERT (Gao et al., 2020)</td>
<td>2020</td>
<td>82.15 (precision)</td>
<td>Bootstrap</td>
<td>Wikipedia</td>
</tr>
</tbody>
</table>

continued on following page
natural language processing with the advent of ELMo (Erhan et al., 2010) and ULMFit (Howard & Ruder, 2018), both of which are based on the LSTM neural network model architecture. However, the difference is that the former uses a three-layer LSTM network, and the latter uses a bidirectional LSTM. Both models support fine-tuning, and the size of the model parameters and the size of the pretrained corpus is much larger than the previous models, and both models have to a large degree, improved their performance in several tasks.

In 2018, the Transformer language model pretraining structure was proposed (Vaswani et al., 2017), and Transformer’s multi-headed self-attention mechanism allows each word to selectively attend to all words preceding it, thus allowing the model to efficiently capture long sentence information without considering expensive circular computation like LSTM. This enables the model to capture more valid information, which is very effective in the field of NLP. The current mainstream language models, including GPT, BERT, BART (Lewis et al., 2020), and T5 (Raffel et al., 2020), all belong to the Transformer architecture. Figure 2 gives the evolution of large language models in the field of natural language processing.

Figure 2. Development of pretrained language models
Prompt Learning Paradigm Features

In the pretraining model with the fine-tuning paradigm, first, the pretrained model is based on a mask masking task that allows the machine to learn to predict the probability of this word and the fine-tuning phase is to add a specific task. In the entity relation extraction task, its fine-tuning task is to go for predicting the probability of the relation classification. At this point, there is a natural gap between the pretraining and fine-tuning tasks because of the different task content, and, essentially, pretraining and prediction do not perform the same kind of tasks (Jacob et al., 2019).

This gap is naturally compensated under the paradigm of pretrained models combined with prompt learning. Prompt learning allows adding some additional information to the original sentence so as to turn the final prediction task into the same type of task as the training task. In entity relation extraction, for example, the original sentence is “Keyi College of Zhejiang Sci-tech University is located in Shaoxing City,” and then the additional information “the relation between the two entities is (mask)” is added so that the prompt template is used for packaging the relation extraction task into the same form as the pretrained task, i.e., the mask masking task. In this way, for the model, the prediction is a repetition of the task of having trained tens of thousands of words, i.e., giving the probability distribution of words at mask positions over the entire word list. We compare these words and types, such as string matching, similarity calculation, etc., to achieve the mapping of relation classifications.

Previously we added additional contextual information in prompt learning that belongs to the prompt template while mapping answers to specific tasks is the verbalizer. In this way, we make up the gap between prediction and training tasks, both doing mask masking tasks, both sequence input and sequence output tasks. Because these tasks are trained with a large corpus, it makes the prompt learning in such a way that the pre-processing model can achieve better performance. On the other hand, we use the same set of data for pretraining, just different prompt template settings for different tasks and different Verbalizer settings, with no need to consider the difference between a large number of NLP tasks.

Selection of Pretrained Models Under the Prompt Learning Paradigm

In terms of model selection, since the model size significantly impacts the model’s effectiveness, we prioritize the mainstream big pretrained models for relation extraction as a typical natural language processing task. First, we consider an auto-regressive generative model like GPT, which sets prompts based on tasks, outputs tokens through sentence endings, and uses these tokens to map relation classifications, and this auto-regressive model is naturally suitable for the prompt learning paradigm. Auto-regressive models like the MetaAI implementation of OPT (Zhang et al., 2022) and GPT-3 with 175 billion parameters both use auto-regressive models (Brown et al., 2020) combined with prompt templates to generate the desired text. This type of model has become the current mainstream model.

The second choice is the BERT model, which performs well in comprehension tasks and classification tasks, modeling RoBERTa, BERT-Large, etc. With this class of models, one can combine the prompt templates to put the mask position at the end of the sentence and let it be wrapped into sequence generation to do relation extraction in the output string. There are also original sequence-to-sequence type models, such as BART and T5. With this type of model, one can add prompt to the input sequence to map it to a specific relation classification. All the current mainstream high-performing big models can be converted into pretrained models suitable for prompt learning, either directly or with a small amount of packaging.

Prompt Template Construction (PTC)

There are four main approaches to prompt template construction, namely the artificial language template construction approach, the structured template approach, the automatically generated prompt
template approach, and the vector representation template approach. Here, we introduce each of them for the relation extraction task.

**Artificial Language Template Construction Approach**

Artificial language templates based on human experiences are the most widely used, and experiments show that human-constructed templates do not work too badly. In the relation extraction task, the constructor has some a priori knowledge related to entity relations and can focus on some features of the task itself. In addition, the big model training is implicitly done with the human experience because a large amount of human-written text is used (Ding et al., 2021). As an example, if the predicted sentence is “Beijing is one of the biggest cities in China,” we use the template “[original text], (entity1) is an/a (mask) of (entity2),” which varies with the input corpus. Then the model will output a token of mask location, such as city, location, etc. The output is then mapped to the location relation classification based on the mapping rules between the output and the type. In this way, it can align the training task with the prediction task, allowing the model to recall the huge amount of data from the pretraining when making predictions.

In addition, some scholars have added rules and knowledge bases to improve prediction accuracy based on the above manual templates (Ye et al., 2022; Petroni et al., 2019). To illustrate the above location relation extraction case, since we can do lexical and named entity recognition on the input text, we can add some rules related to relation classifications, and in the location relation, we can specify the rule that entity1 and entity2 must be place names to improve the correctness of relation extraction. Experiments show that adding additional rule-based constraints to the prompt learning relation extraction task can further improve the results (Han et al., 2022). Many other scholars have applied artificial language template construction approaches to tasks such as question and answer, translation (Schick, T., & Schütze, H., 2021a), and text classification (Schick, T., & Schütze, H., 2021b).

**Structured Template Construction Approach**

To make the input text more intelligible to the model, the prompt template can add structured information (Zhong et al., 2022). In the prompt template for relation extraction, the structure adds the text of the alternative relation classifications, such as “[original text]. Select relation classification in (located in| parent of...),” and this allows the pretrained model to understand the range of the output text. In addition, the example of relation extraction can be added to the prompt template to assist the model in understanding the task goal.

In practice, we can also use multi-prompt template decision-making and then select the output (Yuan et al., 2021). We can make relation predictions for multiple templates separately, and we weigh the predictions to calculate a score and finally select the relation type with the highest score. These prompt engineering methods can reduce labor costs, aggregate multiple weak prompts into one strong prompt, and make the output of relation extraction results more stable. In a recent study, some scholars have further utilized the nearest neighbor approach to assign weights among multiple prompts and achieved good results (Chen, Li, et al., 2022).

**Automatic Generation Prompt Template Approach**

Many human-defined prompt templates were mentioned earlier, but finding an excellent prompt template sometimes requires experience and luck (Richard et al., 2021; Jiang, Xu, et al., 2020). Therefore, an approach to automatically construct prompt templates was proposed (Shin et al., 2020). Firstly, a set of words are selected to generate a collection as a bag of words for trigger words, and then, the original corpus text, the trigger words extracted from the randomly selected bag of words, and the relation classifications are made to splice. This approach may result in the final optimized prompt being placed in the original sentence without semantic meaning, but it may work well. This suggests there is possibly an optimal prompt for the relation extraction task, optimal for the machine.
but completely different from the human-defined prompt considered, such that the optimal prompt fits and generalizes particularly well. However, this idea is not yet fully proven.

In addition, researchers at Princeton University are doing research on the automatic generation of prompt templates. This approach uses the corpus to train and the result text of task prediction to the T5 model to generate prompts and then splicing the text prompt for task prediction, where the one with the highest prediction accuracy is selected as the final selected prompt template (Gao et al., 2021). This way of generating a prompt template is more linguistically appropriate. Recent studies have used the automatic generation of prompt templates for low-resource relation extraction and produced good results (Chia et al., 2022; Liu, Lin, et al., 2022).

**Vector Representation Prompt Template Approach**

The vector representation prompt template is a novel idea. This approach no longer constructs prompt templates from the word perspective, and the authors directly characterize prompts in a vectorized way. In the entity relation extraction task, some old relation vectors are used to initialize it first, the prompt vector is spliced with the corpus and relation classification, and then as the model is trained, the prompt vector changes to a vector where token never appears. Although without comprehensible semantics, this prompt vector has shown good results in experiments with P-tuning (Liu, Ji, et al., 2022), a training process that freezes the model parameters and only goes on to optimize the value of the prompt vector. Some scholars used a similar approach and achieved good results in text generation (Guo et al., 2021; He et al., 2023).

**Summary**

First of all, in the few-shot scenario, the prompt template can produce good results (Le & Rush, 2021). The artificial language prompt template and structured template approaches are the most widely used in practical application scenarios. This model is mostly used in the existing entity relation extraction work. In terms of automatically generated prompt templates, this paper aims to find a set of prompts that are intelligible. We believe the corpus of the pretrained model is artificially produced. The best prompt, if in existence, should be in line with the natural language properties when the experiments are sufficient. The discussion of this issue is intriguing, and we will study it in depth in future work. In addition, although the application of prompt learning works well, the existing research has not been able to explain why it produces the effect and the intrinsic principles, which is also a direction worth studying.

**Verbalizer**

In the relation extraction task, the verbalizer mainly aims to map the generated sequences to specific relation classifications, and some scholars have conducted answer engineering studies in classification tasks (Yin et al., 2019), relation extraction (Chen et al., 2022), named entity recognition (Cui et al., 2021), generation tasks (Alec Radford et al., 2019), and multiple choices tasks (Khashabi et al., 2020). Verbalizers fall into two categories: artificial answer engineering and automatic answer generation engineering.

**Artificial Answer Engineering**

Artificial answer engineering requires some expertise in constructing mapping relations, and in the relation extraction task, some relation tag words are set for the entity relation classification (Jiang, Anastasopoulos, et al., 2020). Then the bag of words of this classification is continuously extended by synonym substitution. The larger this bag of words is, the higher the error tolerance rate is, and some noise is also introduced. Besides, the knowledge base and synonym list can be used to expand the bag of words further. Some relation tag words containing multiple words can be split and the core words can be extracted. Artificial answer engineering essentially takes the output generated by the model and matches the relation distribution according to the human-set mapping to find the attributed
relation classification. Some scholars have introduced an external knowledge base to expand the result set in terms of artificial answer engineering (Hu et al., 2022). For example, if the input text is: “What’s the relation between NER and RE?”, a template “a (mask) question” can be added in front of it, and the system will predict an answer word. We need to match this answer word with the bag of words of our corresponding relation classification. The authors use the knowledge base, synonym lists, and even language models to generate results to expand this bag of words. The implementation proves that this approach works well for verbalizers by removing the noise inside.

**Automatic Answer Generation Engineering**

Similarly, we can generate answer mappings through embedded vectors in the same way (Hambardzumyan et al., 2021). That is, virtual answer mapping vectors are trained. This approach stitches a large number of training samples, virtual answer mapping vectors, and entity relation results. The training is done by the model to allow an embedded vector of virtual answer mappings to be trained for that relation classification. In the prediction phase, the final relation classification attribution is obtained by making a distance calculation between the embedded vector of the model-generated text and the embedding vector of that type to get the closest distance to the relation type (Cui et al., 2022). In the latest study, the authors further optimize the answer mapping in automated answer engineering by using the rich semantic and prior knowledge in relation labels and achieve better results (Chen et al., 2022).

**Training Strategies Under the Prompt Learning Paradigm**

As the mainstream pre-models have become very large (McCloskey & Cohen, 1989) in 2023, some new training strategies have been proposed in the current environment, and three mainstream training strategies are described below.

**Prompting With Optimizing All the Parameters**

In the prompt learning paradigm, full parameter fine-tuning is still possible (Matthew et al., 2019). Only the new data is accessed through a prompt to the input text and the loss is calculated to optimize the model parameters by comparing the resulting output with the target label. This paradigm requires expensive hardware costs and is very difficult to apply in the current large model environment.

**Prompting With a Frozen Pretrained Model**

In contrast to the former, a more suitable approach is to use a prompt training approach with frozen parameters. First, we add some soft prompts to the input corpus of the model, such as tokens of words with no real meaning, and then feed them into the pretraining model for training. In this case, we freeze the parameters of the pretrained model and train only the embedded vectors of these soft prompts. Thus, the goal of driving large models with small parameters is achieved (Lester et al., 2021). Experimental results show that the results are poor on small and medium-sized models, but the results are close to the full parameter tuning on models with more than 10 billion parameters. This can be explained by the powerful fitting ability of big models (Gu et al., 2022). The approach has been equally successful for relation extraction in the medical field (Yeh et al., 2022).

**Summary**

The new paradigm of prompt learning has brought new ideas for relation extraction. The current research shows that the use of prompt template construction, verbalization, and new training strategies are expected to bring new breakthroughs to the task of entity relation extraction. Since this area is still in the exploration stage, persistent efforts are required to explore how to choose the optimal prompt template, find a more suitable answer mapping method for relation extraction, and find a better training strategy. In this paper, the authors believe that this area is one of the most promising research directions at present.
CHALLENGES AND POSSIBLE RESEARCH DIRECTIONS

Although new approaches to the field of relation extraction (RE) are constantly being developed and assessment scores are constantly being improved, the relation extraction task is still far from being truly solved. This is because the real-world scenarios are often rich in relation classifications, sparse in the annotated corpus, and are commonly characterized by multiple relations and cross-sentences. There are still many challenges to fully automating the extraction of relations using machines. We summarize these challenges into three areas.

The Challenge of the Lack of Labeled Data

What is really lacking is large-scale, high-quality training data (Zhang et al., 2019) since annotating a high-quality corpus is costly and the entity relations in real scenarios are numerous and complex. Therefore, the existing corpus size and relation classifications are still limited, to some extent, preventing the existing models from achieving better prediction performance. The distant-supervised and bootstrap approach is a promising way to supplement the corpus by automatically labeling the data with the acquired relations. Multi-instance, reinforcement learning, and rule qualification approaches can reduce the noise problem and semantic drift problems. For now, however, the noise problem and semantic drift problem are yet to be fully addressed, which is a major barrier to the practical application of distant-supervised and bootstrap approaches.

Nowadays, pretrained models and prompt learning paradigms introduce new ideas to solve the problem of lack of data. First, we can explore some ways to allow models to generate more training corpus using big models and prompt templates with high-quality annotated examples. In addition, prompt learning can be used to design efficient filters and semantic drift detectors for the noisy data, allowing distant-supervised learning and bootstrap methods to break through the bottlenecks and play a higher practical application value.

The Challenge of Model Learning Efficiency

Learning efficiency was not a prominent issue originally but rather a long-tail effect in relation extraction tasks in which a large number of relations suffer from corpus sparsity (Han et al., 2018; Koch et al., 2015), forcing the need for more efficient learning efficiency in relation extraction methods. In the past, some scholars have specifically invested in the study of relation extraction in low-resource scenarios (Vinyals et al., 2016; Gao et al., 2019), in which low-resource relation extraction can be handled by means of relation knowledge migration and meta-learning (Ravi & Larochelle, 2016; Finn et al., 2017; Mishra et al., 2018). This type of approach can quickly learn low-resource relation classifications and achieve relation extraction goals. However, the current low-resource measurement data are relatively simple, and some approaches with good measurement scores fall beyond our expectations expected in practical applications. In addition, the extraction performance drops sharply when the number of relation classifications increases. There is probably over-matching in the model learning using low resources rather than actually learning the key features. Therefore, commercial scenarios still require enormous manual verification work.

The introduction of the prompt learning paradigm has allowed us to approach the problem in a new way. On the one hand, we can make learning more efficient by using more efficient prompting with the frozen pretrained approach. On the other hand, based on the experimental experience of (Wei et al., 2021), this area merits further in-depth investigation by combing fine-tuning with prompted data to investigate the predictive ability of the model for unknown relations through the generalization ability obtained by the model in the training of rich relation types with the study of the prompt template.

The Challenge of Handling Complex Relations and Open-Domain Scenarios

Entity relations tend to be more complex in the real world, where there are multiple relations in some sentences and relations nested within each other; other relations may exist across sentences
and chapters. There are also many sentences that do not contain any relations (Marcheggiani & Titov, 2016; Vashishth et al., 2018). In addition, real-world relations are so varied that exhausting all relations artificially poses a challenge. In this paper, we consider the open domain scenario as the ultimate application scenario for relation extraction. Currently, some scholars have studied cross-sentence relation extraction and multiple relation extraction and achieved good results. However, practical applications are still difficult. In fact, the current measurement dataset is much simpler than the real scenario. The complex relation extraction tasks require a stronger understanding and strong knowledge base of models (Cui et al., 2018; Mausam, 2016).

Therefore, we need to continuously enhance the model capability, including the model parameter size continuing to increase and train more data on the one hand. We foresee that the model is likely to grow exponentially in the coming year. On the other hand, based on the new paradigm of learning from pretrained models and prompts, we explore more effective prompt templates and answer mapping methods to improve relation extraction in complex scenarios. We combine the existing tools such as the knowledge base and search engine to further enhance the accuracy of relation extraction. We believe that this is one of the important directions for future relation extraction.

CONCLUSION

This paper systematically introduces relation extraction tasks against the backdrop of big data, from paradigms based on artificial language rule-based templates, and traditional statistical machine learning-based paradigms to neural network model-based paradigms. Among them, the deep model-based relation extraction methods are highlighted, and finally, new research ideas on relation extraction are proposed for the paradigms based on big models. Since big models are too large to be trained easily, prompt learning has become a promising research direction for RE, our work is, therefore, a systematic introduction to this paradigm for RE and compared with traditional paradigms.

The entity relation extraction task was proposed as early as 30 years ago. Historically, methods based on artificial language rules and traditional statistical machine learning played an important role in relation extraction tasks. The paradigm of artificial linguistic rules was the first to be applied, and this class of methods is still used in high-precision classification-less scenarios for it is easy to operate. The traditional machine learning paradigms, on the other hand, have dominated the relation extraction tasks until the introduction of deep learning models. Among them, models such as the maximum entropy model and SVM have been constantly refreshing the best results of relation extraction in supervised approaches. Feature selection has been the most important work in that phase. In that period, distant-supervised methods, bootstrap’s approach, and unsupervised methods have also been introduced into the relation extraction work one after another to open up ideas for relation expansion and unsupervised extraction.

After 2012, neural network models changed the research direction in most areas of artificial intelligence, including the field of relation extraction. The relation extraction approaches based on neural network models have significantly improved the performance of relation extraction, and RNN (recurrent neural network), CNN, RNN (recurrent neural network), and LSTM have been applied to relation extraction tasks with good results that time. At that time, some scholars began to explore more complex relation extraction tasks and relation extraction approaches with low resources. Since 2018, big models based on neural networks have changed the course of relation extraction research once again. Various big model-based approaches have again significantly improved their performance, generating ever better results. Despite the significant progress in relation extraction task, problems like the lack of labeled data, low resource learning efficiency, and the problem of extracting complex and open relations remain major obstacles to relation extraction task.

Our observation of the current situation in the industry reveals the following two trends: (1) the larger the model size, the better the results; and (2) since the paradigm of large models with fine-tuning can no longer be adapted to the current increasing capacities of models, we introduce a
relation extraction paradigm based on big models and prompt learning. The latest research results under this paradigm are also sorted out and analyzed. Several valuable future research directions are proposed to address the above challenges in relation extraction. We hope readers can gain a systematic understanding of the development of relation extraction through this paper, which can help them conduct research on the latest approaches and promising future directions of relation extraction in the new paradigm of prompt learning. We hope to contribute, along with other scholars, to research on the existing problems in the field of relation extraction.

DATA AVAILABILITY

The data used to support the findings of this study are included within the article.

CONFLICTS OF INTEREST

The author declares that there is no conflict of interest regarding the publication of this paper.

FUNDING STATEMENT

This research was funded by the Talent Fund of Keyi College of Zhejiang Sci-Tech University and by the NetEase Frontier Technology Research Project. This work was also supported by NetEase Group, Zhejiang Sci-tech University, and Zhejiang University City College.
REFERENCES


Mausam, M. (2016, July). Open information extraction systems and downstream applications. *Proceedings of IJCAI,* (pp. 4074–4077 ACL.


