Chinese Medicine Prescription Recommendation using Generative Adversarial Network

XUEYAN LI¹, CHUITIAN RONG¹,², XUEMEI SUN¹ and HUABO SUN³

¹School of Computer Science and Technology, TIANGONG UNIVERSITY, Tianjin, China
²Tianjin Key Laboratory of Autonomous Intelligence Technology and Systems, Tianjin, China
³Institute of Aviation Safety, China Academy of Civil Aviation Science and Technology

Corresponding author: Chuitian Rong (e-mail: chuitian@tiangong.edu.cn).

ABSTRACT The theory of traditional Chinese medicine (TCM) is an important part of Chinese culture. In the long history, there are a large number of excellent prescriptions, whose laws have been explored by many studies, but few works directly studied the generation of prescriptions. With the rapid development of deep learning, many applications of text generation using neural networks have emerged. Prescriptions are the doctors' clinical experience and the results of neural networks also come from the accumulated experience. So, it is very feasible to apply deep learning techniques to the recommendation on TCM prescriptions. GAN and its variants have been applied in text generation recently. It has advantages in many aspects, such as the rapid speed of computation, the update of parameters by back propagation without Markov chain and more real data generation with two-players game. We attempted to know the important attributes of prescriptions and use these contents as the training data for variants of GAN to generate prescriptions. Specifically, we attempted to apply SeqGAN (Sequence Generative Adversarial Nets) and CGAN (Conditional Generative Adversarial Nets) to prescription generations. By underlying the knowledge of TCM, the prescriptions with different characteristics can be successfully generated. In the experiments, we conducted the comparative evaluations on the original data with other models. The results showed that applications in the innovation of prescription sequence generations have certain feasibility and significance, even can provide some reference values for the innovation of TCM.

INDEX TERMS TCM, Prescription recommendation, GAN, SeqGAN, CGAN

I. INTRODUCTION

TCM, also known as Han Medicine, is formed by exploration of Chinese for thousands of years. It is a discipline mainly studying human physiology, pathology, diagnosis and diseases. It also contains the valuable experience, carries theoretical knowledge and makes great contributions to people’s medical health. On the other hand, it is of great significance to inheritance of culture. For instance, the precious classics generated in this process are still of great reference value in modern society, such as Inner Canon of the Yellow Emperor (the most authoritative text of early medical theory and drug therapy), Sheng Nong’s herbal classic, Treatise on Cold Pathogenic and Miscellaneous Diseases.

TCM takes Yin-Yang and the Five Elements as the theoretical basis and research diseases through four diagnostic methods. It cures humans by prescriptions, food therapy, acupuncture or moxibustion. Li-Fa-Fang-Yao is the basic theory of TCM, which illustrates the pathogenesis of the disease, the principles and methods of treatment, appropriate prescriptions and the best choice of compatibility of herbs, respectively. Figure 1 explains how to generate prescriptions in that process.

In recent years, there have been many researches about TCM, including using data mining, topic model [19] or

FIGURE 1: Process of Li-Fa-Fang-Yao in TCM
other techniques to explore the compositions of prescriptions. While, few studies did creative works on generations of new TCM prescriptions by using deep model.

GAN [5] has been one of the most popular model and greatly developed since it was proposed. As shown in Figure 2, it was composed of two basic neural networks, namely Generator($G$) and Discriminator($D$). $G$ generates synthetic data based on a given noise. $D$ discriminates the fake data generated by $G$ from the real data. In the training process, $G$ tries to generate data more similarly, $D$ tries its best to distinguish them accurately.

GAN has a good performance in high efficiency of computation and it does not need Markov chain. The update of $G$'s parameters comes from $D$'s back propagation, which greatly promotes the efficiency. On the other hand, it is a minmax two-player game, which can generate more real data. Not only can GAN be applied in image generation, but also in natural language processing. So, we proposed the idea of applying GAN and its variants to generate TCM prescription. Figure 3 briefly shows the process how to generate new prescriptions data by using deep model.

In this study, we applied variants of GAN to make recommendations on prescription generation and our work is the first to explore the deep learning technique on Chinese medicine prescriptions generation. Based on the analysis and comparisons on the neural network and GANs in section II, we decided to apply SeqGAN [20] to generate new prescriptions directly in Section III. In order to generate prescriptions according to specific requirements, we applied CGAN [14] to generate prescriptions with specific functions in Section III. In order to validate the efficiency our methods, we have conducted experiments using the published Chinese medicine prescriptions in Section IV. We also compared the experimental results with the LSTM (Long Short-Term Memory) and VAE (Variational Auto-Encoder). The experimental results are evaluated by word2vec, BLEU and professional major in TCM. Finally, we draw the conclusions that the variants of GAN can generate more real and abundant prescriptions than LSTM and VAE, and the results of CGAN are more specific than SeqGAN. This research has a great significance on the generation of new prescriptions, new herb pairs, and the innovation of traditional Chinese medicine.

II. RELATED WORKS

A. GAN

Deep learning is a new branch in the field of machine learning, which includes convolutional neural network, autoencoder network, deep belief network, and so on. In recent years, they have been widely used in natural language processing, at the same time, many people put forward more abundant models on the basis of neural networks, such as GAN and its variants, which also promoted the development of deep learning.

GAN [5] is one of the deep learning models, which was proposed by Ian J. Goodfellow in 2014 to solve the problem of probability calculation in maximum likelihood estimation. The internal principle of the model is demonstrated by two behaviors, one is counterfeit generating fake currency and the other is trying to identify the counterfeit. They represented two players of GAN: $G$ and $D$. $G$ is trying to improve the technology of making counterfeit banknotes and $D$ is trying to identify counterfeit banknotes. The author used GAN to generate pictures. Take MNIST as example, $G$ uses a mixture of rectifier linear activations and sigmoid activations. $D$ uses maxout activations.

SeqGAN [20] is an improvement on the basis of GAN to generate text sequence, in which $G$ is composed by LSTM or GRU and $D$ is composed by CNN. The RL reward signal comes from $D$ judged on a complete sequence, and is passed back to the intermediate state-action steps using Monte Carlo search. The model is also pre-trained by authors and the experiment of music generation also be conducted.

CGAN [14] is also improved on the basis of GAN to generate data conditionally. The constraint $y$ is added to make the result of $G$ more objective. The input of $G$ is the random noise combined with $y$ in joint hidden representation. After a series of hidden ReLu layers’ mapping and maxout layers’ activation, $G$ is finally sent to sigmoid unit layers to generate data. $D$ also accepts the data with $y$. After maxout layers’ activation, both $G$ and $D$ are mapped to the joint maxout layers, and finally passed to sigmoid.

TextGAN [21] used LSTM as $G$ and CNN as $D$. In the training process, the thought of smooth approximation was also used to solve the differentiable gradient and feature matching was used to improve the convergence of the model. These changes can solve some problems but increase the training time and scale.

MaskGAN [4] was proposed to fill the missing word and used the algorithm of actor-critic to train $G$. But it not performed well in text generation and will be unstable with the increase of the sequence length. To avoid the problems of
mode collapse and training instability, the author selected the method of filling the missing word based on the text.

LeakGAN [6] was proposed to solve the generation of the long text. \( \mathcal{D} \) gave some features to \( \mathcal{G} \) and it will use this additional information to guide the generation of sequences.

**B. APPLICATION OF DEEP LEARNING IN MEDICINE**

The deep learning model is mainly applied in images currently, so the most application in medicine is to explore the molecular structure of drugs, which can be disassembled and combined.

Taking the research of Chen et al. [1] as an example, they pointed out that many applications of pharmaceutical researches have emerged in recent years, especially showing promise in addressing the discovery of drugs. The theories and characteristics of common neural networks were explained in this paper, including DNN [12], CNN [3], RNN [16], AE [9] and so on. Then, they introduced the applications of deep learning about medicine in bioactivity prediction, de novo molecular design, synthesis prediction and biological image analysis.

Lavecchia [10] analyzed some studies and applications of drugs in deep learning, providing a detailed view of the current state-of-the-art in drug discovery and highlighting not only the problematic issues, but also the successes and opportunities for further advances.

Jonathan [17] not only pointed out the current applications and advantages of deep learning in medicine, but also introduced other technologies used in medical researches, such as dropout, drop connect, etc. They also elaborated characteristics and limitations of other convolution networks.

**C. DEEP LEARNING IN TCM**

TCM has a long history and a large number of reference data. The inherent laws of Chinese herbs and prescriptions have been explored by many works. Liang Yao et al. [19] used PTM to study the internal laws of TCM, which is an extension to LinkLDA model. They pointed out that prescription is the product of experience, which can be combined with deep learning. They also introduced the basic theories of TCM, prior knowledge, herb pairs and so on.

Xintian Chen et al. [2] used heterogeneous information network in accurate clustering analysis. They also represented the relationship between the attributes of various parts of TCM with diagrams, which is similar to [7]. Qualitative evaluation and quantitative evaluation have been done. Both of them obtained good results.

Y Jin et al. [8] proposed the establishment of multi-graph convolution network. They also explained that the herb recommendation plays an important role in the generation of prescriptions. At the same time, the induction of syndrome in this process was elaborated, which is also significant to the generation of prescriptions.

Attention herb [11], an innovative model of prescription recommendation proposed by Z Liu, can not only learn the relationship between herbs and symptoms, but also can simulate the process of doctor’s inquiry and prescription generation.

Z Wang et al. [18] proposed a new TCM sequence generation architecture named TCM translator, which used seq2seq to explore the relationships between symptoms and herbs, and finally generated prescriptions through LSTM decoding.

**III. PRESCRIPTION GENERATION USING GAN**

GAN is an unsupervised learning model. It consists two components, as shown in Figure 2. One is \( \mathcal{G} \) for data generation, accepting a random noise \( z \) and outputting \( \mathcal{G}(z) \). Another is \( \mathcal{D} \) for data classification, accepting the real data \( x \) and the fake data \( \mathcal{G}(z) \) and trying to distinguish them.

During the training process, \( \mathcal{D} \) will feed back the results to \( \mathcal{G} \). The goal of training is to make the distribution of the generated samples consistent with the real data. As shown in Figure 4, \( \mathcal{G} \) and \( \mathcal{D} \) play minimax two-player game. For \( \mathcal{G} \), it tries to generate more real data and make \( \mathcal{D} \) cannot distinguish. \( \mathcal{G} \) maximizes the probability of making a mistake of \( \mathcal{D} \). For \( \mathcal{D} \), it tries its best to classify the two kinds of data. Finally, \( \mathcal{G} \) recovers the training data distribution and \( \mathcal{D} \) is equal to 0.5 everywhere. The principle of minimax two-player game as shown in the equation below:

\[
\min_{\mathcal{G}} \max_{\mathcal{D}} V(\mathcal{D}, \mathcal{G}) = \mathbb{E}_{z \sim p_{data}(z)}[\log \mathcal{D}(x)] + \mathbb{E}_{z \sim p_{z}(z)}[\log(1 - \log(D(z)))]
\]

**A. PRESCRIPTION GENERATION USING SEGANN**

Although GAN has a great development and generated many variants, most applications focused on image generation not discrete text. In the process of image generation, the gradient can be adjusted to generate images step by step. However, the structure of text is mostly discrete rather than continuous. So, it is difficult to use gradient policy update. As a result, the information from \( \mathcal{D} \) to \( \mathcal{G}(z) \) cannot be passed and the gradient cannot update. Furthermore, when GAN is used to generate text, its fine-tuning parameters may not work due to the discrete property of text data. Moreover, \( \mathcal{D} \) just evaluates the loss of whole generated sequence not the partial sequence. What’s more, the text is generally generated word by word, GAN can’t control the details of sequence generation.
SeqGAN [20] is an extension of GAN applied to sequence generation. Similar to GAN, SeqGAN is also composed by \( G \) and \( D \). \( G \) is used to generate sequence of words and \( D \) is used to distinguish whether the sentence is real or fake. \( G \) of SeqGAN is composed by LSTM or GRU applied with the reward of RL and Monte Carlo Search. As LSTM has the components for signal back propagation, it can read the sequence and its context. So, LSTM has advantages on processing sequence. In order to deal with the gradient update on discrete data, the reward of RL is applied. Monte Carlo Search is applied to complement non-complete sequences by evaluating the reward scores of all possible words. \( D \) is composed of CNN, which has good performance and efficiency in text classification, which can classify the whole sequence.

LSTM controls the state of transmit through kinds of gates and was showed in Figure 5. In the training process, it mainly consists of three stages, including forget, selective memory and output. The information needs to be remembered will be saved for a long time and unimportant information will be forgot.

The training process of SeqGAN is shown in Figure 6. Before the formal training, there is a pre-train process, \( G \) is pre-trained by maximum likelihood estimation trying to fit the distribution of true data. \( D \), receiving the positive data from real word and the negative data from \( G \), is trained as a binary classifier to distinguish them. Cross entropy is used as the objective function in above pre-training process. After a certain number of training loops, the formal training begins, in which \( G \) and \( D \) are trained alternatively. In the process of the sequence generation, Monte Carlo Search is used to generate kinds of results and can be evaluated whether it is good, and then pass the results to \( D \) to get evaluation scores. The Figure 7 shows how the complete sequence generated by Monte Carlo Search and the how to get the reward. Each circle on the right is a possible result of generating words, then the best policy will be selected according to the score and adjust the policy gradient. \( G \) can be trained by policy gradient, and in this process the probability of choosing the higher value of the reward will be increased and the lower will be decreased.

**FIGURE 5: Gates in LSTM**

![Gates in LSTM](image)

**FIGURE 6: Illustration of SeqGAN**

![Illustration of SeqGAN](image)

**FIGURE 7: Generation by MC Search and Evaluations of Rewards in SeqGAN**

![Generation by MC Search and Evaluations](image)

### B. PRESCRIPTION GENERATION USING CGAN

CGAN [14] is a generative adversarial network with conditional constraints. It can be regarded as a set of generators containing all kinds conditions. As shown in equation, the condition variable \( y \) is fed into \( G \) and \( D \), which can be labels, tags, data, or even pictures. Using this additional condition variable can guide \( G \) to generate specific data.

\[
\min_G \max_D V(D,G) =\mathbb{E}_{x \sim p_{data}(x)} [\log D(x|y)] + \mathbb{E}_{z \sim p_{z}(z)} [\log (1 - \log D(z|y))] \tag{2}
\]

The training process of CGAN is similar to that of GAN, except that the conditional classification label is added to the training data, as shown in the Figure 8. Noise \( z \) and conditions \( y \) are combined in a joint hidden representation in \( G \). True data \( x \) with condition \( y \) and fake data \( G(z) \) with condition \( y \) are both as the inputs of \( D \). Then, they will be embodied again by MLP [14]. Compared with other deep models, CGAN not only has the characteristics of GAN, such as can generate more real data, but also adds the conditional variable on the basis of GAN, which has more advantages on getting more targeted contents.

### C. EVALUATION METHODS

Referring the existing researches about evaluations of text generation, it can be found BLEU and word2vec are often used to calculate the results and similarities. In this condition, the results of our experiment were also made a comprehensive evaluation based on word2vec and BLEU. Besides, we also received judgements and suggestions from professionals major in TCM.

Word2vec [13] is often used in text similarity calculation. Contents of the generated text are inputted and segmented firstly. Then the word vectors of each word will be weighted average. The sentence is expressed as a fixed dimension vector. Finally, the cosine similarity between the generated
and the original data will be calculated. In our experiment, the generated sequence of results was compared with all the true prescriptions and then got the value of similarity.

BLEU [15] is a kind of method used to evaluate the results of text generation. It is still one of the most popular methods to calculate the value between the generated data and the reference composed of the original text. This value represents the accuracy between the generated text and the original text, which ranges from 0.0 to 1.0. If two sentences perfectly match, the score is 1.0, otherwise, the score is 0.0.

In addition to the above two main methods, we also consulted relevant professionals studied in Beijing University of Chinese Medicine, major in TCM. They evaluated how well the generated prescription results performed in our own experiment and gave the judgements and suggestions from the following aspects: feasibility, practicability, security and their own perspectives.

IV. EXPERIMENTS AND EVALUATION

We conducted experiments on dataset collected from Internet and compared results with other popular methods of sequence generation. We developed and trained SeqGAN and CGAN using PyTorch and TensorFlow respectively. SeqGAN is trained on prescription data and used to generate herb sequences as new prescriptions. CGAN is trained on prescription data with their functions as the labels (or conditions). The trained CGAN can generate new prescriptions for specific functions (or conditions). In order to evaluate the generated prescription, we used word2vec and BLEU to evaluate the results. The professionals in TCM were also involved to make their independent evaluations. The following is the details of experiments and the evaluations of results.

A. TRAINING DATA

The data is collected from Sheng Nong’s herbal classic, Canon of the Yellow Emperor 1, Chinese pharmacopoeia2, and so on. Chinese pharmacopoeia is Pharmacopoeia of the People’s Republic of China. It is an authoritative reference compiled by Chinese Pharmacopoeia Commission and also the standard of drugs in China. Sheng Nong’s herbal classic and Inner Canon of the Yellow Emperor are the classics of TCM in ancient. All the data is collected using the Scrapy. Scrapy is a web capture framework for Python, which is used to extract structured data from pages. The framework of it is composed of scrapy engine, scheduler, downloader, spider, item pipeline, downloader middlewares and spider

middlewares. The collected data are extracted and cleaned based on the knowledge of TCM theories.

TCM theories, such as Li-Fa-Fang-Yao, it means in the process of treatment, practitioners firstly focus on the symptoms of patients and the functions of herbs, then the treatment methods and prescriptions are determined. Guided by these classic theories, we collected data and then chose the compositions of prescriptions and functions as the training data. Finally, we obtained more than 50,000 prescriptions.

Medicinal Herb The attributes of medical herb mainly include name, nature, taste, meridian tropism, function and main treatment, etc. As shown in Figure 9, taking Notoginseng Radix et Rhizoma as an example, it tastes sweet and slightly bitter belonging to liver and stomach. The main function is to disperse blood stasis, stop bleeding, relieve swelling and sedation pain, which is often used for the treatment of hemoptysis, hematemesis, blood urination, etc.

Chinese Medicine Prescription The attributes of prescription mainly include herbs, dosage, functions, main treatment and attention, etc. As shown in Figure 10, Erding Keli is composed of Viola, Lobelia, Dandelion, Isatis Root, which has the effect of clearing heat and detoxification. It is commonly used for the treatment of carbuncle, sore throat. Because of the component including sucrose, which may bring about the elevation of glycemia, the diabetic patients should be careful.

B. USING SEQUENCE GAN TO GENERATE PRESCRIPTION SEQUENCES

For SeqGAN, we use LSTM as $G$ and CNN as $D$. $G$ is pre-trained with MLE at first, and $D$ is pre-trained with the data generated by $G$ and the training dataset. Then iterative con-
frontation training is start. In the training of $G$, the maximum length of the training dataset of prescriptions is set to 20 and the learning rate is set to 0.01. It uses Monte Carlo tree search to calculate the behavior value, and then update the policy gradient. In the training of $D$, take the training data as positive examples and the examples generated by $G$ as negative examples to train the discrimination model $D$. Dropout and L2 regularization are used to avoid over-fitting. The objective of optimization is to minimize the cross-loss entropy between the real data and the generated sequences. The hyper-parameters of $G$ and $D$ are often the same, not only the sequence length, but also the convolutional kernel size of $D$. Besides, we set the dimension of embedding as 32 and the batch size as 64. In the training process, we set LSTM network following the distribution $N(0,1)$. In order to prevent the over-fitting on the limited dataset, we trained $D$ and $G$ in a batch manner, such as training $D$ five times then training $G$ once. Once the training process is completed, new prescriptions are generated, then the results are evaluated by different methods.

### TABLE 1: Prescriptions Generated by SeqGAN

| Generated Prescriptions                                                                 | Word2vec | BLEU  | Feasibility |
|-----------------------------------------------------------------------------------------|----------|-------|-------------|
| Lonicerae Japonicae Flos, Moutan Cortex, Aconiti Lateralis Radix Praeparata,             | 0.68     | 0.87  | yes         |
| Frilltillae Cirrhosae Bulbus, Notoginseng Radix Et Rhizoma                              | 0.93     | 0.91  | yes         |
| Glycyrrhyzae Radix Et Rhizome, Angelicae Sinensis Radix, Corydalis Rhizome, Codonopsisradix | 0.94     | 0.89  | yes         |
| Scrophulariae Radix, Glycyrrhyzae Radix Et Rhizoma, Magnoliae Officinalis Cortex        | 0.97     | 0.87  | yes         |
| Achyranthis Bidentatae Radix, Cinnamomi Cortex, Bupleuri Radix, Glycyrrhyzae Radix Et Rhizome, Puerariae Lobatae Radix | 0.88     | 0.91  | yes         |
| Vitis Fructus, Xiongkuan, Dioscoreae Rhizoma, Glycyrrhyzae Radix Et Rhizome, Scrophulariae Radix | 0.84     | 0.87  | yes         |
| Curcumae Rhizome, Conyzae Herba, Nelumbinis Plumula,Picrohisae Rhizome, Angelicae Sinensis Radix, Poris, Glycyrrhyzae Radix Et Rhizoma | 0.97     | 0.87  | yes         |
| Vladimiri Radix, Ophiopogonis Radix, Glycyrrhyzae Radix Et Rhizome, Paoniae Radix Alba, Schisandrae Chinensis Fructus | 0.88     | 0.87  | yes         |
| Hupo, Cypri Rhizome, Ophiopogonis Radix, Armeniacae Semen Amarum                        | 0.71     | 0.88  | yes         |
| Forsythiae Fructus, Cinnamomi Flos, Rehmanniae Radix Praeparata, Polygoni Multiflori Radix, Scutellariae Radix, Paoniae Radix Alba, Paonieae Radix Rubra | 0.69     | 0.87  | yes         |
| Glycyrrhyzae Radix Et Rhizome, Mentheae Haplocalyxis Herba, Salviae Miltiorrhazae Radix Et Rhizome, Margarita, Ophiopogonis Radix, Radhosiae Rubescensis Herba | 0.87     | 0.87  | yes         |
| Coptidis Rhizome, Persicae Semen, Lycii Fructus, Armeiaae Radix, Atractylodis Macrocephalae Rhizome | 0.63     | 0.88  | yes         |

We selected some representative generated prescriptions from the outputs of SeqGAN, as shown in Table 1. It can be found that the generated prescriptions are with high similarities to the real prescriptions and obtain good evaluation judgements from TCM professionals. They commented that these herbs of each prescription have no conflicting content with each other from the perspective of TCM properties. In the other hand, every prescription is reasonable and conventional. The relatively good prescriptions shown in tables are listed in red color.

Taking the second result as an example, the prescription is composed of the following herbs, such as Scrophulariae Radix, Glycyrrhyzae Radix Et Rhizome, Magnoliae Officinalis Cortex. According to the evaluation results, the value of word2vec is 0.93 and BLEU is 0.91, which is reasonable and acceptable. On the other hand, the professional majoring in TCM comments that this prescription has the treatment on discomfort of throat, because of it not only can clear away heat, but also nourish yin and regulate qi. In fact, three herbs in this prescription all have effect on treatment of throat.

### C. USING CONDITIONAL GAN TO GENERATE PRESCRIPTION SEQUENCES

For CGAN, we also used LSTM as $G$ and CNN as $D$. Different from SeqGAN, we added the labels (or conditions) to the training data during training process. In the training process, for $G$, each herb in the prescription was mapped to a vector. The maximum length of the sequence was set to 20, and the hidden layer dimension of the memory cell of LSTM was set to 32. The excessive gradient in back propagation was truncated to 5.0 and the learning rate was set to 0.01 based on the adjustment parameter results. One prescription was composed of numbers sequence through an embedding layer to generate an encoding matrix. In this way we can get the sequence of every prescription. Then they will be input into each cell in LSTM. Combined with a fully connected layer, the probability of each herb will be output finally. For $D$, the generated data and real data are both inputs. We set the labels of real prescriptions as 1 and the generated as 0. On the other hand, all the sequence was concat with the functions. We used one-hot to edit the functions, if one prescription has some function, it will be marked 1 at the corresponding positions, others will be marked as 0. Finally, these data will be input and the value of loss will be calculated by the cross-entropy. The model will be adjusted consistently and get more real data. By doing this, it can generate more specific prescriptions. Because of the discreteness of prescription sequences, we incorporated the labels (one function or multiple functions) of prescription into training sequences. So, the trained CGAN can generate prescriptions for specific conditions. The results of experiment are greatly influenced by the details of training, so we adjusted the training strategy. In this process, $G$ and $D$ were trained by different times to avoid over-fitting, usually $k$ steps of $D$ and one step of $G$. The training process can be can be split into two steps. Firstly, the random noise was sent to $G$, which will imitate the distribution of the real prescriptions. Then, the sample of the generated prescriptions and the real prescriptions were
TABLE 2: Prescriptions Generated by CGAN with Multiple Function Labels

| Functions                                      | Generated Prescriptions                                                                 | Word2vec | BLEU | Feasibility |
|------------------------------------------------|-----------------------------------------------------------------------------------------|----------|------|-------------|
| lowering blood pressure, remit angina pectoris | Blelliae Rhizome, Eucommiae Cortex, Gardeniae Fructus, Ariaenamis Rhiizoma Preparatum  | 0.71     | 0.88 | yes         |
| improve myocardial ischemia, remit angina pectoris | Magnettium, Coicis Semen, Lablab Semen Album                                            | 0.78     | 0.87 | no          |
| stasis-eliminating, relieve pain, throat-clearing, detumescence, rhizoma alismatis | Lonicerae Japomae Flos, Sinomenii Caulis, Pinelliae Rhiizoma Praeparatum, Cistanches Herba, Cum Zinghhere Et Alumine | 0.74     | 0.92 | yes         |
| clearing away heat, relieve pain, suppressing cough, detumescence, radix sieris, warming yang | Ecliptica Herba, Nigella Semen, Ginger Liquid Extract                                   | 0.91     | 0.93 | yes         |
| nourishing yan, relieve pain, detumescence, remove obstruction toit, regulate the flow of vital energy | Xanthii Fructus, Amomi Fructus, Scutellariae Radix, L-Menthol                           | 0.77     | 0.91 | yes         |
| arousing consciousness, regulate the flow of vital energy, remove obstruction toit, relax the muscles | Paonieae Radix Rubra, Angelicae Sinensis Radix, Artemisiae Argyi Folium, Zchnittao, Armeniaceae Semen Amarum, Ginseng Radix Et Rhiizome, Bombyx Bataricus, Uncariae Ramulus Cum Uncis | 0.77     | 0.93 | yes         |
| clearing away heat, suppressing cough, relieve internal heat or fever | Codonopissradix, Forsythiae Fructus, Aconis Radix Cocta, Lonicerae Flos, Acoti Tatarinovii Rhiizome, Angelicae Sinensis Radix, Salviae Militiorrhazae Radix Et Rhiizome, Schizonepetae Herba, Poralae Fruits | 0.80     | 0.92 | yes         |

TABLE 3: Prescriptions Generated by CGAN with One Function Label

| Function                                      | Generated Prescriptions                                                                 | Word2vec | BLEU | Feasibility |
|------------------------------------------------|-----------------------------------------------------------------------------------------|----------|------|-------------|
| relieving pain                                | Angelicae Dahuricae Radix, Notogyeryg Rhizomum Et Radix, Tribuli Fructus, Draconis Sangus, Bupleuri Radix, Pinelliae Rhiizoma Praeparatum Cum Zinghhere Et Alumine, Gingseng Radix Et Rhiizome, Isatidis Folium, Mori Cortex | 0.87     | 0.91 | yes         |
| suppressing cough                             | Ophiopogonis Radix, Ephedra Herba, Scrophulariae Radix, Rehmanniae Radix Praeparata, Paroniae Radix Alba | 0.76     | 0.91 | yes         |
| nourishing yan                                | Lonicerae Flos, Bupleuri Radix, Alismatis Rhiizoma, Carthami Flos, Fritillariae Thunbergii Bulbus, Taxilli Herb | 0.77     | 0.90 | yes         |
| relieving pain                                | Saloniae Militiorrhazae Radix Et Rhiizome, Angelicae Sinensis Radix, Lycici Fructus, Polygonati Rhiizome, Atractylodis Radix, Daturaeflos, Tzchosanbhis Radix, Eucommiae Cortex, Glycyrrhizae Radix Et Rhiizome | 0.78     | 0.87 | yes         |
| relieving pain                                | Armeniaceae Semen Amarum, Sesami Semen Nigrum, Paroniae Radix Alba, Cistanches Herba, Borneodium, Citri Reticulatae Pergcupium, Glycyrrhizae Radix Et Rhiizoma | 0.85     | 0.88 | yes         |
| suppressing cough                             | Taxilli Herba, Scutellariae Radix, Ginseng Radix Et Rhiizoma                            | 0.99     | 0.92 | no          |
| nourishing yan                                | Sim Sulin Pusrix, Bupacae Fructus, Borneodium                                           | 0.72     | 0.87 | yes         |
| suppressing cough                             | Astragali Radix Praeparata Cum Melle, Atractylodis Macrocephalae Rhiizome, Codonopissradix, Chuanxiong Rhiizome, Poria | 0.98     | 0.91 | yes         |
| nourishing yan                                | Margariferi Concha, Chebulae Fructus, Peucedani Radix, Glycyrrhizae Radix Et Rhiizome Praeparata Cum Melle, Coptidis Rhiizoma | 0.85     | 0.87 | yes         |
| relieving pain                                | Sim Felis Pusrix, Ariaenamis Cum Brie, Artemisae Xciropiae Herba, Angelicae Dahuricae Radix, Fritillariae Thunbergii Bulbus | 0.82     | 0.91 | yes         |
| suppressing cough                             | Angelicae Sinensis Radix, Jujubae Fructus, Cinamomum Mullis, Arretinaceae Semen Amarum, Atractylodis Macrocephalae Rhiizome, Zingiberis Rhiizoma Recens, Citri Fructus, Sappoalaoe Rhiizome, Anemathrenae Rhiizome, Uoscoreae Rhiizome | 0.94     | 0.90 | yes         |

sent to \( D \) for classification whether it is true or not, which was relatively difficult and need to be trained more frequently. So, we did the experiment on the different case of \( k \) values. Based on the results, when \( k=1 \) the training ability of \( D \) was always weaker than \( G \) and the results were not satisfying as the evaluation score is lower. When \( k=8 \), the value of loss began with 5.2 and ended with 0.88, which was apparently higher than others. The variation trend of loss for \( k=3 \) was similar to that of \( k=5 \). However, when \( k=5 \) the training process performed more smoothly and the value of loss converged faster to 0.83, which was the best value among them. So, we set \( k = 5 \), which means training \( D \) five times and then \( G \) once.

We conducted two different experiments by adding different number of labels to the training prescriptions. As shown in Table 2 and Table 3 respectively, the results of two kinds of training both have good performance on different evaluations. The relatively good prescriptions shown in table are also listed in red color.

Firstly, CGAN is trained on the prescriptions data accompanied with the multiple functions as labels. Table 2 shows the generated prescriptions for multiple functions. From the perspective of the professional, the effects of generated prescription are relevant with the all labels of function. Each generated prescription has various functions, which are consistent with the content of labels. Taking the second one as an example, the therapeutic effects of labels include improving myocardial ischemia and remitting angina pectoris. The generated prescription contains the herbs of magnetitum, coicis semen and lablab semen albelum, whose score of word2vec and BLEU are 0.78 and 0.87, respectively. From the calculation results, it means the generated prescription is acceptable with high probability. From the perspective of professionals, it has the effect on remitting angina pectoris.
but little effect on improving myocardial ischemia. Taking the last one as an example, the generated prescription has the corresponding functions referred in the training labels, which has a good effect on clearing away heat and suppressing cough and certain effect on relieving internal heat or fever.

Secondly, CGAN is trained on the prescriptions data accompanied with the one of its function as the label. Table 3 shows the generated prescription for one function. Taking relieving pain as an example, one of the possible results contains the herbs of Tiexie and Ecliptae Herba, whose score of word2vec and BLEU are 0.92 and 0.91 respectively. It means the generated prescriptions are reasonable with high probabilities. However, from the perspective of professionals, the function of Ecliptae Herba is to nourish yin, cool blood and stop bleeding, whose main effect is not relieving pain. So, it cannot be simply regarded as a painkiller. Another possible result contains the herbs of Lonicerae Flos, Bupleuri Radix, Carthami Flos, Fritillariae Thunbergii Bulbus, Taxilli Herba, Alismatis Rhizoma, which has a good effect on relieving pain and also clearing away heat.

D. COMPARISON WITH LSTM AND VAE

We also conducted experiments to make comparisons with existing methods, such as LSTM and VAE. LSTM is invented to solve the problem of short-term memory. It has an internal mechanism called gates, which can guide the information flow. VAE is a deep generative neural network, which can learn the data distributions and generate the data. Both of them have been used in sequence generation and also produced some good results.

**Compared with LSTM**

For LSTM, the training data was transformed into one-hot encoding. This model was built with only one layer LSTM. Softmax was used as the full connection layer to output the probability of all words. Then the prescriptions can be generated by 30 epochs training. The Table 4 showed the results generated by LSTM. The first herb in two generated prescriptions using different methods is same. It is obviously observed that the sequence length generated by LSTM is shorter than that generated by SeqGAN. On the other hand, the characteristics of generated prescriptions are not consistent with training data. But the evaluation scores look reasonable, some of them even get 0.92. We guess the reason is that most of results are herb pairs, which is the commonly used herbal collocation.

**Compared with VAE**

For VAE, it is a deep generative neural network. It can also learn the distributions of the data and generate the sequences. However, VAE gets the mean and variance from true data, then extracts samples from the normal distribution. It adapts to learn the data features from its distributions and may cause minor deviations. The GANs are suitable for learning the data distributions of the real data. The above finds have been proved in image generation. The images generated by VAE maybe blurred. This experiment using VAE for prescription generation are also not satisfied, as shown in Table 5. The first herb of generated prescriptions by two methods also are same. Take the first result in Table 5 as example, there is no correlations among the generated herbs. Farfarae Flos is used to cough. Litsea cubeba is used to anorexia. Moghania philippinensis is used to disperse stasis and detoxification. Aconitum kusnezoffii Reichb is used to relieve pain. Melantebrium is used to hemostasis. All the generated herb did not have similar functions and acted on different organs. By comparing the results of the two models, we found that GAN can imitate and learn the distribution of data better than VAE. This is also the reason why the sequence generation of GAN is better than VAE.

To sum up, the results of variants of GAN had a better performance than LSTM and VAE, which can be reflected on the length and evaluation values. The generation results of SeqGAN are more complete and more reasonable compared with the results of LSTM and VAE. Although the evaluation values of SeqGAN and CGAN are similar, the results of CGAN with labels are more objective and targeted than those generated by SeqGAN, and more consistent with the real prescriptions. What's more, referring to the judgements of professionals comprehensively, it is essential to be guided by doctors in clinical practice.

E. JUDGEMENTS OF PROFESSIONALS

The results of this experiment has feasibility and safety to some extent. In clinical practice, the results of this work need to be enhanced. In terms of generated prescriptions, it is feasible to use one sequence directly or combine several of sequences. Actually, in practical treatment, the prescription is generated according to disease pathogenesis, which is mostly a combination of more than twenty kinds of traditional Chinese herbs. So, partial of the generated results are not consistent with the clinical treatment. But at present, the clinical prescription cannot only consist plenty of herbs, but need to be simplified. For example, about 50 percent prescriptions in classics are composed by less than four kinds, and 90 percent are composed by less than eight. In the generation process of TCM prescriptions, it is very feasible to refine the content of prescriptions theoretically through the deep model, which can be further studied.

On the other hand, the concept of toxicity of TCM is elaborated in many works. Chinese pharmacopoeia classifies the toxicity into major toxicity, toxicity and minor toxicity. Under the guidance of the theory of TCM and the reasonable usage, the safety of the above prescriptions is generally well. However, due to individual differences, we should not be too confirm in terms of safety.

V. CONCLUSION AND FUTURE WORK

By investigating the relevant literature, we summarized the existing works and the knowledge about TCM. In order to explore the underlying knowledge of existing prescriptions, we utilized the deep neural network for prescription recommendation. We applied SeqGAN to generate prescriptions directly. In order to generate prescriptions for specific requirements, we applied CGAN to generate prescriptions conditionally. We conducted experiments on real world dataset and
TABLE 4: Comparison of Prescriptions Generated by LSTM and SeqGAN

| GENERATED PRESCRIPTIONS BY LSTM | WORD2VEC | BLEU | FEASIBILITY | GENERATED PRESCRIPTIONS BY SEQGAN | WORD2VEC | BLEU | FEASIBILITY |
|--------------------------------|----------|------|-------------|----------------------------------|----------|------|-------------|
| Asparagi Radix, Angelicae Dahuricae Radix | 0.92 | 0.90 | yes | Asparagi Radix, Bupleuri Radix, Scutellariae Radix, Atractylodis Macrocephalae Rhizoma, Glycyrrhizae Radix Et Rhizoma | 0.85 | 0.86 | yes |
| Glycyrrhizae Radix Et Rhizoma, Citri Reticulatae Pericarpium, Angelicae Dahuricae Radix, Alstonii Radix Et Rhizoma, Notoginseng Radix Et Rhizoma | 0.75 | 0.76 | yes | Glycyrrhizae Radix Et Rhizoma, Nelumbinis Semen, Picrohexine Radix, Angelicae Sinensis Radix, Poria | 0.81 | 0.82 | yes |
| Sojae Semen Germinatum, Cypri Rhizoma, Isatidis Radix | 0.97 | 0.91 | no | Isatidis Radix, Angelicae Sinensis Radix, Forsythiae Fructus, Astragali Radix, Lonicerae Japonicae Flos | 0.91 | 0.93 | yes |
| Notoginseng Radix Et Rhizoma, Borneolum Syntheticum | 0.97 | 0.88 | yes | Atractylodis Macrocephalae Radix, Asparagi Radix, Bupleuri Radix, Scutellariae Radix, Glycyrrhizae Radix Et Rhizoma | 0.85 | 0.86 | yes |
| Cinnamomi Cortex, Bupleuri Radix, Angelicae Dahuricae Radix | 0.89 | 0.88 | yes | Cinnamomi Cortex, Bupleuri Radix, Glycyrrhizae Radix Et Rhizoma, Paerariae Lobatae Radix | 0.81 | 0.79 | no |

TABLE 5: Comparison of Prescriptions Generated by VAE and SeqGAN

| GENERATED PRESCRIPTIONS BY VAE | WORD2VEC | BLEU | FEASIBILITY | GENERATED PRESCRIPTIONS BY SEQGAN | WORD2VEC | BLEU | FEASIBILITY |
|--------------------------------|----------|------|-------------|----------------------------------|----------|------|-------------|
| Farfarae Flos, Litsea cubeba, Moghania philippinensis, Aconitum kusnezoffii Reichb, Melantheritum | 0.78 | 0.73 | no | Farfarae Flos, Angelicae Sinensis Radix, ConYZae Herba, Alli Sativ Bulbus, Suis Felis Pulvis, Lycii Fructus | 0.87 | 0.88 | yes |
| Nelumbinis Semen, Foeniculum vulgare Mill, Natrii Sulph, Bergenia purpurascens | 0.79 | 0.81 | no | Nelumbinis Semen, Picrohexine Radix, Angelicae Sinensis Radix, Poria, Glycyrrhizae Radix Et Rhizoma | 0.81 | 0.82 | yes |
| Vladimiriae Radix, Asteris Radix Et Rhizoma | 0.91 | 0.89 | yes | Vladimiriae Radix, Ophiopogonis Radix, Glycyrrhizae Radix Et Rhizoma, Schisandrae Chinensis Fructus | 0.86 | 0.84 | yes |
| Paroamia Radix Alba, Carchami Flos, Olibanum, L-Menthol | 0.81 | 0.83 | no | Paroamia Radix Alba, Forsythiae Fructus, Rehmanniae Radix Praeparata, Polygoni Multiflori Radix, Scutellariae Radix, Carthami Flos, Paroamiae Radix Rubra | 0.79 | 0.81 | yes |
| Herba Polygoni Chinensis, Commiphora myrrha, Placenta Hominis | 0.83 | 0.87 | yes | Herba Polygoni Chinensis, Menthae Haplocalycis Herba | 0.87 | 0.88 | yes |

The results show our proposed methods can generate more reasonable and more practical prescriptions than compared models. In practice, there are still many factors to be considered, such as toxicity, the effect of treatment and the dose of herbs in one prescription, and so on. For the future research, we will incorporate more theory of TCM into the neural network training process. Also, the neural network model also can be further improved, for example, attention mechanism and controllable text generation techniques can be utilized.

VI. ACKNOWLEDGMENT
This work was supported by the project of Natural Science Foundation of China (No.61402329, No.61972456) and the Natural Science Foundation of Tianjin (No.19JCYBJC15400).

REFERENCES
[1] Chen, H., Engkvist, O., Wang, Y., Olivecrona, M., Blaschke, T.: The rise of deep learning in drug discovery. Drug Discovery Today 23(6), 1241–1250 (2018)
[2] Chen, X., Ruan, C., Zhang, Y., Chen, H.: Heterogeneous information network based clustering for precision traditional chinese medicine. BMC Medical Informatics Decis. Mak. 19(S6):264 (2019)
[3] Duvenaud, D., Maclaurin, D., Aguilera-Iparraguirre, J., Gomez-Bombarelli, R., Hirzel, T., Aspuru-Guzik, A., Adams, R.P.: Convolutional networks on graphs for learning molecular fingerprints. In: Cortes, C.,
XUEYAN LI is currently pursuing the Master’s degree in TIANGONG UNIVERSITY. Her research interests include data mining and big data analysis.

CHUITIAN RONG is an associate professor at TIANGONG UNIVERSITY. His research interests mainly include database system, intelligent information retrieval, big data analysis.

HUABO SUN, born in 1980, Ph.D., researcher. His main research focuses on big data analysis of civil aviation safety.