Adaptive Prompt Learning-Based Few-Shot Sentiment Analysis

Pengfei Zhang · Tingting Chai · Yongdong Xu

Accepted: 13 March 2023 / Published online: 28 March 2023
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract
In the field of natural language processing, sentiment analysis via deep learning has a excellent performance by using large labeled datasets. Meanwhile, labeled data are insufficient in many sentiment analysis tasks, and obtaining these data is time-consuming and laborious. Prompt learning devotes to resolving the data deficiency by reformulating downstream tasks with the help of prompt. The model performance of this method depends on the quality of the prompt. This paper proposes an adaptive prompting (AP) construction strategy using seq2seq-attention structure to acquire the semantic information of the input sequence. Our method of dynamically constructing adaptive prompts can not only improve the quality of prompt, but also can effectively generalize to other fields by constructing a pre-trained prompt with existing public labeled data. The experimental results on FewCLUE datasets demonstrate that the proposed method AP can effectively construct appropriate adaptive prompt regardless of the quality of hand-crafted prompt and outperform the state-of-the-art baselines.

Keywords  Natural language processing · Sentiment analysis · Adaptive prompt learning · Seq2seq-attention

1 Introduction

Nowadays, deep learning (DL) has been widely used in natural language processing, computer vision, information security, Internet of things and other fields to solve all kinds of problems. In the field of computer vision, Bai et al. [1] proposed a new model TransFusion based on the transformer, which is used to solve the LiDAR camera fusion problem in 3D object detection. In the field of information security, Lu et al. [2] applied the Deep Belief Network to the Cyber-Attack Detection task. In the field of Internet of things, Chen et al. [3] applied...
extreme learning machine (ELM), Elman neural network (ENN) and long short term memory neural network (LSTM) to the wind speed prediction task. The application of deep learning in these areas has achieved good results.

In the field of natural language processing, deep learning methods have also achieved good performance in many tasks. In the task of sentiment analysis, supervised deep learning methods have achieved excellent performance. At the same time, the model effectiveness of sentiment analysis depends on large-scale high-quality labeled data which is insufficient. In addition, manually marking large-scale data is time-consuming and laborious. It is difficult to obtain desirable labeled data to train the model. In order to address data acquisition issue, there are two kinds of methods. The first kind of methods are to use large-scale unsupervised data and a small amount of supervised data for learning, such as semi-supervised learning method. Yu et al. [4] proposed a semi-supervised machine learning method for tweet sentiment classification. Jiang et al. [5] proposed an emotion space model (ESM) based on emoticon, which uses emoticons to construct word vector from unlabeled data.

The second kind of methods are to learn general features from large-scale data and then adjust them on specific tasks, such as fine-tuning pre-trained model and prompt learning. Fine-tuning is to adjust the parameters of the pre-training model according to different downstream tasks, so that the model can be applied to downstream tasks. Araci et al. [6] proposed a FinBERT language model based on BERT, which can handle the sentiment classification task in the financial field by fine-tuning the BERT model. With the increasing scale of the pre-trained language model, the hardware and data requirements of fine tuning are also rising. Therefore, researchers have explored a smaller but more efficient method which is called prompt learning. The prompt learning method is to transform the input and output of downstream tasks into an acceptable form of the pre-trained model, so that the model can be used for downstream tasks.

Prompt learning solves the problem of lacking data by using prompt to adjust downstream tasks. This method relies on the quality of the prompt. There are two main methods to construct prompt: hand-crafted prompt method and automated prompt method. Hand-crafted prompt method is generally based on human natural language knowledge. Petroni et al. [7] designed manually the cloze templates for the knowledge probing task in the LAMA dataset. Brown et al. [8] designed prefix templates for question answering, translation and probing tasks for common sense reasoning. In general, hand-crafted prompt method is intuitive, but it requires researchers to have rich experience and linguistic knowledge. In some sentiment analysis tasks, it is even more difficult to construct a prompt manually than to label some samples manually. In order to solve the problems of hand-crafted prompt method, many researchers began to explore how to automatically learn the appropriate prompt. Liu et al. [9] proposed P-tuning, which abandons the assumption that the generated prompt must be natural language and learns the best prompt by using unused tokens in the vocabulary of model. All in all, these automated learning prompt methods improve the efficiency of constructing prompt, but they do not make full use of the semantic information of input text in the process of generating prompt, which limits the model performance to a certain extent.

In this work, an adaptive prompt method (AP) by introducing seq2seq-attention structure is proposed to achieve state-of-the-art performance in low resource tasks. It can automatically generate prompt without artificial design. Moreover, the introduced seq2seq-attention structure can make full use of the semantic information of the input text to generate matching prompt. In addition, its ability of prompt construction can be further improved by pre-training on the existing labeled datasets in other fields.
2 Related Work

2.1 Sentiment Analysis

Sentiment analysis originates from the analysis of subjectivity in sentences [10]. Due to the emergence of a large number of network resources, the research of sentiment analysis has become an active field since 2000 [11]. There are three kinds of sentiment analysis methods: sentiment dictionary based sentiment analysis, machine learning based sentiment analysis and deep learning based sentiment analysis. Early sentiment analysis mainly focused on building a sentiment dictionary for text classification. It was constructed manually by summarizing words containing sentiment tendencies, and labeling the sentiment polarity and intensity of these words to varying degrees. Therefore, it is necessary to build a high-quality sentiment dictionary [12]. SentiWordNet [13] is the earliest sentiment dictionary. Cai et al. [14] solved the polysemy problem of sentiment words by constructing a sentiment dictionary in a specific domain. The same sentiment word in the sentiment dictionary may express different meanings at different time, in different languages or in different domains, so this method does not perform well in cross domain or cross language. In addition, the sentiment dictionary method often ignores the contextual semantic relationship, so it has great limitations.

Machine learning based sentiment analysis mainly relies on NLP researchers or engineers to use their domain knowledge to define and extract significant features from the original data, such as n-gram features, and then use traditional machine learning classifiers such as support vector machine, naive Bayes and maximum entropy for supervised learning [15]. Li et al. [16] proposed a multi-label maximum entropy (MME) model for sentiment classification over short text. Li [17] builds a model with the prior knowledge of the categorization information in order to extract meaningful features from the unstructured texts by using TF-IDF, short for term frequency-inverse document frequency. The machine learning based sentiment analysis mainly lies in the extraction of sentiment features and the combination selection of classifiers. The combination selection of different classifiers has a certain impact on the results of sentiment analysis. Such methods often can not make full use of the context information when conducting sentiment analysis on the text content, which has a negative impact on their classification accuracy.

In recent years, with the development of deep learning theory, neural network has gradually matured in the field of sentiment analysis. Deep neural network can effectively capture the high-level semantic information of text without complex feature engineering, and the expression ability index of the model is times better than that of the shallow model. Among them, convolutional neural network and recurrent neural network are the most widely used [18]. Li [19] proposed the BLSTM and CNN Stacking Architecture (BCSA) to enhance the ability to recognition emotions. Besides, Chen [20] proposed HUSN which is a novel sentiment classification algorithm that utilizes user’s review habits to enhance hierarchical neural networks. The model proposed by Sadr [21] uses the tree structure of recurrent neural network to replace the pooling layer in the convolutional network in order to capture long-term dependencies and reduce the loss of local information. Compared with the other methods, deep learning based sentiment analysis methods have significant advantages in text feature learning. They can actively learn features, and actively retain the information of words in the text. Therefore, they can better extract the semantic information of corresponding words and expressions, and achieve better results of the sentiment classification.
2.2 Pre-trained Model

The purpose of pre-trained language models (PLMs) is to use a large number of texts that have appeared in people’s life to train the model, so that the model can learn the probability distribution of each word in these texts, so as to model the model that conforms to these text distributions.

Traditional PLMs technology aims to learn word embedding. These models usually have low computational efficiency because downstream tasks no longer need to use them, such as skip gram [22] and glove [23]. Although these pre-trained word vectors can capture the semantic meaning of words, they are context independent and can not capture the high-level concepts of text, such as grammar and semantics. Elmo [24] proposed a context sensitive text representation method, which constructs the text representation through the deep bidirectional language model, which effectively solves the problem of polysemy. In 2018, Devlin et al. [25] proposed BERT (bidirectional encoder representations from transformers) pre-trained language model. The model trains massive corpus through bidirectional transformer encoder and uses masked language model (MLM) to generate in-depth bidirectional language representation. After pre-training, you only need to add an additional output layer for fine-tuning to achieve the performance of state of the art in a variety of downstream tasks. In this process, there is no need to make task specific structural modifications to Bert.

BERT has opened a new era, and a large number of pre-trained language models have emerged since then. Roberta [26] has longer training time, larger batch, longer sequence and more data on the basis of retaining the original BERT architecture. At the same time, it deletes the prediction of the next sentence and uses dynamic masking. Albert [27] solves the problems of high memory consumption and slow training speed of Bert. ERNIE [28] introduced the knowledge masking strategy, including entity level masking and phrase level masking, to replace the random masking in Bert. In addition, Bert also can be used in other languages. Farahani [29] proposed a monolingual Bert for the Persian language which is lighter than the original multilingual Bert model. Bert-wwm [30] is for Chinese which not only masks continuous entity words and phrases, but also masks all continuous words that can form Chinese words.

2.3 Prompt Method

With the increasing volume of pre-trained language model, the hardware requirements, data requirements and actual cost of fine-tune are also rising. In addition, the rich and diverse downstream tasks also make the design of pre-training and fine-tuning stage cumbersome and complex. Therefore, researchers hope to explore smaller, lighter, more universal and efficient methods. Prompt method is an attempt in this direction which includes hand-crafted prompt method and automated prompt. Prompt method is a technology that adds additional text to the input segment in order to better use the knowledge of the pre-trained language model. In the aspect of hand-crafted prompt methods, Schick et al. [31] designed pattern exploiting training (PET), which is a semi-supervised training task. The input example is redefined as the phrase of cloze to help the language model understand the given task. By selecting suitable prompt, hand-crafted prompt method can use an unsupervised pre-trained language model to solve various downstream tasks, and its performance in few-shot tasks is much better than the supervised and semi-supervised methods.

In the aspect of automated prompt methods, Jiang et al. [32] proposed a mining based method, which can automatically find templates of a given set of training input and output.
This method finds the intermediate word or dependency path between input and output in a large text corpus containing input and output strings, and uses the frequently occurring intermediate word or dependency path as a template. Davison et al. [33] designed an input (head relation tail) template using LM by studying the tasks related to the knowledge base. Gao et al. [34] proposed a method to generate prompt automatically which is called LM-BFF. It uses a generative model T5 to automatically generate prompts for cloze questions, including template generation and label word generation. LM-BFF incorporates demonstrations during training to help the model better distinguish samples. Liu et al. [9] proposed a method called P-tuning, which abandons the conventional requirement that “the template is composed of natural language” and uses the token never seen in the model to form the template. It transforms the construction of the template into a continuous parameter optimization problem and realizes the automatic construction of the template. Wang et al. [35] proposed EFL method which reformulates the fine-tuning task into textual entailment and designs fine-grained textual descriptions for the labels.

In the hand-crafted prompt methods, the accuracy of the model depends very much on the quality of the constructed template, and the effects of different templates may vary greatly. For some tasks, it is not so easy to discover an optimal prompt manually. In the automated prompt methods, the prompt is constructed by the model and does not rely on manual work. But both of those methods can not use the semantic information of input texts in prompt construction process.

To solve the above problems, this paper proposes a template construction method by introducing seq2seq-attention structure which can dynamically generate matching template vectors and makes full use of the original text information. At the same time, with the idea of pre-trained model, we design a template construction strategy based on pre-training, which can make full use of the public sentiment analysis datasets of high resources fields to learn general prompts and apply them to other field of low resources.

The main contributions of this paper include three aspects:

1. We propose an adaptive prompt method by introducing seq2seq-attention structure. This method has the advantages of both hand-crafted prompt method and automated prompt method. And it can make full use of semantic information of input text, which is the main innovation point of this paper.
2. The experimental results on the FewCLUE dataset show that the proposed method is effective in the sentiment analysis few-shot task.
3. We proposed to pre-train the adaptive prompt module in high resources tasks, and migrate to or fine-tune in low resource tasks which can effectively play a significant effect in low resource tasks.

3 Methodology

In this section, we propose an adaptive prompting method based on seq2seq-attention (AP) and introduce its implementation. We introduce seq2seq-attention structure to generate adaptive templates from input, and then use the pre-training model to realize sentiment analysis.
3.1 Adaptive Prompt Learning

Our work is based on adaptive prompt learning model, which improves traditional hand-crafted prompt learning method (HPL). The HPL model includes input layer, hand-crafted prompt, pre-trained language model and output layer.

Given a pre-trained model \( M \), vocabulary \( V \), a input sequence \( X \) of length \( n; \{x_1, x_2...x_n \} \), verbalizer \( W; \{w_1,w_2...w_l \} \), a hand-crafted prompt \( P; \{p_1,p_2...p_i,[MASK],p_{i+1}...p_m \} \), where value of \([MASK]\) comes form \( W \). Firstly, The input sequence \( X \) and manual prompt \( P \) form a template \( t \). Then, the template \( t \) will be mapped into \( e(t) \): \( \{e(p_1), e(p_2)...e(p_i), e([MASK]), e(p_{i+1})...e(p_m), e(x_1), e(x_2)...e(x_n)\} \) by pre-trained model embedding layer \( e \) (where each token \( p_i \) will be mapped into \( e(p_i) \)). After that, the pre-trained model \( M \) is used to calculate the probability values of \([MASK]\) in \( e([MASK]) \) to select the best word in \( W \) which has the maximum probability. For example, for the sentiment calculation of “The weather is very good”, hand-crafted prompt “it is [MASK]”, verbalizer {“good”, “bad”}, and then the traditional prompt model will construct the template “It is [MASK], The weather is very good. “Finally, \( M \) will return the predictive value. The hand-crafted prompt method is as shown in Fig. 1a.

While the strategy of traditional hand-crafted prompt learning is intuitive and work well in some sentiment tasks, there are also two issues with this approach. (1) It is hard for human to discover optimal prompts in all tasks. (2) Even if in the same task, the optimal prompts of different input sequences are different. Usually, it is hard for algorithms to dynamically find “best” prompt for a special input sequence in task. We consider automatic adaptive prompt design to solve the defects of manual prompt design. As shown in Fig. 1b, we use adaptive prompt layer to generate prompt instead of manual craft prompt. In order to strengthen the relationship between the prompt and the input sequence \( X \), we consider to use the text information of the input sequence \( X \) to automatically generate the corresponding prompt, that is, generate an adaptive prompt. In traditional automated methods, there is no direct correlation between prompt and input \( X \). We trained the adaptive prompt using the context information of input \( X \). We think the model structure of the adaptive prompt layer should meet the following two conditions: (1) the output sequence (adaptive prompt embedding sequence) and input sequence (word embedding vector of input \( X \)) of the adaptive prompt layer can be unequal in length; (2) the adaptive prompt layer can obtain the semantic information of the input sequence \( X \) and generate adaptive prompts according to its semantics. Based on the above conditions, we can choose the seq2seq structure as the basic structure of the adaptive prompt layer. In the seq2seq structure, all input information will eventually be encoded into a content vector \( C \). \( C \) will store all sequence information. Once the input sequence length
Fig. 2 Seq2seq-attention structure, as the adaptive prompt layer of our model. Input is embedded words of $X$, and output is prompt embedding vector sequence. The model is divided into two parts: encoder and decoder. The encoder and decoder in this paper both use GRU structure. The word embedding sequence of $X$ and the randomized parameter value $h_0$ are input into the encoder to obtain the hidden state sequence $\{h_1, h_2..., h_n\}$. The randomized parameter value $y_0$ and the last hidden state of the encoder hidden layer $h_n$ are the input of the decoder, and the first hidden state of the decoder is $h'_1$. We use the soft attention mechanism to calculate attention scores. For the decoder hidden state $h'_t$ and the encoder hidden state $h_i$, we calculate the attention score $w_{ti}$ according to the formula $w_{ti} = \text{softmax}(h_i h'_t)$. And then we calculate the content vector $C_t$ according to the formula $C_t = \sum_i w_{ti} h_i$. Content vector $C_t$ and the last decoder hidden state $h'_t$ are concatenated into $y_t$. Finally, $y_1, y_2..., y_s$ constitutes an adaptive prompt embedding sequence. It should be noted that $\{y_1, y_2..., y_s\}$ here corresponds to $\{h_1, h_2..., h_s\}$ in Fig. 1b.

is too long, the proportion of key information will be reduced, resulting in the loss of key information. Therefore, in order to ensure the integrity of key information and enhance the relationship between prompt sequence and input sequence $X$, we introduce attention mechanism to better capture the details of $X$. Therefore, we use seq2seq-attention structure as adaptive prompt layer to generate adaptive prompt as shown in Fig. 2.

In general, we use seq2seq-attention structure as the adaptive prompt layer to generate an adaptive prompt sequence. And each vector dimension of the sequence is consistent with the output vector dimension of the embedding layer of the pre-trained model. Meanwhile, the adaptive prompt’s embedding vectors are continuous which enables us to find a better continuous prompt beyond the original vocabulary could express [9]. In addition, adaptive prompt layer can capture text information of input $X$ by attention structure (the yellow part in the Fig. 2), which can make the generated prompt fitter with the input text.

3.2 Hybrid Prompt Learning

The automated learning prompt method has various advantages in most tasks, such as wide application range, strong generalization ability and stable and balanced performance. However, it may fall into local optimal solution in many cases. The hand-crafted prompt method has excellent performance in some cases, but it is unstable and requires the participation of experienced experts. In order to combine the advantages of the two methods, we design the
hybrid prompt model (HAP) composed of a hand-crafted part and automated part as shown in Fig. 3.

The hybrid prompt model combines the word vector generated by hand-crafted prompt and the prompt layer. Through the hybrid prompt embedding layer, the template can be represented a triple \( <X, P, h> = \{e(p_1), e(p_2), \ldots, e([MASK]), e(p_{i+1}), \ldots, e(p_m), h_1, h_2, \ldots, h_s, e(x_1), e(x_2), \ldots, e(x_n)\} \), where \( P \) is the hand-crafted prompt, \( h \) is the adaptive prompt embedding vector sequence, \( X \) is the input text sequence. And the final model prediction result set \( y = \{p(i), i \in W\} \) is calculated by pre-trained model.

Both adaptive prompt and hand-crafted prompt effect the final result \( y \). The result showed that (4. 4. 1 for details), the model can learn to adjust the weights of \( P \) and \( h \) to generate better output results. When the hand-crafted prompt is "good", the hybrid prompt model will automatically increase the weight of the hand-crafted prompt. Otherwise, it will increase the weight of the adaptive prompts. Therefore, theoretically, the model has the advantages of both hand-crafted prompt method and automatic prompt method. When a "good" hand-crafted prompt can be found, the adaptive prompt part generates auxiliary prompt to further improve the model effect. Even if the "good" hand-crafted prompt cannot be found, the adaptive prompt part also can generate excellent prompt. Thus, the combination of two parts can effectively improve the quality of generated prompt.

4 Experiments and Results

4.1 Database

We evaluate AP method mainly on the EPRSTMT (E-commerce Product Review Dataset for Sentiment Analysis) task of public dataset FewCLUE. EPRSTMT is labelled as Positive or Negative and collected by ICIP Lab of Beijing Normal University. The datasets used in migration experiment include social media public sentiment dataset (more than 100,000 data with emotional labels on Sina Weibo, and about 50,000 positive and negative comments respectively), hotel comment data (more than 7,000 hotel review data, more than 5000 positive and 2,000 negative comments), user comments data by a takeout platform (4,000 positive and 8,000 negative user comments collected by a takeout platform), online shopping data which have 7 categories (books, fruits, shampoo, water heater, milk, clothes and hotels) and
Table 1  Accuracy of different methods under different hand-crafted prompts in Chinese datasets

| Prompt     | Zero-shot | HPL  | HAP  |
|------------|-----------|------|------|
| __开心(happy) | 75.7%     | 83.8%| 84.1%|
| __高兴(glad)  | 70.2%     | 79.3%| 83.9%|
| __好(good)   | 64.9%     | 80.3%| 82.8%|
| __行(OK)     | 51.8%     | 78%  | 82.3%|

more than 60,000 comments in total, with about 30,000 positive and negative comments respectively. The dataset in English field includes 7,000 movie data (about 3,500 positive and 3,500 negative data respectively).

4.2 Hyper-Parameters Setting

In order to fully obtain all the information of the sentence, the maximum length of a sentence is set to twice the median length of the sentences in the dataset. In the experiment, the pre-trained model adopts Roberta-wwm-ext model. The batch size value is set to 5 and output length of adaptive prompt layer is set to 2 in Chinese and 4 in English. The model adopts Adam optimizer and adopts different learning rates for different optimization methods.

4.3 Optimization Strategy

The model consists of two trainable parts, one is the pre-trained model parameters, the other is the seq2seq-attention parameters. Based on this, our optimization methods can be divided into two categories: one is to fine-tune all parameters (prompt+LM tuning). In this setting, prompt-relevant parameters can be fine-tuned together with the all or some of the parameters of the pre-trained models [36]. And the other is to fine-tune only the seq2seq-attention part (fixed-LM prompt tuning). In the scenario where additional prompt-relevant parameters are introduced besides parameters of the pre-trained model, fixed-LM prompt tuning updates only the prompts’ parameters using the supervision signal obtained from the downstream training samples, while keeping the entire pre-trained LM unchanged[36].

4.4 Results Analysis

4.4.1 Prompt+LM Tuning

In this experiment, we use the method of full model parameter adjustment to test the dataset from the FewCLUE in Chinese (32 data in training set and 600 data in test set) and the dataset of movie field in English (32 data in training set and 600 data in test set). In this method, the fine-tuning of pre-trained model plays a leading role in the overall model training, and the seq2seq part plays an auxiliary role. That is, the hand-crafted prompt plays a major role, while the automated template is a supplement and enhancement to the hand-crafted prompt. In this case, the learning rate is 1e-5. The results are shown in Tables 1 and 2.

Among them, the zero-shot method only uses prompt to construct the template, and then predicts through the pre-trained model without fine-tuning the parameters of the pre-trained model. We use the results of zero-shot to judge the quality of hand-crafted prompt. In this case, it can be seen that the quality of hand-crafted prompt has a greater impact on HPL.
Table 2  Accuracy of different methods under different hand-crafted prompts in English datasets

| Prompt                  | Zero-Shot | HPL  | HAP  |
|------------------------|-----------|------|------|
| It was __              | 62.3%     | 72.7%| 77.8%|
| Just __!               | 60.2%     | 75.7%| 78.3%|
| It makes me feel __ that | 72%       | 74.5%| 80%  |

method, but less impact on HAP method. HAP method can also have a higher accuracy when the hand-crafted prompt is not good. On the other hand, when the hand-crafted prompt is good, HAP model can also play a better role than HPL model. Thus, the model can learn to adjust the weights of hand-crafted prompt and adaptive prompt to return better results.

4.4.2 Fixed-LM Prompt Tuning

In order to further test the ability of the seq2seq-attention part of the AP method, we canceled the hand-crafted prompt in this part of the experiment, only used seq2seq-attention to generate the prompt, and frozen the parameters of the pre-trained language model. Therefore, the goal of seq2seq-attention structure is to learn the embedding representation of adaptive prompt in pre-trained language model, which makes \( h \) behaves like the sequence of real text through the embedding layer.

We designed experiments on large-scale data (microblog data) and small sample data (FewCLUE data). In the experiments on large-scale data, the accuracy of the model is more than 92%, while in the small sample data, the accuracy of the model is only about 65%.

For the good performance of the experiment in the case of large data and the poor performance of small sample data, we consider that in the case of large data, due to the sufficient samples, we can learn the adaptive prompt and its embedding representation in pre-trained model through seq2seq-attention structure. In the case of insufficient samples, seq2seq-attention structure is difficult to learn two parts at the same time, therefore, resulting in over fitting.

Embedding representation can only be learned under large-scale data. And this experiment has shown that the model has the ability to learn adaptive prompt with sufficient samples. Therefore, in order to verify that the seq2seq-attention structure can learn the generality of embedding representation of pre-trained model and adaptive prompt, we have done migration experiment.

4.4.3 Migration Experiment

Sentiment analysis includes different fields, such as catering, e-commerce and film. Although there are some differences between these fields, they are generally a classification of sentiment. The reason why the past models can not be used directly across fields is that the words and language structures used for emotional expression in different fields are very different, resulting in different parameters of word vector layer and full connection layer. Therefore, a good automated prompt construction structure should be able to learn the adaptive prompt in the general field and perform well in the unknown field.

In this experiment, we set a mixed data experiment. In this experiment, the training set mixes the sentiment analysis datasets of 7 categories (books, fruits, shampoo, water heater, milk, clothes and hotels) of online shopping field, microblog field, takeout field, hotel field.
Fig. 4 The accuracy of datasets in different fields (left), and the accuracy of test data (right)

Table 3 Main results of different learning mechanisms on FewCLUE. Values with * are retrieved from Xu [37] et al.

| Method        | FineTuning | PET     | LM-BFF  | P-tuning | EFL     | AP   |
|---------------|------------|---------|---------|----------|---------|------|
| Accuracy      | 65.4%*     | 86.7%*  | 85.6%*  | 88.3%*   | 84.9%*  | 88.7%|

Table 4 Comparison of experimental results

|                | PET                | AP                | Pre-AP          |
|----------------|--------------------|-------------------|-----------------|
| 67.8% (avg)    | 69.8% (avg)        | 78.2%             |

We use the dataset of e-commerce field (FewCLUE) as the test set. In this case, the learning rate is 2e-6. The result of each epoch is as shown in Fig. 4.

It can be seen that the model can learn to construct general adaptive prompt in the mixed fields, and can play a good effect in other fields, with an better accuracy of 88.7%, much higher than the results in 4.1.1 and other methods on FewCLUE datasets. The experimental results are shown in Table 3.

It proves that the model can learn the adaptive prompt in the general fields and perform well in the unknown field.

4.4.4 Pre-train Experiment

Pre-training is an application of transfer learning. It uses almost unlimited text to learn the context sensitive representation of each member of the input sentence. It implicitly learns the general grammatical and semantic knowledge, and migrates the knowledge learned from the open domain to the downstream tasks to improve the low resource tasks. We hope to learn the general expression of adaptive prompt from the large-scale sentiment analysis dataset through pre-training, so as to better solve the sentiment classification in the case of small sample data. In this experiment, we pre-train the model in the sentiment analysis datasets of microblog field, takeout field, hotel field and online shopping field. And fine-tuning in the movie field (32 data in training set and 600 data in test set), in this case, the learning rate is 5e-6. The experimental results are shown in Table 4.

The results show that the accuracy of the Pre-AP model is much higher than that of PET and AP models, which means the pre-training method is feasible in AP model.
4.5 Discussion

This paper demonstrates the effectiveness of our proposed method through four experiments. In the first experiment, we think the reason why the HAP method is better than the HPL method is that the HAP method can learn to adjust the weight of the hand-crafted prompt and adaptive prompt to return better hybrid prompt. When the hand-crafted prompt is not good, our model tends to give higher weight to adaptive prompt, which makes the hybrid prompt better than the pure hand-crafted prompt. Otherwise, our model tends to give the hand-crafted prompt higher weight. Since there is no perfect hand-crafted prompt, the adaptive prompt will also supplement the hand-crafted prompt, so the hybrid prompt generated by our model will still have a better effect than the pure hand-crafted prompt. Since the HPL method only depends on the hand-crafted prompt, the quality of the hand-crafted prompt has a great impact on the model effect. However, the HAP method depends on both hand-crafted prompt and adaptive prompt and the weight of dependence on them will be adjusted according to the result, so the quality of hand-crafted prompt has less impact on the HAP method.

In the second experiment, the seq2seq-attention structure performs well in large-scale data but poorly in small sample data. We think the reason is that when the samples are sufficient, the adaptive prompt and its embedding representation in the pre-trained model can be learned through the seq2seq-attention structure. While in the case of insufficient samples, the seq2seq-attention structure is difficult to learn two parts at the same time, resulting in poor experimental results.

In the third experiment, we proved that the AP model perform better than other methods on FewCLUE dataset. We think the reason is that our model can make full use of the semantic information of input text to generate adaptive prompt, while other models ignore the association between prompt and the semantics of input text. This experiment also proves that the AP model can learn to construct general adaptive prompt in the mixed fields and can perform well in other fields. The parameter quantity of our model is mainly composed of three parts: the parameters of seq2seq-attention, the parameters of Robertsa-wwm-ext model and the parameters of the MLP layer. The parameters of Robertsa-wwm-ext model accounts for the majority, about 125 M.

In the fourth experiment, the experimental results show that the accuracy of the Pre-AP model is much higher than that of PET model and HAP model, which shows the feasibility of the pre-training method in the AP model.

5 Conclusion

This paper introduces the method of sentiment analysis, analyzes the shortcomings of prompt learning. The accuracy of the HPL method depends very much on the quality of the hand-crafted prompt. When the quality of the hand-crafted prompt is poor, the accuracy of the HPL method will also be very low. In addition, the previous automatic prompt methods often ignored the relationship between prompt and input text semantic information when generating prompt, and this relationship plays a great role in improving the accuracy of the model. Then we proposed the adaptive prompt model. The advantages of the model can be summarized as follows:
1. The hand-crafted prompt and automated prompt are combined in the model. This improvement can solve the problem that the model accuracy of HPL method is very dependent on the quality of hand-crafted prompt. Regardless of the quality of the hand-crafted prompt, our model can achieve good results.

2. Seq2seq-attention structure is introduced to make full use of context information to generate adaptive prompt. This is the main innovation of the paper. We use the seq2seq-attention structure to capture the semantic information of the input text, so as to generate a better prompt.

3. The proposed model AP learns to construct a general adaptive prompt by using the sentiment analysis datasets in the fields which have sufficient samples.

4. Pre-trained prompt method in the field of sentiment analysis is proposed.

Future research work will be carried out in-depth research from the following aspects to provide directions for further improving the performance of the model: (1) find a better parameter fine-tuning method based on pre-trained prompt; (2) extend this method to other fields of natural language processing, such as text classification, machine reading comprehension.

References

1. Bai X, Hu Z, Zhu X, et al (2022) Transfusion: robust lidar-camera fusion for 3d object detection with transformers. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 1090–1099
2. Lu KD, Zeng GQ, Luo X et al (2021) Evolutionary deep belief network for cyber-attack detection in industrial automation and control system. IEEE Trans Industr Inf 17(11):7618–7627
3. Chen MR, Zeng GQ, Lu KD et al (2019) A two-layer nonlinear combination method for short-term wind speed prediction based on ELM, ENN, and LSTM[J]. IEEE Internet Things J 6(4):6997–7010
4. Yu Z, Wong R K, Chi C H, et al (2015) A semi-supervised learning approach for microblog sentiment classification. In: 2015 IEEE international conference on smart city/SocialCom/SustainCom (SmartCity). IEEE, pp 339–344
5. Jiang F, Liu YQ, Luan HB et al (2015) Microblog sentiment analysis with emoticon space model. J Comput Sci Technol 30(5):1120–1129
6. Araci D F, Genc Z (2019) Financial sentiment analysis with pre-trained language models. arXiv preprint arXiv:1908.10063
7. Petroni F, Rocktäschel T, Lewis P, et al (2019) Language models as knowledge bases?. arXiv preprint arXiv:1909.01066
8. Brown T, Mann B, Ryder N et al (2020) Language models are few-shot learners. Adv Neural Inf Process Syst 33:1877–1901
9. Liu X, Zheng Y, Du Z, et al (2021) GPT understands, too. arXiv preprint arXiv:2103.10385
10. Rebecca B, Wiebe J, Thomas O H (1999) Development and use of a gold standard data set for subjectivity classifications. In: Proceedings of the 37th annual meeting of the association for computational linguistics (ACL–99). Association for Computational Linguistics
11. Tong R M (2001) An operational system for detecting and tracking opinions in on-line discussion. In: Proceedings of the ACM SIGIR workshop on operational text classification
12. Xue Y, Li Q, Jin L, et al (2014) Detecting adolescent psychological pressures from micro-blog. In International conference on health information science. Springer, Cham
13. Baccianella S, Esuli A, Sebastiani F (2010) SentiWordNet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. Lang Resources Eval. European Language Resources Association (ELRA)
14. Cai Y, Yang K, Huang D et al (2019) A hybrid model for opinion mining based on domain sentiment dictionary. Int J Mach Learn Cybern 10(8):2131–2142
15. Lafferty J, Mccallum A, Pereira F C N (2002) Conditional random fields: probabilistic models for segmenting and labeling sequence data. In: Proceedings of ICML
16. Li J, Rao Y, Jin F et al (2016) Multi-label maximum entropy model for social emotion classification over short text. Neurocomputing 210(19):247–256
17. Li G, Lin Z, Wang H, et al (2020) A discriminative approach to sentiment classification. Neural Process Lett 51(2)
18. Lecun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436
19. Li D, Sun L, Xu X, et al (2021) BLSTM and CNN stacking architecture for speech emotion recognition. Neural Process Lett (1)
20. Chen J, Yu J, Zhao S, et al (2021) User’s review habits enhanced hierarchical neural network for document-level sentiment classification. Neural Process Lett (2)
21. Sadr H, Pedram MM, Teshnehlab M (2019) A robust sentiment analysis method based on sequential combination of convolutional and recursive neural networks. Neural Process Lett 50(6):2745–2761
22. Le Q, Mikolov T (2014) Distributed representations of sentences and documents. In: International conference on machine learning. PMLR, pp 1188–1196
23. Pennington J, Socher R, Manning C (2014) Glove: global vectors for word representation. In: Conference on empirical methods in natural language processing
24. Peters ME, Neumann M, Iyyer M, Gardner M, Clark C, Lee K, Zettlemoyer L (2018) Deep contextualized word representations. arXiv preprint arXiv:1802.05365
25. Devlin J, Chang M W, Lee K, et al (2018) Bert: pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805
26. Liu Y, Ott M, Goyal N, et al (2019) Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692
27. Lan Z, Chen M, Goodman S, et al (2019) Albert: a lite bert for self-supervised learning of language representations. arXiv preprint arXiv:1909.11942
28. Zhang Z, Han X, Liu Z, et al (2019) ERNIE: enhanced language representation with informative entities. In: Proceedings of the 57th annual meeting of the association for computational linguistics
29. Gharachorloo M, Farahani M, Farahani M et al (2021) Parsbert: transformer-based model for persian language understanding. Neural Process Lett 53(6):3831–3847
30. Cui Y, Che W, Liu T et al (2021) Pre-training with whole word masking for Chinese bert[J]. IEEE/ACM Trans Audio Speech Lang Process 29:3504–3514
31. Schick T, Schütze H (2020) Exploiting cloze questions for few shot text classification and natural language inference. arXiv preprint arXiv:2001.07676
32. Jiang Z, Xu FF, Araki J et al (2020) How can we know what language models know? Trans Assoc Comput Linguist 8:423–438
33. Davison J, Feldman J, Rush A M (2019) Commonsense knowledge mining from pretrained models. 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), pp. 1173–1178
34. Gao T, Fisch A, Chen D (2020) Making pre-trained language models better few-shot learners. arXiv preprint arXiv:2012.15723
35. Wang S, Fang H, Khabsa M, et al (2021) Entailment as few-shot learner. arXiv preprint arXiv:2104.14690
36. Liu P, Yuan W, Fu J, et al (2021) Pre-train, prompt, and predict: a systematic survey of prompting methods in natural language processing. arXiv preprint arXiv:2107.13586
37. Xu L, Lu X, Yuan C, et al (2021) Fewclue: a Chinese few-shot learning evaluation benchmark. arXiv preprint arXiv:2107.07498

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.