

Part-of-Speech Tagging for Traditional Chinese Medicine Clinical Pulse Feeling and Palpation Information Based on Hidden Markov Model Prompt-Tuning

Linfeng Dai

School of Software Engineering, Chengdu University of Information Technology, Chengdu 610225, China

Abstract: This paper proposes a Traditional Chinese Medicine Clinical Pulse Feeling and Palpation Information part-of-speech tagging method based on Hidden Markov Model Prompt-Tuning, aimed at addressing the challenges of scarce annotated samples and the issue of words with multiple parts of speech in Traditional Chinese Medicine Clinical Pulse Feeling and Palpation Information part-of-speech tagging tasks. By designing specific prompt templates, this method transforms the downstream task into a cloze task for the pretrained language model during its pretraining phase, thereby bridging the gap between pretraining and fine-tuning objectives and enhancing the knowledge representation ability of the pretrained language model for downstream tasks. This approach also integrates the part-of-speech information of adjacent words into the prompt templates and uses the Hidden Markov Model to model the dependencies between adjacent words' parts of speech, effectively improving the overall performance of the model in the Traditional Chinese Medicine Clinical Pulse Feeling and Palpation Information part-of-speech tagging task, particularly in identifying words with multiple parts of speech. Furthermore, empirical evidence demonstrates the outstanding performance of the proposed method in this part-of-speech tagging task, achieving an F1 score of up to 85.52%.

Keywords: Prompt Tuning, Part-of-Speech Tagging, Hidden Markov Model, Pre-training Language Model, Traditional Chinese Medicine Clinical Pulse Feeling and Palpation Information.

1. Introduction

In Traditional Chinese Medicine (TCM), doctors synthesize and analyze information from the four diagnostic methods (inspection, listening and smelling, inquiring, pulse feeling and palpation) to identify the cause, location, nature, and development trend of a patient's illness, and to determine the appropriate treatment method[1]. Among these, pulse feeling and palpation is an indispensable diagnostic approach, comprising both pulse diagnosis and palpation diagnosis. This method involves the physician using their hands or fingers to touch, feel, press, or tap certain parts of the patient's body to diagnose diseases and understand the condition[2].

Part-of-speech tagging is a fundamental task in the field of natural language processing (NLP), with the goal of assigning appropriate part-of-speech tags to each word based on context[3]. Part-of-speech tagging of Traditional Chinese Medicine Clinical Pulse Feeling and Palpation Information can aid downstream tasks in more accurately identifying and extracting key elements from pulse diagnosis information, thereby enhancing the quality and efficiency of data processing in intelligent medical systems.

However, unlike part-of-speech tagging tasks in general domains, the current task of part-of-speech tagging for Traditional Chinese Medicine Clinical Pulse Feeling and Palpation Information faces two major challenges:

1. High Specialization of Traditional Chinese Medicine Clinical Pulse Feeling and Palpation Information: Collecting and organizing a large amount of pulse diagnosis information with part-of-speech tags requires significant human, material, and time resources. Thus, the lack of annotated data becomes one of the challenges hindering the development of part-of-speech tagging tasks for Traditional Chinese Medicine

Clinical Pulse Feeling and Palpation Information.

2. High Frequency of Words with Multiple Parts of Speech: Words with multiple parts of speech are quite common in Traditional Chinese Medicine Clinical Pulse Feeling and Palpation Information, accounting for about 14.93% of the entire dataset. An example of such a phenomenon is shown in Figure 1. In this example, the character '起' appears in two different pieces of pulse diagnosis information but has different parts of speech in different contexts: in one instance, it is tagged as a modal verb (VV), while in another, it is used as a predicative adjective (VA). In the current task, the presence of words with multiple parts of speech adds complexity to part-of-speech tagging, posing another challenge to the development of part-of-speech tagging tasks for Traditional Chinese Medicine Clinical Pulse Feeling and Palpation Information[4].

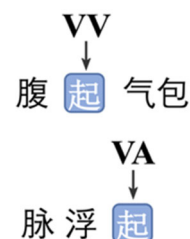


Figure 1. Examples of the Phenomenon of Words with Multiple Parts of Speech in Traditional Chinese Medicine Clinical Pulse Feeling and Palpation Information

In summary, in the current task, the scarcity of annotated data and the prevalence of words with multiple parts of speech collectively constitute key obstacles to the research progress in part-of-speech tagging for Traditional Chinese Medicine Clinical Pulse Feeling and Palpation Information.

with “[CLS]”. Similarly, y_{i-1} represents the part of speech of the preceding word, and for the first word x_1 in the sentence, “[CLS]” is also used to represent its part of speech. Using the prompt template constructor, each word x_i in the input sentence $X = \{x_1, x_2, \dots, x_n\}$ is transformed into a corresponding prompt sentence. Therefore, for a sentence containing n words, this function will output n prompt sentences. Each prompt sentence includes five new input components: the original sentence X , the current word x_i , the word preceding the current word x_{i-1} (replaced with “[CLS]” if x_i is the first word), the part of speech of the word preceding x_i y_{i-1} (also replaced with “[CLS]” if x_i is the first word), and the masked token [MASK] in the prompt template.

In this paper, information about the previous word and its part of speech is incorporated into the prompt template, prompting the pretrained language model to consider the part of speech of the preceding word when predicting the part of speech for the current word. Additionally, the inclusion of the original sentence emphasizes the importance of the global context of the text to be annotated for predicting the part of speech of the current word.

Table 1 shows the prompt sentences constructed using the prompt template when $X = \{\text{“脉”, “浮”, “起”}\}$. During the construction of the prompt sentences, since the part-of-speech tagging has not yet begun, the part of speech of the preceding word is unknown. Therefore, this paper uses “[P]” as a placeholder to temporarily represent the part of speech, which will be replaced once the part of speech for that word is predicted.

Table 1. Sentence X and prompt sentences generated by prompt template T

输入的样例句子
脉 浮 起
通过提示模板 T 生成的提示句子
在句子“脉浮起”中，词语“脉”的前文如果是由 “[CLS]” 词性的词语 “[CLS]” 来修饰，那么词语“肘”的词性是 “[MASK]”
在句子“脉浮起”中，词语“浮”的前文如果是由 “[P]” 词性的词语 “脉” 来修饰，那么词语“浮”的词性是 “[MASK]”
在句子“脉浮起”中，词语“起”的前文如果是由 “[P]” 词性的词语 “浮” 来修饰，那么词语“起”的词性是 “[MASK]”

2.3. Pre-trained Language Model Layer

In the Prompt-Tuning paradigm of this paper, the task of the Pre-trained Language Model Layer is to leverage its extensive pretrained knowledge and representation ability to more comprehensively capture the contextual information of the prompt sentences and to fill in the masked token. In the model, the default output of the masked language model is the probability of inserting each word in the vocabulary into the “[MASK]” position. However, for part-of-speech tagging tasks, the model needs the probability of inserting each tag from the dataset's tag set into the “[MASK]” position. Therefore, this paper first adds the dataset's tags into the vocabulary of the masked language model. Then, based on the positions of the part-of-speech tags in the vocabulary, the output of the masked language model is precisely trimmed to obtain the probability vectors for each part-of-speech tag. These vectors serve as the final output of the Pretrained Language Model and are inputted into the Hidden Markov Model Layer.

2.4. Hidden Markov Model Layer

The Hidden Markov Model (HMM) is a statistical model

composed of observation probabilities (emission probabilities) and state transition probabilities. In this paper's task, the state transition probabilities are learnable parameters, representing the probability of transitioning from one part-of-speech state to another. The observation probabilities correspond to the pretrained language model's predicted probabilities for the '[MASK]' position.

In part-of-speech tagging tasks, the HMM can model the dependencies between the parts of speech of adjacent words. It considers the dependency of the current word on the previous word, aligning with the chain prediction of the current part-of-speech tagging. This approach more accurately captures the correlations between the parts of speech of adjacent words.

There are mainly two decoding methods in the prediction part of the HMM: approximation algorithms and the Viterbi algorithm. Approximation algorithms are simpler to compute but cannot guarantee that the predicted sequence of part-of-speech tags is the most likely overall state sequence. Additionally, approximation predictions may encounter situations where adjacent transitions have a probability of zero. The Viterbi algorithm, on the other hand, uses dynamic programming to find the path with the highest probability and is more accurate than approximation algorithms, but it also has higher time complexity. In the current task of part-of-speech tagging for Traditional Chinese Medicine Clinical Pulse Feeling and Palpation Information, since the length of the pulse diagnosis text is relatively short, the time complexity of using the Viterbi algorithm is within an acceptable range. Therefore, this paper utilizes the Viterbi algorithm for decoding.

The process of part-of-speech tagging a text using the Hidden Markov Model (HMM) and the Viterbi algorithm is manifested as a single traversal from the first word to the last word in the sequence. A step in this traversal involves part-of-speech tagging for a single word in the text to be annotated. After completing the current step, the part-of-speech tag of the current word is added to the prompt sentence of the next word and fed back into the Pretrained Language Model Layer for cloze completion of the masked token. The output of this cloze task serves as the data for the next iteration of the Viterbi algorithm. The part-of-speech tagging task for the current text is considered complete only after the Viterbi algorithm has traversed all the words in the text. The specific algorithmic process and its pseudocode are presented as Algorithm 1.

Algorithm 1 hint tuning algorithm based on hidden Markov

输入：提示句子具体数量 N ，一组提示句子 prompts, bert 计算函数 get_score, 模型配置 Config, 状态转移矩阵 transition。
 输出：viterbi 过程矩阵 trellis, 预测结果 labels, bert 函数输出的 loss。

- 初始化 viterbi 过程矩阵 $trellis \leftarrow [N, Config.class_nums]$
- 初始化存储 viterbi 路径的路径矩阵 $pre_index \leftarrow [N, Config.class_nums]$
- 初始化保存总损失的变量 $total_loss$
- for** each $index \in [1 \dots N]$ **do**
- $score, loss \leftarrow get_score(prompts[index])$
- $total_loss \leftarrow total_loss + loss$
- if** $index = 0$ **then**
- $trellis[0] \leftarrow score$
- else**
- $trellis_cur \leftarrow [0 \dots Config.class_nums]$
- for** each $score_idx \in [1 \dots Config.class_nums]$ **do**
- $temp \leftarrow []$
- for** each $trellis_idx \in [0 \dots Config.class_nums]$ **do**
- $item \leftarrow trellis[index-1][trellis_idx]$
- $+ transition[trellis_idx][score_idx] + core[score_idx]$
- $temp \leftarrow append(item)$
- end for**
- $max_value \leftarrow max(temp)$
- $max_index \leftarrow argmax(temp)$

```

19.     pre_index[index][score_idx] ← pre_index[index-
1][max_index] + [score_idx]
20.     trellis_cur ← append(max_value)
21.     end for
22.     trellis[index] ← trellis_cur
23.     end if
24.     cur_pred_label_id ← max(trellis[index]) + 1
25.     if index != N - 1 then
26.         next_prompt ← prompts["input_ids"][index+1]
27.         for each i,x ∈ next_prompt do
28.             if x = Config.PLB then
29.                 next_prompt[i] = cur_pred_label_id
30.             end if
31.         end for
32.     end if
33. end for
34. labels ← pre_index[N-1][argmax(trellis[N-1])]
35. return trellis,labels,total_loss

```

3. Experiment

In the test dataset, the method proposed in this paper was compared with CRF and BiLSTM+CRF models under the BERT[11], MC-BERT[12], and BART[13] pretrained language models. Notably, MC-BERT is a model obtained through fine-tuning on an unannotated domain-specific Traditional Chinese Medicine clinical records dataset in prior work. Subsequent sections of this chapter will detail the datasets used in the experiments, experimental settings, evaluation metrics, and the results of the experiments.

3.1. Dataset

All the experimental data in this paper are derived from the corpus of Traditional Chinese Medicine Clinical Pulse Feeling and Palpation Information extraction, constructed in previous work[4]. In that prior study, detailed part-of-speech tagging was already performed on Traditional Chinese Medicine Clinical Pulse Feeling and Palpation Information, with each word being assigned a corresponding part-of-speech tag. Given the high quality and specificity of these data, this paper chose to use this dataset directly for the part-of-speech tagging task experiments. In this dataset, after deduplication required for the part-of-speech tagging task, 944 tagged data entries were obtained. The dataset was divided into training, validation, and test sets in a 6:2:2 ratio. In the training set sampling portion, training samples with 5, 10, 15, 20, and 25 data entries were randomly selected to train the model. Additionally, to ensure objectivity in model evaluation, each set of training samples was sampled five times and the average of the training results was taken.

3.2. Metrics

This paper employs typical metrics used in classification tasks—Precision, Recall, and F1-Score—to measure the quality of the method proposed. The specific formulas for these metrics are as follows:

$$\text{Precision} = \frac{TP}{TP+FP} \quad (1)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (2)$$

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

Where TP denotes the number of samples predicted to be true and actually true, FN denotes the number of samples predicted to be inverse but actually true, and FP denotes the number of samples predicted to be true and actually inverse.

3.3. Experimental Settings

In the comparative experiments and the proposed method, the learning rate is set to 2e-5, batch size to 64, maximum sentence length to 256, and the number of training epochs is 100. The optimizer used during model training is AdamW. To verify the effectiveness of the method proposed in this paper, the following models were chosen for comparative experiments: BERT+CRF, BERT+BiLSTM+CRF, MC-BERT+CRF, MC-BERT+BiLSTM+CRF, BART+CRF, and BART+BiLSTM+CRF.

3.4. Experimental Results and Analysis

3.4.1. Comparison with Standardization Methods

Table 2 lists the average precision (Pavg), average recall (Ravg), and average F1 score (F1avg) for various models trained on the Traditional Chinese Medicine Clinical Pulse Feeling and Palpation Information part-of-speech tagging dataset. These models include the proposed method with different Pre-trained Language Models (PLMs) and three other fine-tuning-based part-of-speech tagging models (PLM+BiLSTM, PLM+BiLSTM, and PLM+CRF). All values are represented as percentages.

The experimental results show that all baseline models demonstrate a trend of performance improvement as the number of samples increases (from 5-shot to 25-shot), indicating that the models can learn more from a larger number of samples, thereby improving the accuracy of part-of-speech tagging. In different experimental groups corresponding to different pretrained language models, the model proposed in this paper consistently outperforms the other fine-tuning-based baseline models. However, it is observed that the best-performing model varies under different pretrained language models, which indicates that the part-of-speech tagging model proposed in this paper has stronger adaptability and generalizability compared to traditional fine-tuning-based part-of-speech tagging models.

From Figure 3, it can be seen that the MC-BERT model ranks second in terms of performance. However, it is noteworthy that as the number of training samples increases, the performance of MC-BERT significantly improves, maintaining an average F1 score gap of within 2% compared to the BART model. This result indicates that although MC-BERT is slightly less effective than BART in scenarios with fewer data samples, it can demonstrate performance comparable to BART when the sample size is moderately increased, highlighting MC-BERT's suitability and effectiveness for part-of-speech tagging tasks in Traditional Chinese Medicine Clinical Pulse Feeling and Palpation Information.

Furthermore, it is important to note the significant difference in parameter size between the MC-BERT and BART models: MC-BERT has a parameter size of 390.286MB, while BART is 1436.215MB. This difference in parameter size is particularly crucial for applications in low-resource environments. Even though MC-BERT is slightly inferior to BART in certain performance metrics, its relatively smaller parameter size means that it can serve as a more efficient alternative in environments with limited memory. Using MC-BERT as a pretrained language model provides a viable way to balance resource constraints with performance needs, within an acceptable range of model performance loss. Therefore, MC-BERT offers a solution that is acceptable in terms of loss and requires relatively fewer resources.

Regarding the recognition of words with multiple parts of

speech, it has been statistically found that the model achieves a recognition rate of 38.81% for such words, which is 8.55% higher than the baseline models. This demonstrates the effectiveness of the proposed method in addressing the issue of words with multiple parts of speech.

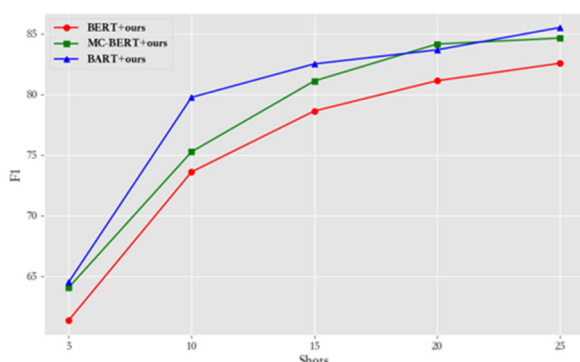


Figure 3. Curve of the change of F1 values of different pre-training language models with training scale

To assess the role of the Hidden Markov Model (HMM) in the proposed model, an ablation study was conducted. In this study, the HMM component was removed, and the results were compared with those of the complete model. As shown in Table 3, the ablation study revealed that the removal of the HMM led to a decrease in the average F1 score of the proposed model across different pretrained language models and training set sizes. This decrease in performance indicates the crucial role that the HMM plays in the part-of-speech tagging task for Traditional Chinese Medicine Clinical Pulse Feeling and Palpation Information.

This study presents a part-of-speech tagging method based on Hidden Markov Model Prompt-Tuning, aimed at addressing part-of-speech tagging tasks under low-resource conditions. This method, using the concept of sequence labeling combined with dynamic contextual information and transforming the part-of-speech tagging task into a cloze task, effectively bridges the gap between pretrained models and downstream tasks. It solves the issue of words with multiple parts of speech in Traditional Chinese Medicine Clinical Pulse Feeling and Palpation Information. Additionally, the Prompt-Tuning approach enhances the model's performance in low-resource scenarios. The experimental results demonstrate that the proposed method significantly outperforms control models under various training set size conditions, offering a new perspective and solution for low-resource part-of-speech tagging tasks. Future research will focus on exploring soft prompts to more comprehensively evaluate the performance of the Hidden Markov Model integrated Prompt-Tuning model in part-of-speech tagging, aiming to provide a more accurate method for part-of-speech

tagging.

References

- [1]. Zheng, Hongxin. 2016. Basic Theories of Traditional Chinese Medicine. China Press of Traditional Chinese Medicine.
- [2]. Li, Candong. 2016. Diagnostics of Traditional Chinese Medicine. China Press of Traditional Chinese Medicine.
- [3]. Brill E. Transformation-based error-driven learning and natural language processing: A case study in part-of-speech tagging[J]. Computational linguistics, 1995, 21(4): 543-565.
- [4]. Wang Y, Jiang W, Jiang Y, et al. On Corpus Construction and Joint Modeling for Clinical Pulse Feeling and Palpation Information Extraction and Lexical Analysis of Traditional Chinese Medicine[C]//Proceedings of the 22nd Chinese National Conference on Computational Linguistics. 2023: 352-363.
- [5]. Zhang H, Yang Z, Fu Y, et al. BERT 4EVER@ EvaHan 2022: Ancient Chinese Word Segmentation and Part-of-Speech Tagging based on Adversarial Learning and Continual Pre-training[C]//Proceedings of the Second Workshop on Language Technologies for Historical and Ancient Languages. 2022: 150-154.
- [6]. Pathak D, Nandi S, Sarmah P. Part-of-speech tagger for Assamese using Ensembling approach[J]. ACM Transactions on Asian and Low-Resource Language Information Processing, 2023, 22(10): 1-22.
- [7]. Han J, Zhao S, Cheng B, et al. Generative prompt tuning for relation classification[J]. arXiv preprint arXiv:2210.12435, 2022.
- [8]. Chen, Deguang, Ma, Jinlin, Ma, Ziping, et al. 2021. A Survey of Pretraining Techniques in Natural Language Processing. Journal of Computer Science and Exploration, 15(8), 1359.
- [9]. Liu P, Yuan W, Fu J, et al. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing[J]. ACM Computing Surveys, 2023, 55(9): 1-35.
- [10]. Tian Y, Guo Y. Ancient chinese word segmentation and part-of-speech tagging using data augmentation[C]//Proceedings of the Second Workshop on Language Technologies for Historical and Ancient Languages. 2022: 146-149.
- [11]. Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint arXiv:1810.04805, 2018.
- [12]. Wang Y, Li K, Jiang Y, et al. Four Diagnostic Description Extraction in Clinical Records of Traditional Chinese Medicine with Batch Data Oversampling[C]//Proceedings of the 21st Chinese National Conference on Computational Linguistics. 2022: 611-622.
- [13]. Lewis M, Liu Y, Goyal N, et al. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension[J]. arXiv preprint arXiv:1910.13461, 2019.