IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received August 20, 2019, accepted September 8, 2019, date of publication September 16, 2019, date of current version October 4, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2941503

AttentiveHerb: A Novel Method for Traditional Medicine Prescription Generation

ZHI LIU[1,2,3], ZEYU ZHENG[1,2], XIWANG GUO[4,5], (Member, IEEE), LIANG QI[6], (Member, IEEE), JUN GUI[1,7], DIANZHENG FU[1], QINGFENG YAO[1,2,3], AND LUYAO JIN[8]

[1]Department of Digital Factory, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China
[2]Institutes for Robotics and Intelligent Manufacturing, Shenyang 110016, China
[3]University of Chinese Academy of Sciences, Beijing 100049, China
[4]Helen and John C. Hartmann Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, NJ 07102, USA
[5]Computer and Communication Engineering College, Liaoning Shihua University, Fushun 113001, China
[6]Department of Intelligent Science and Technology, Shandong University of Science and Technology, Qingdao 266590, China
[7]School of Information Science and Engineering, Shenyang University of Technology, Shenyang 110870, China
[8]School of Computer Science and Engineering, Northeastern University, Shenyang 110819, China

Corresponding authors: Zeyu Zheng (zhengzeyu@sia.cn) and Liang Qi (qiliangsdkd@163.com)

This work was supported in part by the National Key Research and Devleopment Program of China under Grant 2018YFF0214704, in part by the Liaoning Province Education Department Scientific Research Foundation of China under Grant L2019027, and in part by the Liaoning Province Dr. Research Foundation of China under Grant 20170520135.

ABSTRACT In this paper, we propose a novel intelligent model, called AttentiveHerb, for simulating the doctor's inquiry and prescription that is composed by a series of herbs. It can automatically simulate some principles and learns the interaction between symptoms and herbs from clinical records of traditional herbal medicine. This model consists of two different attention mechanisms for distinguishing the main symptoms and paying different attention to different symptoms. By experiments, in terms of the predicted prescriptions, 51% of the total cases are in full accordance with the labels; in 1.09% of cases, all herbs of a label can be found in the predicted prescription and the predicted prescription have other additional herbs; in 15.4% of cases, all herbs of a predicted prescription can be found in their corresponding label; in 22.41% of cases, several herbs in each predicted prescription overlap with its label; and 10.1% of cases are completely different from the label. In summary, 67.49% of the predicted prescriptions are close to the labels, and 89.9% contain the same herbs with the labels, which indicates that the prescriptions generated by our model are close to those by doctors. Besides, our model can recommend herbs that do not appear in the label prescriptions but are useful for relieving symptoms. It shows that our model can learn some interactions between herbs and symptoms. With enough normalized traditional herbal medical records, this model works more accurately. This study also provides a benchmark for the upcoming researches in intelligent inquiry and prescription generation of traditional herbal medicine.

INDEX TERMS Attention mechanism, deep learning, neural network, sequence learning, traditional herbal medicine.

I. INTRODUCTION

The effective of China's traditional herbal medicine has been well recognized by more and more people and nations. The effect of the herbal medicine on symptom remission have been studied by many researchers around the world [1], [2]. According to the records of the World Health Organization, approximately 4 billion people in the world are using herbal medicine to treat diseases [3], [4]. In China, traditional herbal medicine is one of the most significant ways to treat a disease, and the treatments of Traditional Chinese Medicine (TCM) mainly include herbs, acupuncture, cupping. There are a lot of classics and well-known traditional herbs. As shown in Figure 1, the *Inner Canon of Huang di* is the first medical instructive classic with a 2500-year history, which has always been of a significant value in the research of contemporary TCM. Recently, there are a lot of contributions to contemporary medicine provided by the researches on traditional herbs. The Artemisia annua Linn is a herb from which Artemisinin is extracted by Youyou Tu and her group

The associate editor coordinating the review of this manuscript and approving it for publication was Kathiravan Srinivasan.

FIGURE 1. ***The Inner Canon of Huangdi*** **(the left picture) and Artemisia annua Linn (the right picture) are the representations of TCM, which are the first instructive classic in history and extracting artemisinin respectively.**

who won the Nobel Prize in 2015. However, since traditional herbal medicine is taught only between the master and an apprentice, it takes dozens of years for a TCM practitioner to master the TCM principles, the pros and cons of herbs, and the compatibility between herbs (i.e., some herbs can cooperate with each other, while others may counteract with each other). Thus, researches of intelligent TCM are extremely of importance. The researches of artificial intelligence in TCM have never stopped, and are becoming popular especially in the recent five years with the booming of deep learning (DL) technology.

The initial attempts in intelligent TCM are mainly the rule-based expert systems [5]–[8]. They are mainly divided into two forms: (1) researchers exploit the traditional-machine-learning algorithms to find the relations between classic prescriptions and diseases; and (2) rule-based expert systems convert traditional classics and experiences of famous experienced TCM practitioners into rules, performing a series of decision-making such as determining syndrome, rules of treatment, and providing a whole classic prescription. In terms of DL methods, Liu *et al.* [9] suggest that methods for solving syndrome diagnosis should combine cognitive mechanisms of the brain. They exploit deep learning and multi-label learning to substitute relatively shallow structure methods for chronic gastritis. Peng *et al.* [10] regard the diagnosis of Parkinson in TCM as a multi-label classification and propose a latent Dirichlet allocation multi-label method, which learns global correlations between labels. Xie *et al.* [11] predict the relationship between classic TCM prescriptions and diseases by using traditional neural language process (NLP) technologies such as SVM with TF-IDF and neural network.

Wang *et al.* [12] propose a conditional probabilistic model to jointly analyze symptoms, diseases and herbs from TCM records. In their work, the conditional probabilistic model can discover the asymmetric causal relation among symptoms, diseases and herbs such as symptom-disease and herb-disease associations, which is one of the pattern recognition algorithms not for prescription generation. Ruan *et al.* [13] propose a novel cluster, called THCluster, by utilizing random walks, Bayesian rules and expectation maximization (EM) for clustering the categorization of herbs on a heterogeneous information network. In THCluster, the random walk model estimates the available probabilities of entities (e.g. herbs); the Bayesian rules and EM estimate the posterior probabilities of entities (e.g. herbs). However, the THCluster is a clustering method, which does not aim to automatically and comprehensively generate a TCM prescription. Huang *et al.* [14] propose an algorithm, called PaReCat, by integrating a herb-symptom dictionary, diffusion component analysis (DCA) and agglomerative clustering method, aiming to subcategorize patients with a general disease into more detailed groups corresponding to variations of that disease. The PaReCat is to categorize patients into different groups that can be treated more precisely but not to focus on the TCM prescription generation. From [12] to [14], these works are typically using clustering methods to discover some useful patterns among patients' records. Based on a heterogeneous information network, Ruan *et al.* [15] propose an unsupervised analysis model, named AMNE, for discovering the complex interactions between herb compositions and corresponding symptoms by utilizing autoencoder, link prediction and clustering methods. AMNE is not a prescription generation study but a herb-symptom correlations discovery study. Thus, it does not focus on TCM prescription generation. Yao *et al.* [16] propose a topic model, named PTM, to explicitly describe the process of TCM prescription generation based on the TCM theories. PTM recommends herbs/symptoms to symptoms/herbs, discover the common useful patterns in prescriptions by using herb compatibility and herbs efficacy. Different from ours, they focus on the prescription and discover the herb-symptom interaction inversely. PTM is not based on symptoms to recommend herbs but based on the herb-symptom interaction. In [16], it also provides a meaningful TCM dataset. Hu *et al.* [17] propose a dual-CNN and latent Dirichlet allocation (LDA) based model to construct a prescription from tongue images, which is consisted of two separated CNNs where the smaller CNN is regarded as auxiliary therapy topic feature extractor and trained by the multi-task learning strategy. However, in this form of prescription generation, no sequential information between herbs and symptoms are taken into account which is important in TCM prescription generation as described in our work and [18]. It only uses tongue information to generate herbs, which can only learn the projection from tongue features to herbs. Li *et al.* [18] use seq2seq framework to generate TCM prescription from symptom of a patient. In order to solve the repetition problem caused by seq2seq framework, they utilize the coverage mechanism and soft loss function. Even through it has the same foundation with ours, its main purpose is different. In our work, our goal is that the designed two attention mechanisms can empower our model the ability of learning the herb-symptom interaction knowledge, the compatibility of herbs and the TCM principles such as treatment based on syndrome differentiation, while [18] relieves repetition problem in prescription generation. We have added the performance comparison between ours and the model proposed in [18], which is showed in the following section. Because different herbs can cure the same symptoms, the current conventionally used evaluation metrics are not reasonable to estimate generated TCM prescriptions such as precision, recall, and F1. These metrics strongly depend on the labels (real prescriptions) which cannot correctly evaluate the efficacy of herbs that do not appear in real prescriptions but appear in generated ones. Thus, instead of using these metrics, we invited human doctors to evaluate the model's prediction.

The aforementioned studies are important for the development of intelligent TCM. However, the initial attempts cannot generate a useful prescription since they have not considered the TCM principles (e.g., principles of integrity about the human body, relevance and systematization over viscera, the sophisticated relationships between symptoms and syndrome, and the interactions between symptoms and herbs). Such deficiency also leads to several serious flaws intrinsically, such as the generation of fixed rules and rules created by human, and the inability to accurately describe complex rules. Besides, the fixed rules cannot be adjusted dynamically. With the emergence and development of deep learning, most recent works are incorporating the learning of a few TCM principles into their model and focus on the determination of syndrome, which makes a significant progress in intelligent TCM. However, they still cannot model the necessary principles or generate effective prescriptions. Additionally, other prescription-generation studies cannot model the TCM principles or characteristics, such as the compatibility between herbs and temporal relationship between symptoms. Despite all this, the method of deep learning offers us a series of inspirations.

The breakthrough in deep learning in recent years makes it possible to model these complex TCM principles. Deep learning was proposed by Hinton and his group [19] in 2006. In the recent 5 years, the deep-neural-network-based methods have made a large number of breakthroughs in Computer Vision (CV) and NLP including image captioning [20], subjectivity classification [21], image understanding [22], machine translation [23], reading comprehension [24], relation classification [25], abstract generation [26], bioinformatics analysis [27], visual question answering [28].

From the perspective of artificial intelligence, the inquiry and prescription generation can be regarded as a sequence process task (sequence process task aims to analysis and process sequential data such as language, text, and audio). The input of the model is a series of symptoms that patients tell the TCM practitioners, and the output of the model is herbs and their doses. We simulate this process as a translation from symptoms to herbs and regard it as a sequence to sequence (seq2seq) [29] task and integrate attention mechanism [30] into seq2seq model to enhance the quality (effectiveness and diversity of predicted herbs) of the model's output. Basically, an attention mechanism is a function which takes a sequence (e.g. a sentence) as the input and generates another same-length vector which is called attention weights. Each weight in the vector corresponds with an element in the sequence, and represents different importance of each element in the sequence; then a context vector can be generated through a dot-product between the attention weights and the sequence. The attention mechanism has been widely used in NLP tasks. The proposed model can learn TCM principles and automatically generate a useful TCM prescription according to the learned principles and interactions between herbs and symptoms. We call the whole process the ***TCM prescription generation task*** in this paper. An example of a complete herbal prescription is shown in Table 1.

TABLE 1. An example of a complete TCM prescription including both symptoms and their corresponding prescription.

Symptoms	Prescription
difficulty falling asleep, restless sleep, dreaminess, fatigue and lack of strength, amnesia, acrohypothermy, pink tongue	sliced processed aconite 60g, rhizoma zingiberis 20g, prepared radix glycyrrhizae 5g, radix codonopsis pilosulae 30g, milkvetch root 30g, angelica 15g, Atractylodes macrocephala 15g, cinnabar prepared tuckahoe with pine 15g, high aspiration 15g, radices morindae 20g stir-baked semen ziziphi spinosae 15g, costus root 15g, Arillus longanae 15g, epimedium brevicornum 20g, semen boitae 15g, semen amomi 15g, dried tangerine or orange peel 15g, Cuscuta sinensis Lam. 15g,

The main contributions of our work are summarized as follow:

(1) We propose a novel method to dynamically model the TCM inquiry and prescription generation processes. This method can be extended to any form of sequence-based diagnosis-treatment processes.
(2) The proposed model can learn some widely-used TCM principles and the interaction between symptoms and herbs automatically from clinical medical records.
(3) Experimental results demonstrate that our model can not only generate the similar predicted prescription with the label prescription, but also recommend herbs that not appear in the label prescription but are of great benefit to symptoms. This study also provides a baseline for the upcoming researches in intelligent TCM.
(4) Our model can guarantee the effect of predicted prescriptions with small-scale samples.

The rest of this paper is organized as follow. The details of the method are presented in Section II. Section III gives experimental results and analysis. The conclusion and future work are discussed in Section IV.

II. METHODS

A. INPUT AND OUTPUT FORMULATION

In this work, we use a recurrent-neural-network-based seq2seq model. Regarding the form of the symptoms and prescriptions in TCM records, we take the input and output from the idea of machine translation [23]. Each symptom

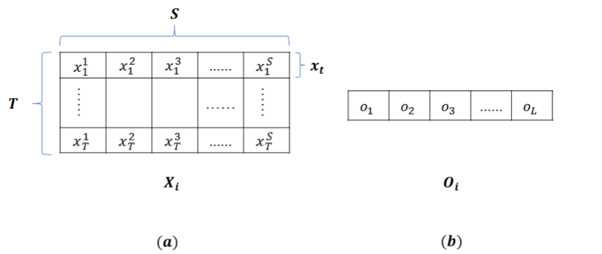

FIGURE 2. The standard input and output of proposed model: (a) each input sample is a matrix X, and each row of X denoted by x_t represents a symptom formulated as a word vector with S dimensions. (b) The output is a herb sequence containing L herbs.

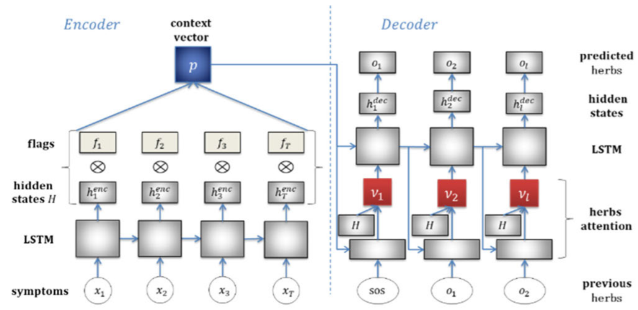

FIGURE 3. An illustration of our model. Our model is built based on the seq2seq framework and is enhanced with a two-stage attention mechanism. In EncLSTM, each symptom is encoded and combined with the main-subordinate flags to compute the context vector p. Then, GenLSTM, whose first hidden state is initialized by p, computes the herb attention to generate each herb step by step, where x_1, x_2, and x_3 represent symptoms insomnia, vexation, and dreaminess respectively, and o_1, o_2, and o_t represent herbs Greenish Lily Bulb, Poria cocos, and the fruit of Chinese wolfberry, respectively.

is provided in the form of a word or a phrase vector, and each herb is presented as a word or a phrase. As shown in Figure 2(a) and (b), the input symptoms are denoted by a matrix X with the dimension of $T * S$, where T represents the number of symptoms and S represents the dimension of a vector of each symptom. Each row in $X$is denoted as x_t ($t \in \{1, \ldots, T\}$) which means the t-th symptom. The output denoted by $\mathrm{O} = \{o_1, o_2, \ldots, o_L\}$ is a sequence of L herbs, where o_l ($l \in \{1, \ldots, L\}$) represents the l-th herb. $D = [(X_i, O_i)]_{i \in N}$ represents the whole data set.

B. ATTENTIVEHERB: END-TO-END TCM PRESCRIPTION GENERATION MODEL

AttentiveHerb is a model which is able to generate herbs at every step to relieve symptoms of patients. The symptoms are firstly fed into a Long Short Term Memory Network (LSTM) [31] one by one (called EncLSTM for short), which will map these symptoms into hidden features. Then, another LSTM (called GenLSTM for short) takes these features as input and produces herbs. In EncLSTM and GenLSTM, we respectively propose two different attention mechanisms to model TCM principles and partial temporal orders among symptoms in order to improve the final prescription's quality:

1) Main-subordinate symptom attention. It is used to distinguish the main symptoms from the subordinates and keep more information of the main symptoms in the process.

2) Dynamical herb attention. It considers different symptoms when generating herbs each time.

In summary, AttentiveHerb consists of four steps: (1) EncLSTM encodes each symptom; (2) Main-subordinate symptom attention distinguishes the main and subordinate symptoms; (3) Dynamical herb attention pays different attention to different symptoms; and (4) GenLSTM generates herbs by using the both attentions. The framework of AttentiveHerb is shown in Figure 3.

C. SYMPTOMS ENCODER (EncLSTM) AND HERBS DECODER (GenLSTM)

We use a normal LSTM as the EncLSTM to model the information and temporal dependency of symptom sequence. We use a normal LSTM instead of bidirectional LSTM that is widely used in NLP for two reasons. First, the model is easily overfitting when the model's complexity is too high whereas the dataset is relatively small. Additionally, it is good enough to encode the information and generate relatively accurate herbs in our experiments by using normal LSTM. Second, bidirectional LSTM is designed to enrich the linguistic information based on the past and future context to solve language tasks. However, TCM prescription generation is not a complete language task. It does not include all the sequential dependency like machine translation. The bidirectional LSTM will also increase computational cost of the training and testing processes. EncLSTM encodes a sequence from x_1 to x_T, generating each hidden state h_t^{enc}, and produces a set of hidden states $H^{enc} = \left\{h_1^{enc}, \ldots\ldots, h_T^{enc}\right\}$.

During the prescription generation, GenLSTM takes v_l and h_l^{dec} to get herbs, where v_l is computed by dynamical herb attention and h_l^{dec} is initialized by p. Both v_l and p will be described in detail in following content. The first input of GenLSTM is a special start-of-sequence character and generation process will proceed until the special end-of-sequence character is encountered. In this process, the l-th input for computing v_l is assigned by the previous output, i.e., $x_l = o_{l-1}$. The reason we use this way to generate herbs is that TCM practitioners always decide what herbs would be added according to the patient's symptoms and the herbs that were prescribed previously. It also reflects the principle of compatibility between herbs.

D. MAIN-SUBORDINATE SYMPTOM ATTENTION

TCM practitioners always pay more attention to the herbs that can treat the main symptoms when they make prescriptions for patients. After that, they will consider comprehensively what herbs should be added in the prescription to relieve other subordinate symptoms and the toxicity of previous herbs, which means that symptoms occur in a chronological order. This principle is reasonable and effective because many subordinate symptoms appear because of the occurrence of the main symptoms. When the main symptoms are cured, the subordinate symptoms can be relieved as long as the

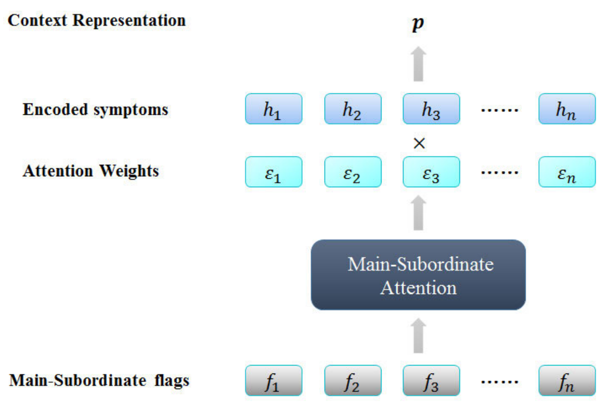

FIGURE 4. The main-subordinate attention. This attention model uses a set of main symptom flags *F* to produce corresponding attention weights and then generate intermediate representation *p* by calculating the weighted sum between attention weights and hidden states produced by EncLSTM.

patients take a little or even no herbal medicine. Therefore, paying more attention to the main symptoms is of quite importance. Although LSTM can encode necessary information of the symptoms, it cannot distinguish which symptoms are the main symptoms. Therefore, we propose the main-subordinate symptom attention to enable our model to differentiate the main symptoms from subordinate ones. This attention mechanism has the ability to recognize the main symptoms and keep more information about the main symptoms when producing the intermediate context representation p. Before we compute p, we calculate a set of attention weights which contain T elements with each one denoted by ε_t. For each symptom, there is a corresponding index f indicating whether this symptom is the main symptom, i.e., $f = 1$ if it is the main symptom and otherwise $f = 0$. Every sequence of the symptoms has a corresponding sequence of index $F = \{f_1, \ldots, f_T\}$, and then the attention weights are calculated by the following equation:

$$\varepsilon_t = \frac{exp\left(W_\varepsilon \cdot h_t^{enc}\right)}{\sum_{j=1}^{T} exp\left(W_\varepsilon \cdot h_j^{enc}\right)} * \frac{exp\left(f_t\right)}{\sum_{j=1}^{T} exp\left(f_j\right)} \quad (1)$$

Each ε_t is a scalar computed through the hidden states H^{enc} produced by EncLSTM. W_ε is a set of trainable parameters that are jointly trained with other model's parameters. Our objective is to enable model make the self-adjusting attention weights for the main and subordinate symptoms and distinguish whether it is the main one guided by F. Once these attention computations are completed, we can obtain the intermediate context representation p as a weighted sum of hidden states H^{enc} from EncLSTM, i.e., $p = \sum_{j=1}^{T} \varepsilon_j \times h_i^{enc}$. The interpretation of this operation is that ε_t means how much attention should be paid to different symptoms. A diagram of this attention mechanism is presented in Figure 4.

E. DYNAMICAL HERB ATTENTION

Before the herb generation in GenLSTM, we compute the dynamical herb attention. Our goal is to capacitate our model to generate proper herbs according to symptoms. However, incorporating the whole symptoms as context to generate each herb is not quite effective and efficient. Thus, we propose a dynamical herb attention to model the interactions between symptoms and herbs, and the relationships between herbs. The attention is represented as $\boldsymbol{\lambda}$ and takes the previously predicted herb o_{l-1} and the decoder hidden states h_{l-1}^{dec} as input. $\boldsymbol{\lambda}$ is a vector that contains T scalars. Each value in $\boldsymbol{\lambda}$ corresponds to a symptom, and it means the extent of the herb in the l-th step affecting to each symptom. The bigger value of the t-th element in $\boldsymbol{\lambda}$, the greater the effect of the l-th herb on t-th symptom. $\boldsymbol{\lambda}$ is calculated as follows:

$$\boldsymbol{\lambda} = softmax\left(W_\lambda \cdot \left[o_{l-1}, h_{l-1}^{dec}\right]\right) \quad (2)$$

$$softmax = \frac{exp\left(x_t\right)}{\sum_{j=1}^{T} exp\left(x_j\right)} \quad (3)$$

Under the premise of referring to the previously generated herb and H^{enc}, the model shows which symptoms can be mainly relieved by the herb produced in the l-th step through the corresponding herb attention. After $\boldsymbol{\lambda}$ is obtained, we can obtain a new input v to GenLSTM, i.e., $v = \sum_{j=1}^{T} \lambda_j \times h_j^{enc}$.

TABLE 2. The properties of our dataset. Noticed that, the range of the lengths of both symptoms and prescription are relatively large, which induce a big challenge for traditional methods to encode symptoms and generate herbs.

Properties of Data Set	
The minimum number of symptoms included in symptoms sequence	1
The maximum number of symptoms included in symptoms sequence	33
The average number of symptoms included in symptoms sequence	17
The minimum number of herbs contained in label prescription	1
The maximum number of herbs contained in label prescription	40
The average number of herbs contained in label prescription	18.6
Total Number of Symptoms	5784
Total Number of Herbs	1751

III. EXPERIMENTS

A. DATA SET AND EXPERIMENTS SETTINGS

Because of the lack of formal and open source TCM digital records, we cooperate with researchers from the Institute of Information on Traditional Chinese Medicine, China Academy of Chinese Medical Sciences (CACMS) and collect an insomnia dataset. Each record in this dataset is validated through clinical practice. Then, we extract the key information from these records to obtain a directly trainable dataset which consists of symptoms, diagnosis of TCM, syndrome, rules of treatment, prescription, and therapeutic effect. The total number of samples in our dataset is 6428. We divide them into a training set and a test set, where the former has 5785 samples (about 90%) and the latter has 643 samples (about 10%). Some properties about the dataset are shown in Table 2.

We calculate and display the distribution of the length of the symptom sequence and the distribution of the number

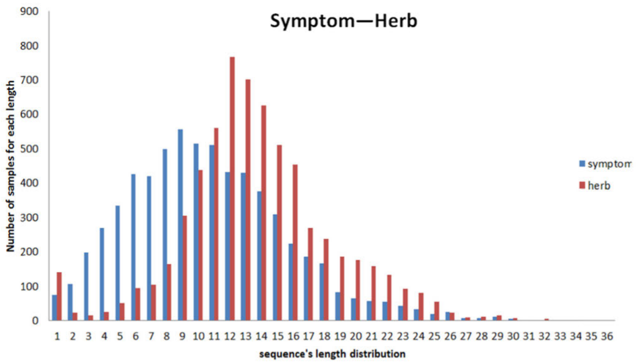

FIGURE 5. The distribution of number of symptoms and herbs and its corresponding number of samples. The number of symptoms and herbs has a large range, and the number of samples containing more than 26 herbs is much less than samples containing less than 26 herbs.

TABLE 3. Parameters settings used in AttentiveHerb and LSTM models.

Parameters Name	Illustration	Value
lstm_layers	the number of LSTM layers	2
lstm_hidden_dim	the dimensions of LSTM layers' hidden	256
attention_hidden_dim	the dimensions of attention layers' hidden	256
embeded_dim	embedding dimensions for both symptoms and herbs	300
batch_size	the size of batch	16
lr_rate	the learning rate	0.001
dropout_p	the rate of dropout	0.5
clip_p	the gradient clip rate	0.9
num_steps	the total number of training steps	30,000

of herbs in each prescription. As shown in Figure 5, most samples have less than 26 herbs or symptoms. It shows that the number of both symptoms and herbs that is more than 26 is not common in clinical practice. The model is prone to overfitting if the samples with more than 26 symptoms or herbs are included in training set. In order to prevent it, we remove these samples.

In our experiments, we choose three baseline models for comparison: 1) LSTM seq2seq, a normal seq2seq framework implemented by two normal LSTM models which are called LSTM-encoder and LSTM-decoder (encoder and decoder for short), respectively; 2) an attention-based LSTM seq2seq, where an attention layer is added into the decoder; and 3) the model proposed in [18] (coverage-soft-model for short) which also can be seen as attention-based LSTM seq2seq since all the architecture updates occur in decoder by the attention-like manner. All LSTM-based models share the same LSTM settings as in Table 3. For the coverage-soft-model, we implement coverage-soft-model by using the exact parameters proposed in [18]. All models are trained by Adam algorithm on a P100 GPU.

B. MODEL EVALUATIONS

Since we regard the TCM prescription generation as a generative task instead of a classification task, there are no proper evaluation metrics for the predicted prescription. It is not reasonable to simply use the metrics such as precision, recall, accuracy or BLEU [32] which is popular in machine translation to evaluate the predicted prescription. Because the TCM prescription does not completely contain the relationship in the languages, but contains some TCM principles that consider the compatibility between different herbs, such as "Traditional Chinese medicine formulation" (also called "Jun Chen Zuo Shi").[1] It is also not reasonable to only count the number of herbs that appear in label prescription since the same symptoms may be caused by different reasons. Therefore, it is necessary to find the real cause to use the symptomatic herbs under the guidance of syndrome differentiation. The model may select different effective herbs according to the learned interactions between herbs and symptoms (different TCM practitioners may recommend different herbs for the same patient but all prescribed herbs are provided under the guidance of syndrome differentiation). Only counting the number of herbs according to the label prescription cannot reflect this diversity in useful herbs. Thus, we invite TCM doctors to evaluate the generated prescriptions of all compared models, then the statistics and analysis of the results are conducted under the help of human doctors instead of simply comparing the scores like accuracy, precision, or BLEU.

We invite TCM doctors from different TCM institutes such as the Puji Outpatient Department of Traditional Chinese Medicine in Liaoning, China to evaluate generated prescriptions independently. All involved doctors estimate generated prescriptions from the following aspects: 1) Herbs Effectiveness (HE for short); it evaluates whether the herbs in generated prescriptions are useful for relieving the given symptoms. The range of HE score is [0, 5], and the higher scores, the more effective of these herbs. 2) Herbs compatibility (HC for short); it assesses if the herbs in a generated prescription can assist each other and has less antagonism. The range of HC score is [0, 5], and the higher scores, the higher compatibility of these herbs, and vise verse. The doctors evaluate them according to the TCM theories, principles, and their experience. The estimations are listed in Table 4.

As shown in Table 4, applying LSTM seq2seq directly on the TCM prescription generation can obtain some results and generate some useful herbs. However, the herbs compatibility performance is seldom learned by the model. When adding the attention layer into the decoder, the model gains 38% and 80% improvements on HE and HC respectively. The use of coverage mechanism and soft loss on the basis of attention-based LSTM seq2seq also improves the HE and HC scores. Nevertheless, because the goal of the coverage mechanism and soft loss is to relieve repetition problem of herb

[1] Herbs in a prescription can be distributed into 'Jun' medicine, 'Chen' medicine, 'Zuo' medicine, 'Shi' medicine. 'Jun' (Principal herbs) treats the main cause of the disease. 'Chen' (assistant herbs) enhances the effects of 'Jun' and treats accompanying symptoms. 'Zuo' (complement herbs) reduces or eliminates possible toxic effects of the 'Jun' and 'Chen' herbs and treats accompanying symptoms. 'Shi' (guide and mediator herbs) helps to deliver or guide the above three kinds of herbs to the target organs, mediating effects of all above kinds of herbs.

TABLE 4. Comparison results for prescription generation of various models with human metrics. Higher values of these metrics indicate a better performance.

Model	HE	HC	Total score
Label	4.7	4.2	8.9
LSTM seq2seq	2.1	0.5	2.6
Attention-based LSTM seq2seq	2.9	0.9	3.8
Coverage-soft-model	3.1	1.2	4.3
AttentiveHerb	**4.3**	**3.1**	**7.8**

TABLE 5. The ablation study of the AttentiveHerb model. Higher values of these metrics indicate a better performance.

Model	HE	HC	Total score
AttentiveHerb	**4.3**	**3.1**	**7.8**
Main-subordinate attention only	3.4	2.2	5.6
Dynamic herb attention only	2.9	0.9	3.8
Without any attentions	2.1	0.5	2.6

TABLE 6. Comparison results for prescription generation of various hyperparameters of the AttentveHerb model. Higher values of the metrics indicate a better performance.

Model	Lstm layer	Lstm hiddens	Attention hiddens	Embeded hiddens	Total scores
AttentiveHerb-1-128-128-300	1	128	128	300	6.3
AttentiveHerb-1-256-256-300	1	256	256	300	6.5
AttentiveHerb-1-512-512-300	1	512	512	300	6.8
AttentiveHerb-2-128-128-300	2	128	128	300	7.1
AttentiveHerb-2-256-256-300	**2**	**256**	**256**	**300**	**7.8**
AttentiveHerb-2-512-512-300	2	512	512	300	7.3

recommendation, the improvements are limited. As described in Section II. D, we design the main-subordinate attention to enable the AttentiveHerb to differentiate the main cause and keep more main cause information throughout the process, integrating these kinds of information into the prescription generation procedure. The experiments demonstrate that the main-subordinate attention improves both HE and HC scores with a big margin especially on HC score.

In order to investigate the effects of the two proposed attention mechanisms, we conduct the ablation study of our model. The results are listed in Table 5. As shown in the table, different attentions have different improvement on HE and HC. Our experiments show that the main-subordinate attention has more effect than the dynamic herb attention, which means that the main cause information carries more useful pathology features. It demonstrates that keeping more useful pathology features in the first place are more effective than the later complex computation on mixed features.

We also investigate the performance of our model on different hyperparameters to find the relatively better results while considering the computational costs. All results are listed in Table 6. We name our models with the values of different hyperparameters for convenience. For example, AttentiveHerb-2-256-256-300 means that the proposed model has 2 LSTM layers with 256 hidden dimensions in both decoder and encoder, 2 attention layers with 256 hidden dimensions, and input embedding layers with 300 dimensions. Since the performance differences of DL model mainly come from the information encoding layers (i.e., LSTM layers and attention layers), we conduct various experimental settings on our model. All variants of AttentiveHerb are trained with the same training settings as described in Table 3. As shown in Table 6, with the increase of the capacity of LSTM layer and attention layer, the performance of AttentiveHerb is also increasing. For setting of 512 with 2 LSTM layers, the performance reduction is mainly caused by the limitation of training samples since AttentiveHerb overfits the training set when the total training steps are completed.

C. EXPERIMENTS ANALYSIS

We analyze all the predicted prescriptions in test set. It shows that all predicted prescriptions can be divided into five groups. The prescriptions in the first group are fully in accordance with label prescription, which means every single herb matches the herbs in label prescription. In the second group, the predicted prescriptions contain several more herbs on the basis of including all the herbs in label prescriptions. In the third group, herbs of predicted prescriptions are all in the list of label prescriptions, and the label have some additional herbs. In the fourth group, the predicted prescription contains herbs that are included in label prescription, whereas the rest of the herbs are not in the label prescriptions. The last one is that the predicted prescription is totally different from label prescription. Table 7 shows the cases of all five groups which can be found in Appendix. The first part, an example of the first group for instance, shows that when a symptom set is given, the model may generate a prescription that is completely consistent with the label prescription. The herbs in the predicted prescription that is different from the label prescription are marked in bold, and herbs that considered as 'Jun' medicine are marked with underlines.

We now discuss why our model chooses different herbs like the fourth part in Table 7. We visualize the attention weights of herb attention to show the interaction model learned. From Figure 6, it is a symptom-herb attention relationship, and each value in each column represents the extent to which this herb can relieve each symptom. In other words, the model decides which symptoms the herbs mainly act on. The greater the value of the symptom, the more effective this herb treat the symptom. After performing researches on

TABLE 7. Four different sets of symptoms, corresponding label and predicted prescriptions.

Symptoms	Label Prescription	Predicted Prescription
Insomnia, Vexation, Dreaminess, Dizziness, Tinnitus, Amnesia, Red Tongue with Little Coating, Thready Rapid Pulse	Greenish Lily Bulb, Poria Cocos, The Fruit of Chinese Wolfberry, The Fruit of Chinese Magnoliavine, Herba Patrinia, Caulis Polygoni Multiflori, Polygala Root, Acorus Gramineus Soland, Pinellia Ternary, Selfheal, Coptis Chinensis Franch, Ligusticum Wallichii, Rhizoma Anemarrhenae, Liquorice, Semen Zizyphi Spinosae, Radix Curcumae, Raw Dragon's Teeth,The Root of Red-Rooted Salvia	Greenish Lily Bulb, Poria Cocos, The Fruit of Chinese Wolfberry, The Fruit of Chinese Magnoliavine, Herba Patrinia, Caulis Polygoni Multiflori, Polygala Root, Acorus Gramineus Soland, Pinellia Ternary, Selfheal, Coptis Chinensis Franch, Ligusticum Wallichii, Rhizoma Anemarrhenae, Liquorice, Semen Zizyphi Spinosae, Radix Curcumae, Raw Dragon's Teeth, The Root of Red-Rooted Salvia
Insomnia, Whitish Greasy Coating on the Tongue, Fat, Pale and Dark Red Tongue Slippery and Sink Pulse, Exhaustion, Abdominal Distension and Anorexia	Pinellia Ternary, Concha Ostreae, Gastrodia Elata Blume, Poria Cocos, Dried Tangerine or Orange Peel, Liquorice	Pinellia Ternary, Concha Ostreae, Gastrodia Elata Blume, Poria Cocos， Dried Tangerine or Orange Peel， Liquorice, **Malt, Medicated Leaven**
Fatigued Spirit and Lack of Strength, Nervousness and Palpitation, Insomnia, Dizziness, Red Tip of Tongue, Thin White Tongue Coating, Dry Eyes, Thready Rapid Pulse	Semen Zizyphi Spinosae, Rhizoma Anemarrhenae, Ligusticum Wallichii, Poria Cocos, Bombyx Batryticatus, Dragon's Teeth, Concha Ostreae	Semen Zizyphi Spinosae, Rhizoma Anemarrhenae, Ligusticum Wallichii, Poria Cocos, Dragon Tooth, Concha Ostreae
Insomnia, Spontaneous Perspiration, Palpitation, Intolerance of Cold and Cold Limbs, Heaviness of Arms and Legs, Plump Pale Tongue, Whitish Greasy Coating on the Tongue, Deep Thready Pulse,	Cassia Twig, Radices Paeoniae Alba, Fructus Ziziphi Jujubae, Radix Aconiti Lateralis Preparata, Astragalus Mongholicus, Bighead Atractylodes Rhizome, Concha Ostreae, Dragon's Teeth, Prepared Radix Glycyrrhizae,	**Angelica Sinensis**, Radices Paeoniae Alba, Fructus Ziziphi Jujubae, Crude Astragalus Mongholicus **Radix Pseudostellariae, Radix Ophiopogonis, The Root of Red-Rooted Salvia**
Insomnia and Dreamful Sleep, Dizziness, Headache Frequently, Vexation and Irritability, Eat a Normal Diet, Pale Red Tongue, Whitening of The Middle Part of Tongue Coating, Small And Wiry Pulse	Magnetite, The Shell of Abalone or Sea-Ear, Fossil Fragments, Vermilion Poria Cocos, Curcuma Aromatica, Fructus Aurantii, Semen Zizyphi Spinosae, Caulis Polygoni Multiflori, The Root of Red-Rooted Salvia, Cortex Albizziae, Calamus	**Radix Bupleuri, Rhizoma Anemarrhenae, Radices Paeoniae Alba, Prepared Rehmannia Root, Prepared Rhizoma Pinellize Without Adju-Vant , Rhizoma Cyperi, Polygala Root, Flower of Silktree Albizzia**

herbal medicines in depth [33], querying the experienced TCM doctors and researchers of CACMS, we find that the substitutions appear in the predicted prescription have certain rationality. In Figure 6, the x-axis presents herbs predicted by our model and y-axis represents the given symptoms. Each herb's column represents an attention distribution for all

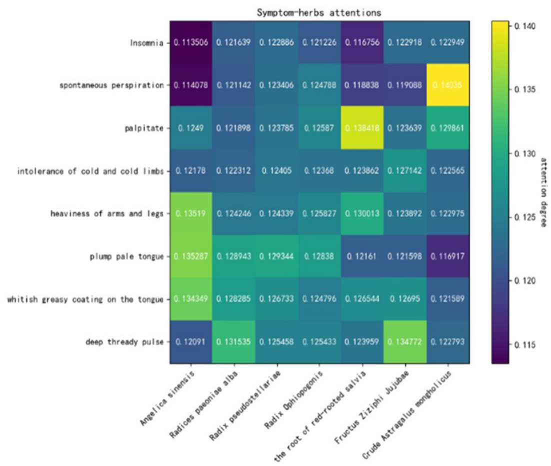

FIGURE 6. Herb-symptom attention visualization. The x-axis presents herbs predicted by our model and y-axis represents the given symptoms. Each herb's column represents an attention distribution for all symptoms. Different values in the same column correspond to the effectiveness of the herb represented by this column for different symptoms. The greater the value of a symptom, the better effect our model deems of this herb on treating this symptom.

symptoms. Different values in the same column correspond to the effectiveness of the herb for different symptoms. The greater the value of a symptom, the better effect our model deems of this herb on treating this symptom. For instance, our model holds that the *Angelica Sinensis* and *Fructus Ziziphi Jujubae* can relieve the *Deep Thready Pulse* according to the two values in the bottom row of Figure 6. The TCM knowledge shows that these two herbs are able to enrich the blood and relief the deficiency of Yin[2] or blood deficiency, which is the main cause of *Deep Thready Pulse*. However, the *Angelica Sinensis,*an effective herb to relieve deep thread pulse predicted by our model, does not appear in the label prescription. Our model also regards that the *Root of Red-Rooted Salvia* can relieve the *Palpitate*which is mainly caused by the deficiency of blood or stasis of phlegm and blood, leading to poor circulation of pulse around the heart. According to the TCM knowledge [33], the *Root of Red-Rooted Salvia* is able to activate blood circulation and remove blood stasis. The Crude *Astragalus Mongholicus* and *Radix Pseudostellariae* have the effect of tonifying and nourishing Qi, the vitality of all tissues and organs in TCM theory, and lifting Yang to invigorate spleen, which can effectively relieve *Spontaneous Perspiration*. As the third and last columns of the second row show, our model makes the same predictions.

We also analyze all symptoms and the corresponding predicted prescriptions that are categorized into the fifth set and try to find out whether these herbs are useful or not since

[2]The Yin-Yang theory is a guiding theory for medical practice, and is also one of the basic theories of TCM. Based on its concept, TCM holds that the relationship between Yin and Yang, a mutually opposite and interrelated relationship, runs through everything such as nature and human body, and is also the root and law of the occurrence, development and change of human physiology and pathology. The application of the Yin-Yang theory in TCM is mainly embodied in: explaining human body's organization, physiological activities, and pathological changes and guiding the diagnosis of the disease

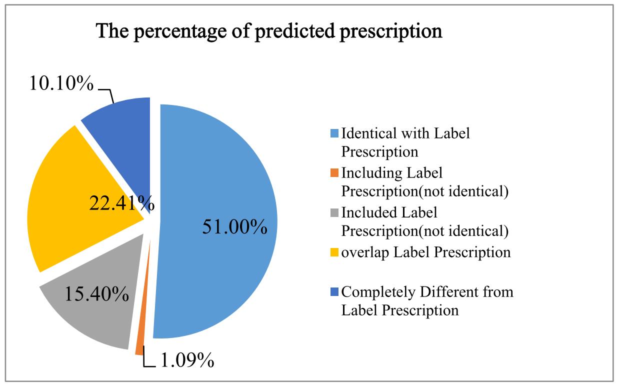

FIGURE 7. The statistics of our experiments.

prescriptions in the fifth group are totally different from the labels. According to the TCM knowledge and the results from the experienced TCM doctors and researchers of CACMS, [33], We can find out that even though all herbs in this kind of prescriptions are different from herbs in the label prescriptions, they are effective for relieving the corresponding symptoms which is the real cause and are suitable for the therapy in TCM. Taking the last part in Table 7 as an example, the syndrome of these symptoms is Qi and Blood Deficiency. Under this condition, *Insomnia* and *Dreamful Sleep* may occur, and *Dizziness*, *Frequent Headache* and *Vexation and Irritability* will be caused by *Insomnia and Dreamful Sleep*. We can see that the real cause of all the aforementioned symptoms is the deficiency of Qi and Blood. Therefore, the main treatment is tonifying Qi and Blood for relieving *Insomnia and Dreamful Sleep*. In the predicted prescription, *Radices Paeoniae Alba*, *Prepared Rehmannia Root*, *Polygala Root* and *Flower of Silktree Albizzia* can effectively nourish blood; the *Rhizoma Cyperi* and *Flower of Silktree Albizzia* can regulate Qi. By using these herbs, the main symptoms can be relieved. The other herbs in the prescription, such as Radix Bupleuri, *Rhizoma Anemarrhenae*, *Prepared Rhizoma Pinellize without Adju-Vant* and *Rhizoma Cyperi*, can relieve other symptoms, such as depression and eliminating stagnation. More sample analysis and herb selection are presented in the Appendix.

We now study whether our model learns some TCM principles such as compatibility between herbs and treatment based on the syndrome. After a series of herbs researches [33] and in-depth discussions with researchers of CACMS and experienced TCM doctors, we learn that the main cause of *Insomnia* is the malfunction of viscera, the disharmony of Qi and blood, imbalance of Yin and Yang, etc. The pathogenic organ is mainly the heart and involves other organs. The syndromes of insomnia are mainly concentrated on the deficiency of Yin or Blood, which can be relieved by the treatment law of reinforce deficiency and reduce excessiveness. The deficiency means the lack of necessity for the human body, whereas excessiveness means the redundancy of active pathogenic factors and body wastes. Notice that, herbs (e.g. *Angelica Sinensis, Astragalus Mongholicus, Radix Pseudostellariae* and the*Root of*

TABLE 8. The comparison between the label prescription, our model's prediction and the basic model's prediction. All useful herbs are in bold. This is a case from fourth group described in subsection 3.2. We can see that even though the herbs our model predicted and the label herbs have some differences, the predicted prescription is still effective and efficient for relieving these symptoms. This means that our model has an ability to create. However, the herbs that the basic seq2seq model predicted are unable to relieve the symptoms, and only a few herbs have an effect on these symptoms.

Symptoms	Insomnia, Facies Dolores, Dry Stool, Dizziness and Tinnitus, Palpitate Red Tongue and Yellow and Dry Coating on The Tongue, Taut Rapid Pulse
Label Prescription	**The Rhizome of Chinese Goldthread, Stir-Fried Fructus Aurantii Immaturus, Scutellaria Baicalensis, Bambusae Caulis Im Taeniam, Lophatherum Gracile,** Curcuma Aromatic, **Concretio Silicea Bambusae, Fritillaria Thunbergii Miq., The Fruit of Chinese Wolfberry, Pericarpium Trichosanthis, Polygala Root, Acorus Gramineus Soland., Fructus Corni, Raw Fossil Fragments, Raw Concha Ostreae**
AttentiveHerb model	**Peach Kernel, Chinese Herbaceous Peony, Raw Liquorice,** Processed Rhizoma Cibotii, **The Root of Kudzu Vine, Raw Concha Ostreae,** Cinnamomum Cassia Presl, **Pinellia Ternata, Prepared Radix Glycyrrhizae, Plantago Seed,** Malt, **Radix Pseudostellariae, Gardenia Jasminoides Ellis, Carthamus Tinctorious**
Basic seq2seq model	Rainworm, Radix Paeoniae Rubrathe, Leech, **Steamed Fructus Schisandrae with Vinegar,** Angelica Sinensis, Akebia Quinata Decne, Fructus Aurantii Immaturus, **Arillus Longanae, Acorus Gramineus Soland, Raw Liquorice**

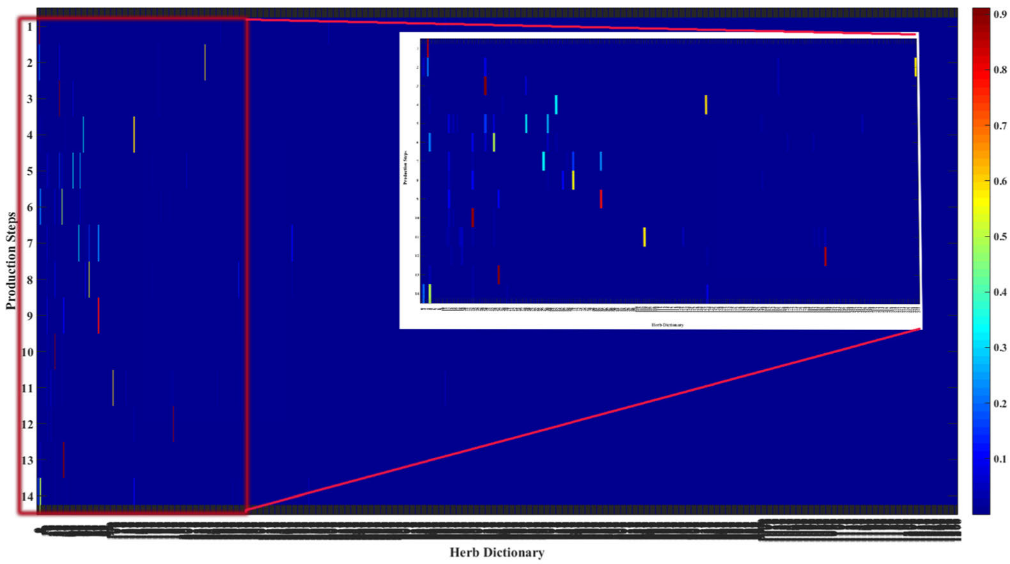

FIGURE 8. The whole process of predicted prescription generation. We demonstrate the whole process that our model selects the 14 herbs from all 1271 candidate herbs in the overall picture and the amplified image on the right top of the Figure shows the main area of the selected herbs. The x and y axis represent the prescription generation process and herb index respectively. We also provide a part of the check list of all herbs in Table 9 since the whole check list is too large to show them all.

Red-Rooted Salvia) in the predicted prescription have the effect of replenishing Qi and nourishing blood, which are in accordance with the TCM treatment law. These herbs also have a cooperative effect with each other such as Qi and blood supplementation. It also reflects that our model can learn certain TCM principles of compatibility between herbs intrinsically.

We also collect the statistics of the five types of cases, and the details are shown in Figure 7. We notice that the herbs in predicted prescriptions that are fully accordance with the label prescriptions account for 51% (the first group). The predicted prescriptions which predict several more herbs on the basis of all included herbs in label prescriptions account for 1.09% (the second group). The predicted prescriptions whose predicted herbs are all in the list of label prescriptions but have fewer herbs account for 15.4% (the third group). Based on the above three cases, the proportion of the predicted prescriptions which are close to the label prescriptions accounts for 67.49% in total. A small set of predicted prescriptions that only several herbs are same with labels account for 22.41% (the fourth group). Thus, the predicted prescriptions that include the same herbs with label

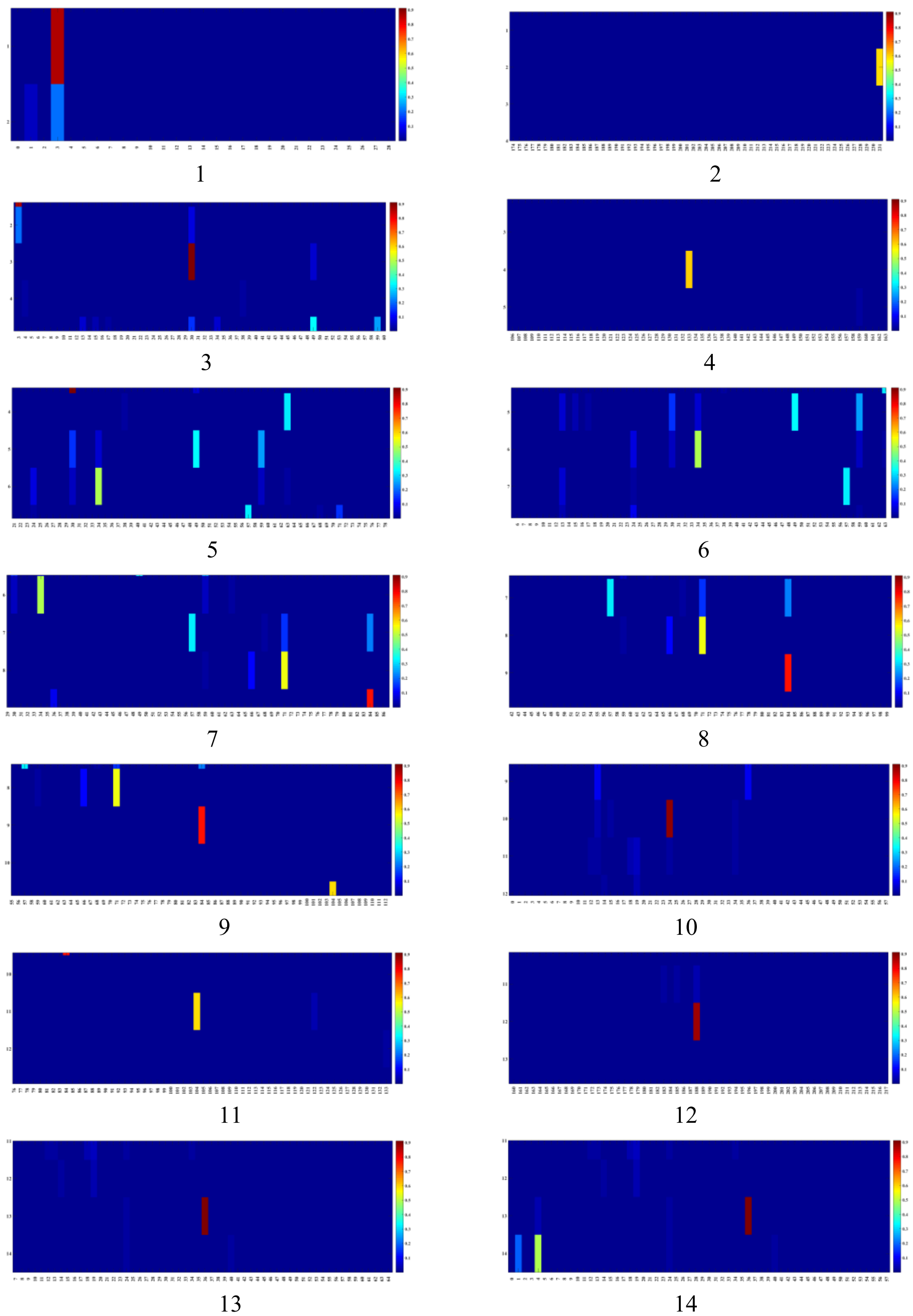

FIGURE 9. We amplify each herb generation from the whole process presented in Figure 8. Each picture represents a step and the color in each position represents the corresponding herb selection probability computed by our model. The variance of color is from blue to red; the closer the color of each position is to the red, the greater probability that the corresponding herb will be selected, and vice versa.

prescriptions account for 89.9% in total. The rest 10.1% are prescriptions that are completely different from the label ones (the fifth group). Three noteworthy aspects can be drawn from these analyses. First, our model has a strong inclination to refer to the label prescription when generating the testing prescription, which has the same manner as human; doctors always refer to the classical prescription and add or remove herbs from it according patients' personal conditions. Second, our model can learn the interactions between herbs and symptoms. Third, our model could learn that different herbs can relieve the same symptoms. That's why the prescriptions predicted by the proposed model are partly the same as the real label prescription.

IV. CONCLUSION

TCM is the accumulation of human experience in the medical value of herbs for thousands of years. This work is the first step to intelligentize the experience. In this paper,

TABLE 9. The partial herb-index check list.

Id	Chinese Name	English Name	Id	Chinese Name	English Name
1	黄花蒿	Artemisia annua Linn	43	神曲	Medicated Leaven
2	石菖蒲	Acorus gramineus Soland	44	黄芪	Milkvetch Root
3	黄芪	Astragalus Mongholicus	45	瓜蒌皮	Pericarpium trichosanthis
4	龙眼肉	Arillus Longanae	46	桃仁	Peach kernel
5	当归	Angelica Sinensis/ Angelica	47	制狗脊	Processed rhizoma cibotii
6	白术	Bighead Atractylodes Rhizome/ Atractylodes Macrocephala	48	半夏	Pinellia ternata
7	天虫	Bombyx Batryticatus	49	炙甘草	Prepared Radix Glycyrrhizae
8	竹茹	Bambusae caulis im taeniam	50	车前子	Plantago seed
9	夜交藤	Caulis Polygoni Multiflori	51	茯苓	Poria Cocos
10	黄连	Coptis Chinensis Franch	52	远志	Polygala Root
11	牡蛎	Concha Ostreae	53	熟地黄	Prepared Rehmannia Root
12	朱茯苓	Cinnabar Prepared Tuckahoe With Pine	54	清半夏	Prepared Rhizoma Pinellize Without Adju-Vant
13	秋葵	Cuscuta Sinensis Lam.	55	生龙骨	Raw Fossil Fragments/ Raw Dragon's Teeth
14	桂枝	Cassia Twig	56	生甘草	Raw liquorice
15	郁金	Curcuma Aromatica	57	生牡蛎	Raw Concha Ostreae
16	合欢皮	Cortex Albizziae	58	太子参	Radix pseudostellariae
17	广木香	Costus Root	59	知母	Rhizoma Anemarrhenae
18	菖蒲	Calamus	60	郁金	Radix Curcumae
19	生黄芪	Crude Astragalus Mongholicus	61	白芍	Radices Paeoniae Alba
20	红花	Carthamus tinctorious	62	黑附片	Radix Aconiti Lateralis Preparata
21	天竺黄	Concretio silicea bambusae	63	麦冬	Radix Ophiopogonis
22	肉桂	Cinnamomum Cassia Presl	64	柴胡	Radix Bupleuri
23	陈皮	Dried Tangerine or Orange Peel	65	香附	Rhizoma Cyperi
24	龙齿	Dragon's Teeth	66	巴戟天	Radices Morindae
25	淫羊藿	Epimedium Brevicornum	67	党参	Radix Codonopsis Pilosulae
26	浙母贝	Fritillaria thunbergii Miq.	68	干姜	Rhizoma Zingiberis
27	山茱萸	Fructus corni	69	柏子仁	Semen Boitae
28	大枣	Fructus Ziziphi Jujubae	70	炒枳实	Stir-fried fructus aurantii immaturus
29	枳壳	Fructus Aurantii	71	制附片	Sliced Processed Aconite
30	龙骨	Fossil Fragments	72	砂仁	Semen Amomi
31	合欢花	Flower of Silktree Albizzia	73	黄芩	Scutellaria baicalensis
32	栀子	Gardenia jasminoides Ellis	74	夏枯草	Selfheal
33	百合	Greenish Lily Bulb	75	酸枣仁	Semen Zizyphi Spinosae
34	天麻	Gastrodia Elata Blume	76	薏苡仁	Semen Coicis
35	芍药	Herbaceous Peony	77	炒枣仁	Stir-baked Semen Ziziphi Spinosae
36	败酱草	Herba Patriniae	78	黄连	The rhizome of Chinese goldthread
37	炒枳实	Immaturus	79	枸杞子	The Fruit of Chinese Wolfberry
38	川芎	Ligusticum Wallichii	80	葛根	The Root of Kudzu Vine
39	甘草	Liquorice	81	丹参	The Root of Red-Rooted Salvia
40	淡竹叶	Lophatherum gracile	82	石决明	The Shell of Abalone or Sea-Ear
41	磁石	Magnetite	83	五味子	The Fruit of Chinese Magnoliavine
42	麦芽	Malt			

we propose a TCM prescription generation task and a novel model to solve it. In TCM theory, the most salient principles are holism and treatment based on syndrome differentiation, which means TCM regards all viscera as a harmonious, highly related, circular unity, and the effective treatments are made by finding the real cause through different clinical

TABLE 10. The partial symptom-index check list.

Id	Chinese Name	English Name
1	腹胀纳差	Abdominal distension and anorexia
2	健忘	Amnesia
3	恶寒	Aversion to Cold
4	身体疼痛	Body Aches
5	多梦	Dreaminess
6	头晕	Dizziness
7	眼睛干涩	Dry eyes
8	脉沉细	Deep thready pulse
9	大便干燥	Dry stool
8	乏力	Exhaustion
9	食纳可	Eat a normal diet
10	发烧	Fever
11	痛苦面容	Facies dolores
12	舌体胖质淡边尖黯红	Fat, pale and dark red tongue
13	神疲乏力	Fatigued spirit and lack of strength
14	头痛	Headache
15	四肢困重	Heaviness of arms and legs
16	时作头痛	Headache frequently
17	失眠	Insomnia
18	失眠多梦	Insomnia and dreamful sleep
19	畏寒肢冷	Intolerance of cold and cold limbs
20	心慌心悸	Nervousness and palpitation
21	心悸	Palpitation
22	舌质淡红	Pale red tongue
23	舌淡胖	Plump pale tongue
24	舌红少苔	Red tongue with little coating
25	舌尖红	Red tip of tongue
26	舌红苔黄干	Red tongue and yellow and dry coating on the tongue
27	脉沉滑	Slippery and sink pulse
28	自汗	Spontaneous perspiration
29	脉弦细	Small and wiry pulse
30	耳鸣	Tinnitus
31	脉弦数	Taut rapid pulse
32	脉细数	Thready rapid pulse
33	苔薄白	Thin white tongue coating
34	心烦	Vexation
35	心烦易怒	Vexation and irritability
36	苔稍白腻/苔白微腻	Whitish greasy coating on the tongue
37	苔中白	Whitening of the middle part of tongue coating

symptoms. For instance, when people get cold, some symptoms may be appeared such as *Fever*, *Headache*, *Body Aches* and *Aversion to Cold*. The syndrome may be an *Anemofrigid Cold* or an *Anemopyretic Cold*, depending on the personal fitness and cause of illness. TCM practitioner will use different herbs to treat the superficially same symptoms because of the different syndromes. Thus, finding the real cause is a significant precondition for making an effective prescription.

TABLE 11. The partial TCM concepts and items-index check list.

Id	Chinese Name	English Name and/or Explanation
1	阴阳	**Yin-Yang.** The Yin-Yang theory is a guiding theory for medical practice, and is also one of the basic theories of TCM. Based on its concept, TCM holds that the relationship between Yin and Yang, a mutually opposite and interrelated relationship, runs through everything such as nature and human body, and is also the root and law of the occurrence, development and change of human physiology and pathology. The application of the Yin-Yang theory in TCM is mainly embodied in: explaining human body's organization, physiological activities, and pathological changes and guiding the diagnosis of the disease [34].
2	气	**Qi**. The theory of Qi in TCM is a systematic theory that studies the concept, generation, distribution, function of Qi in human body and its relationship with viscera, body essence, blood and body fluid. TCM holds that Qi is a very vigorous and running subtle substance in the human body, and is one of the basic substances to constitute and maintain human life activities. Qi is constantly running, promoting and regulating the metabolism of the human body, maintaining the life process of the human body [35].
3	君臣佐使	**Traditional Chinese medicine formulation (also called "Jun Chen Zuo Shi")**. Herbs in a prescription can be distributed into 'Jun' medicine, 'Chen' medicine, 'Zuo' medicine, 'Shi' medicine. 'Jun' (Principal herbs) treats the main cause of the disease. 'Chen' (assistant herbs) enhances the effects of 'Jun' and treats accompanying symptoms. 'Zuo' (complement herbs) reduces or eliminates possible toxic effects of the 'Jun' and 'Chen' herbs and treats accompanying symptoms. 'Shi' (guide and mediator herbs) helps to deliver or guide the above three kinds of herbs to the target organs, mediating effects of all above kinds of herbs [36].
4	整体论	**Holism.** Traditional Chinese medicine holds that the human body is an organic whole, and the various components of the human body are inseparable in structure, coordinated in function, complementary to each other, and affected each other in pathology. Therefore, local pathological changes and global pathological reactions should be unified in the view of TCM that local pathological changes will cause global pathological changes. Local pathological changes, often with the whole body viscera, qi and blood, the rise and fall of yin and yang, by viscera, tissues and organs in physiology, pathology and mutual influence [37].
5	辨证论治	**Treatment based on syndrome differentiation.** Doctors use various medical methods to analyse the nature of the disease, causes, patient's situation and make a correct judgement. Then, the treatments are made according to differentiation of symptoms and signs analysis results [38].
6	风寒感冒	Anemofrigid Cold
7	风热感冒	Anemopyretic Cold

The TCM practitioner's inquiry and the process of prescribing reflect these principles such as symptoms collection, syndrome differentiation, treatment determination, and prescription generation, which are the abilities we try to enable our model to possess.

Besides, there are also some important principles about the compatibility between herbs like "Jun Chen Zuo Shi". It is the fundamental rule for constructing a prescription. In addition, we also have noticed that the different symptoms occur chronologically. When insomnia occurs to a patient, for example, the headache, delirious, fidget may follow. Although not all symptoms have a strict temporal order, they can be divided into several small parts that symptoms in each part may have such strong temporal relationship. Thus, when TCM practitioner composes a prescription, they need to consider the effect, assistance and rivalry relations of herbs according to patient's symptoms instead of simply listing a set of herbs. These complex principles about pathogeny diagnosis, herbs selection and partial temporal order among symptoms may be the major reasons why there are few studies on the prescription generation.

In our work, we propose a two-stage attention mechanism to model these principles as described above. Our main-subordinate attention is proposed to differentiate the main symptoms from the subordinate ones. It tells the encoder which are the main symptoms and keeps more main symptoms information through the EncLSTM. According to the main-subordinate attention, our model will differentiate the real syndrome that leads to the main symptoms in the upcoming process, which can mimic the principle of treatment based on syndrome differentiation. The way that EncLSTM and main-subordinate attention encode symptoms also naturally consider the partial temporal order among symptoms. Before each herb is produced, our dynamic herb attentions are computed to pay different degrees on symptoms. The GenLSTM takes the herb attention v, main-subordinate context vector p as input to generate herbs one by one according to different symptoms. Different herbs are working on different symptoms. In most of our testing samples, the later herbs added to the prescription play a small role in relieving symptoms, which can simulate the principle of compatibility between herbs.

The proposed model regards the prescription generation as a translation from 'symptom language' to 'herb language' in appearance and models the TCM principles intrinsically. Our experiments show that our model can learn the interactions between herbs and symptoms and adjust the predicted prescription based on these learned interactions. The proposed model contains certain creativity. It is noteworthy that the proposed model is not only restricted in TCM but a universal idea for any kind of treatment problem with which the treatment processes are similar. We hope that this work can be a fundamental baseline for future researches and attract more researchers to pay much attention to TCM.

There are still some challenges in TCM prescription generation, such as the lack of training samples and effective evaluation criteria for the predicted prescription. In our future work, we will concentrate on finding a proper evaluation method that can assess the predicted generation's effectiveness, novelty and diversity. Meanwhile, we will collect more formal digital TCM clinical medical cases to enlarge our dataset. The herb's dosage modeling is also a necessary and challenging work because the predicted prescription is incomplete without the corresponding dosage. The TCM knowledge is also a useful context information during the generating prescription. However, there is rarely open-source, formalized or structured TCM knowledge base. It is time-consuming to collect and construct a useful knowledge base. We will study how to use the TCM knowledge in prescription generation in our future work.

APPENDIX

Since most of the herb and symptom items used in this paper are more common in Chinese and in order to make the concepts and items about traditional Chinese medicine more clear, we try our best to provide the necessary explanations and several tables for explaining the concepts and items, helping the upcoming researches in intelligent TCM and interested researchers who are interested in to reproduce our work. Part of the check list of symptoms and herbs, TCM concepts are listed in Table 9, Table 10, Table 11 respectively.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their constructive advice. They are also grateful for the hard works of the doctors such as Haojian Guo, Jingkuan Lan, Haiyan Jiang from Puji Outpatient Department of Traditional Chinese Medicine in Liaoning, China, who help us to estimate the generated prescriptions.

REFERENCES

[1] A. Helfenstein and P. Tammela, "Analyzing user-generated online content for drug discovery: Development and use of MedCrawler," *Bioinformatics*, vol. 33, no. 8, pp. 1205–1209, 2017.

[2] M. Li, S. Wang, G. Xie, X. Ma, T. Chen, and W. Jia, "polyPK: An R package for pharmacokinetic analysis of multi-component drugs using a metabolomics approach," *Bioinformatics*, vol. 34, no. 10, pp. 1792–1794, 2018.

[3] S.-Y. Pan, S.-F. Zhou, S.-H. Gao, Z.-L. Yu, S.-F. Zhang, M.-K. Tang, J.-N. Sun, D.-L. Ma, Y.-F. Han, W.-F. Fong, and K.-M. Ko, "New perspectives on how to discover drugs from herbal medicines: CAM's outstanding contribution to modern therapeutics," *Evidence-Based Complementary Alternative Med.*, vol. 2013, Jan. 2013, Art. no. 627375.

[4] N. R. Farnsworth, "Screening plants for new medicines," in *Biodiversity*. [Online]. Available: http://www.ciesin.org/docs/002-256c/002-256c.html

[5] X. Li, F. Li, Y. Wang, P. Qian, and X. Zheng, "Computer-aided disease diagnosis system in TCM based on facial image analysis," *Int. J. Funct. Inform. Personalised Med.*, vol. 2, no. 3, pp. 303–314, 2009.

[6] X. Zhang, J. Wang, B. Liang, H. Qi, Y. Zhao, J. Lu, L. Xu, X. Deng, X. Li, L. Wang, X. Tan, Y. Mao, G. Zhang, J. Wang, X. Li, and Y. Wang, "Mining the prescription-symptom regularity of TCM for HIV/AIDS based on complex network," in *Proc. IEEE Symp. Comput. Intell. Big Data (CIBD)*, Orlando, FL, USA, Dec. 2014, pp. 1–8.

[7] S. K. Poon, J. Poon, M. McGrane, X. Zhou, P. Kwan, R. Zhang, and D. M.-Y. Sze, "A novel approach in discovering significant interactions from TCM patient prescription data," *Int. J. Data Mining Bioinf.*, vol. 5, no. 4, pp. 353–368, Jan. 2011.

[8] W. Chen and S. He, "Application of data mining technology in TCM diagnosis and treatment," in *Proc. IEEE Int. Conf. Bioinformatics Biomed. (BIBM)*, Nov. 2017, pp. 1350–1353.

[9] G.-P. Liu, J.-J. Yan, Y.-Q. Wang, W. Zheng, T. Zhong, X. Lu, and P. Qian, "Deep learning based syndrome diagnosis of chronic gastritis," *Comput. Math. Methods Med.*, vol. 2014, no. 3, 2014, Art. no. 938350.

[10] Y. Peng, C. Tang, G. Chen, J. Xie, and C. Wang, "Multi-label learning by exploiting label correlations for TCM diagnosing Parkinson's disease," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Nov. 2017, pp. 590–594.

[11] D. Xie, W. Pei, W. Zhu, and X. Li, "Traditional Chinese medicine prescription mining based on abstract text," in *Proc. IEEE 19th Int. Conf. e-Health Netw., Appl. Services*, Oct. 2017, pp. 1–5.

[12] S. Wang, E. W. Huang, R. Zhang, X. Zhang, B. Liu, X. Zhou, and C. Zhai, "A conditional probabilistic model for joint analysis of symptoms, diseases, and herbs in traditional Chinese medicine patient records," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2016, pp. 411–418.

[13] C. Ruan, Y. Wang, Y. Zhang, J. Ma, H. Chen, U. Aickelin, S. Zhu, and T. Zhang, "THCluster: Herb supplements categorization for precision traditional Chinese medicine," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Nov. 2017, pp. 417–424.

[14] E. W. Huang, B. Liu, S. Wang, X. Zhou, R. Zhang, and C. Zhai, "PaReCat: Patient record subcategorization for precision traditional chinese medicine," in *Proc. 7th ACM Conf. Bioinf., Comput. Biol., Health Inform.*, 2016, pp. 443–452.

[15] C. Ruan, Y. Wang, Y. Zhang, and Y. Yang, "Exploring regularity in traditional Chinese medicine clinical data using heterogeneous weighted networks embedding," in *Proc. Int. Conf. Database Syst. Adv. Appl.*, 2019, pp. 310–313.

[16] L. Yao, Y. Zhang, B. Wei, W. Zhang, and Z. Jin, "A topic modeling approach for traditional Chinese medicine prescriptions," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 6, pp. 1007–1021, Jun. 2018.

[17] Y. Hu, G. Wen, H. Liao, C. Wang, D. Dai, and Z. Yu, "Automatic construction of chinese herbal prescriptions from tongue images using CNNs and auxiliary latent therapy topics," *IEEE Trans. Cybern.*, to be published.

[18] W. Li, Z. Yang, and X. Sun, "Exploration on generating traditional chinese medicine prescription from symptoms with an end-to-end method," 2018, *arXiv:1801.09030*. [Online]. Available: https://arxiv.org/abs/1801.09030

[19] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006.

[20] A. Karpathy and L. Fei-Fei, "Deep visual-semantic alignments for generating image descriptions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3128–3137.
[21] X. Zhang, S. Huang, J. Zhao, X. Du, and F. He, "Exploring deep recurrent convolution neural networks for subjectivity classification," *IEEE Access*, vol. 7, pp. 347–357, 2019.
[22] M. Sabokrou, M. Fayyaz, M. Fathy, Z. Moayed, and R. Klette, "Deep-anomaly: Fully convolutional neural network for fast anomaly detection in crowded scenes," *Comput. Vis. Image Understand.*, vol. 172, pp. 88–97, Jul. 2018.
[23] R. Sennrich, O. Firat, K. Cho, A. Birch, B. Haddow, J. Hitschler, and M. Nadejde, "Nematus: A toolkit for neural machine translation," in *Proc. 15th Conf. Eur. Chapter Assoc. Comput. Linguistics, Softw. Demonstration*, 2017, pp. 65–68.
[24] S. Liu, S. Zhang, X. Zhang, and H. Wang, "R-trans: RNN transformer network for Chinese machine reading comprehension," *IEEE Access*, vol. 7, pp. 27736–27745, 2019.
[25] X. Guo, H. Zhang, H. Yang, L. Xu, and Z. Ye, "A single attention-based combination of CNN and RNN for relation classification," *IEEE Access*, vol. 7, pp. 12467–12475, 2019.
[26] P. Nema, M. Khapra, A. Laha, and B. Ravindran, "Diversity driven attention model for query-based abstractive summarization," in *Proc. 55th Annu. Meeting Assoc. Comput. Linguistics*, 2017, pp. 1063–1072.
[27] A. Elbasir, B. Moovarkumudalvan, K. Kunji, P. R. Kolatkar, R. Mall, and H. Bensmail, "DeepCrystal: A deep learning framework for sequence-based protein crystallization prediction," *Bioinformatics*, vol. 35, no. 13, pp. 2216–2225, 2019.
[28] P. Anderson, X. He, C. Buehler, D. Teney, M. Johnson, S. Gould, and L. Zhang, "Bottom-up and top-down attention for image captioning and visual question answering," 2017, *arXiv:1707.07998*. [Online]. Available: https://arxiv.org/abs/1707.07998
[29] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," 2014, *arXiv:1406.1078*. [Online]. Available: https://arxiv.org/abs/1406.1078
[30] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," 2014, *arXiv:1409.0473*. [Online]. Available: https://arxiv.org/abs/1409.0473
[31] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[32] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "BLEU: A method for automatic evaluation of machine translation," in *Proc. 40th Annu. Meeting Assoc. Comput. Linguistics*, Jul. 2002, pp. 311–318.
[33] National Administration of Traditional Chinese Medicine. [Online]. Available: https://www.dayi.org.cn/
[34] X. N. F. Qianjin and L. Yaming, "Correlation of symmetry theory between Yinyang and physiology," Tech. Rep., 2003.
[35] *Yin Nature and Yang Nature of Qi and Theory of Qi in Traditional Chinese Medicine*, 2001.
[36] "Discussion on concept of 'the Monarch and his subjects, assistant and envoy' and applied principle from documents," *Chin. J. Basic Med. Traditional Chin. Med.*, vol. 10, no. 5, pp. 63–65, 2004.
[37] C. L. Z. Jusheng, "Philosophical origins of traditional chinese medicine's holism and western medicine' S reductionism," Tech. Rep., 1999.
[38] "Formula to syndrome being the soul of Chinese medical treatment principles," *Shanghai J. Traditional Chin. Med.*, vol. 40, no. 8, pp. 52–54, 2006.

ZHI LIU received the B.S. degree in software engineering from Changchun University, Changchun, China, in 2014, and the M.S. degree from the School of Computer Science and Engineering, Northeastern University, Shenyang, China, in 2017. He is currently pursuing the Ph.D. degree with the Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang. His current research interests include machine learning, deep learning, intelligent medical, natural language processing, and computer vision.

ZEYU ZHENG received the B.S. degree in mechanical engineering from Zhejiang University, Zhejiang, China, in 1997, and the Ph.D. degree from The Graduate University for Advanced Studies, Japan, in 2005. He is currently a Professor with the Institute of Automation, Chinese Academy of Sciences, Shenyang, China. He has authored nearly 50 technical articles in journals and conference proceedings. His research interests include intelligent optimization algorithm, big data processing technology, data mining, project management, and complex systems.

XIWANG GUO received the B.S. degree in computer science and technology from the Shenyang Institute of Engineering, Shenyang, China, in 2006, the M.S. degree in aeronautics and astronautics manufacturing engineering from Shenyang Aerospace University, Shenyang, in 2009, and the Ph.D. degree in system engineering from Northeastern University, Shenyang, in 2015. From 2016 to 2018, he was a Visiting Scholar with the Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, NJ, USA. He is currently an Associate Professor with the College of Computer and Communication Engineering, Liaoning Shihua University. He has authored over 30 technical articles in journals and conference proceedings, including the IEEE Transactions on Cybernetics, IEEE Transactions on Systems, Man, and Cybernetics: Systems, IEEE Transactions on Intelligent Transportation Systems, and IEEE/CAA Journal of Automatica Sinica. His current research interests include petri nets, remanufacturing, recycling and reuse of automotive, and intelligent optimization algorithm.

LIANG QI (S'16–M'18) received the B.S. degree in information and computing science and the M.S. degree in computer software and theory from the Shandong University of Science and Technology, Qingdao, China, in 2009 and 2012, respectively, and the Ph.D. degree in computer software and theory from Tongji University, Shanghai, China, in 2017. From 2015 to 2017, he was a Visiting Student with the Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, NJ, USA. He is currently with the Shandong University of Science and Technology. He has (co)authored over 45 articles in journals and conference proceedings, including the IEEE Transactions on Intelligent Transportation Systems, IEEE/CAA Journal of Automatica Sinica, IEEE Transactions on Systems, Man, and Cybernetics: Systems, IEEE Transactions on Computational Social Systems, IEEE Transactions on Automation Science and Engineering, and IEEE Transactions on Cybernetics. His interests include petri nets, machine learning, optimization, and intelligent transportation systems. He received the Best Student Paper Award-Finalist in the 15th IEEE International Conference on Networking, Sensing and Control (ICNSC'2018).

JUN GUI received the B.S. degree in industrial automation from North China Electric Power University, Baoding, China, in 2000, the M.S. degree in sensor system technology from the Karlsruhe University of Applied Sciences, Karlsruhe, Germany, in 2004. He is currently pursuing the Ph.D. degree in data analysis and mining with the Shenyang Institute of Automation, Chinese Academy of Sciences. He is also an Assistant Professor with the Information Science and Engineering School, Shenyang University of Technology. He has authored over ten articles in journals and conference proceedings. His current research interests include data analysis, data mining, knowledge discovery, and intelligent systems.

DIANZHENG FU received the B.S. degree in environmental engineering from the Dalian University of Technology, Dalian, China, in 2009, and the Ph.D. degree in energy and environmental engineering from North China Electric Power University, Beijing, China, in 2015. He is currently an Associate Research Fellow with the Digital Factory Department, Shenyang Institute of Automation, Chinese Academy of Sciences. He has authored over 30 technical articles in journals and conference proceedings, including *Energy Conversion and Management*, *Energy*, *Stochastic Environmental Research and Risk Assessment*, *Environmental Engineering and Management Journal*, and *Water Resources Management*. His current research interests include system analysis, system optimization, and data analysis techniques. He is also a Reviewer of a series of journals, including *Applied Energy*, *Journal of the Energy Institute*, *Energy*, *Energy Conversion and Management*, and *Applied Sciences-Basel*.

QINGFENG YAO received the B.S. degree from the University of Science and Technology Liaoning of Engineering, Anshan, China, in 2017. He is currently pursuing the master's degree with the Shenyang Institute of Automation, Chinese Academy of Sciences. His current research interests include neural networks and reinforcement learning.

LUYAO JIN is currently pursuing the B.S. degree with the School of Computer Science and Engineering, Northeastern University, Shenyang, China. She got Honorable Mention at the Maths Contest in Modeling, Consortium for Maths and Applications, in April 2017.