Promoting Appropriate Medication Use Leveraging Medical Big Data

Longbiao CHEN\textsuperscript{1}

\textsuperscript{1}Affiliation not available

October 28, 2022

Abstract

Inappropriate medication use has become an important factor affecting the safety of rational medication. Most traditional medical anomaly detection systems are based on rules to regulate inappropriate medication use. In this paper, we model the complex relationships among patients, diseases, and medicine based on medical big data to promote appropriate medication use. More specifically, we first construct the medication knowledge graph based on the historical prescription big data of tertiary hospitals and medical text data. Second, based on the medication knowledge graph, we employ a Gaussian Mixture Model (GMM) to represent patients in groups as physiological features. For diagnostic features, we employ the pre-training word vector BERT to enhance the semantic representation between diagnoses. And to reduce adverse drug interaction caused by combination drug use, we employ a graph convolution network to transform drug interaction information into drug interaction features. Finally, we employ the sequence generation model to learn the complex relationship among patients, diseases, and medicine and provide an appropriate medication evaluation for prescribing by doctors in small hospitals from drug list and medication course of treatment. In this paper, we leverage the MIMIC-III dataset and the dataset of a tertiary hospital in Fujian Province to verify the validity of the model. The results show that our method is more effective than other baseline methods in the accuracy of medication regimen prediction of rational medication. In addition, it has achieved high accuracy in the appropriate medication detection of prescriptions in small hospitals.

Hosted file
draft-v10.tex available at https://authorea.com/users/518137/articles/592404-promoting-appropriate-medication-use-leveraging-medical-big-data
[Diagram of a neural network model with encoder and decoder blocks, showing the input and output layers.]

[Another diagram illustrating a multilayer perceptron with encoding and decoding stages and their corresponding features and probabilities.]
RESEARCH ARTICLE

Promoting Appropriate Medication Use Leveraging Medical Big Data

Shiwang HUANG, Xiaohai CAI, Zhiming LIN, Yunting SHAO, Longbiao CHEN, Chenhui YANG, Linghong HONG*
Fujian Key Laboratory of Sensing and Computing for Smart Cities, School of Information Science and Engineering, Xiamen University, Xiamen 361005, China

Abstract According to the statistics of the World Health Organization (WHO), inappropriate medication use has become an important factor affecting the safety of rational medication. In the gray area of medical insurance supervision, such as designated drugstores and designated medical institutions, there are lots of appropriate medication phenomena about "Supporting Doctor with Medicine". Most traditional medical anomaly detection systems are based on rules to regulate inappropriate medication use, and the audit approach is inflexible and not suitable for clinical environments that require intelligent auditing. In this paper, we model the complex relationships among patients, diseases, and medicine based on medical big data to promote appropriate medication use. More specifically, we first construct the medication knowledge graph based on the historical prescription big data of tertiary hospitals and medical text data. Second, based on the medication knowledge graph, we employ a Gaussian Mixture Model (GMM) to represent patients in groups as physiological features. For diagnostic features, we employ the pre-training word vector BERT to enhance the semantic representation between diagnoses. And to reduce adverse drug interaction caused by combination drug use, we employ a graph convolution network to transform drug interaction information into drug interaction features. Finally, we employ the sequence generation model to learn the complex relationship among patients, diseases, and medicine and provide an appropriate medication evaluation for prescribing by doctors in small hospitals from drug list and medication course of treatment. In this paper, we leverage the MIMIC_III dataset and the dataset of a tertiary hospital in Fujian Province to verify the validity of the model. The results show that our method is more effective than other baseline methods in the accuracy of medication regimen prediction of rational medication. In addition, it has achieved high accuracy in the appropriate medication detection of prescriptions in small hospitals.

Keywords rational use of drugs, appropriate medication, NLP, knowledge graph, transformer

1 Introduction

Rational use of medicines is safe, effective, affordable and appropriate to treat or cure the patient [1]. Irrational medication use is a major problem worldwide. World Health Organization (WHO) estimates that more than half of all medicines are prescribed, dispensed, or sold inappropriately, and that half of all patients fail to take them correctly [2]. In addition, in the gray area of medical insurance supervision, such as designated pharmacies and medical institutions, there may be "huge prescriptions for minor diseases" healthcare fraud [3]. The inappropriate drug use behaviors such as overuse, underuse, or misuse of medicines not only waste medical resources but also lead to significant patient harm in terms of medication errors (ME) and adverse drug events (ADE) [1]. WHO is committed to promoting rational use of medicines clinical physicians and pharmacists, so as to ensure that "patients receive the appropriate medicines, in doses that meet their own individual requirements, for an adequate period of time." [1].

One of the key challenges in rational medication is the
rational and appropriate use of medicines. Compared to the safety, effectiveness, and economics of rational medication, the evaluation of rational appropriate use of medicines is more complicated, involving hyper-medication, under-medication, and inappropriate medication.

To address these issues, hospitals assign experienced investigators to investigate Medicare fraud detection. However, this method becomes time-consuming and inefficient due to a large amount of data collection. Improvements in data mining and deep learning tools have led to a shift in attention to developing automated systems for fraud detection. Various deep learning-based medical anomaly detection systems (DADs) have been developed and deployed in hospitals to reduce the incidence of improper drug use. After analyzing a large number of medical records, the Diagnosis Related Groups (DRGs) payment system [4] based on disease type have been launched by The National Medical Insurance Administration (NMIA) to specify uniform drug delivery rules and prevent excessive medical treatment. However, the single and rigid pharmaceutical rules is not able to achieve more accurate personal medication, which also poses a major challenge to the promotion of DRGs [5]. Therefore, we need a more flexible and intelligent method for the evaluation of rational appropriate medication.

Fortunately, with the emergence of medical consortiums and the sinking of medical resources, the professional prescribing experience of tertiary hospitals can be accessed, providing us with new perspectives to address the problems existing in designated pharmacies and designated medical institutions. Therefore, we integrate the clinical medication experience of tertiary hospitals and medical knowledge, and transfer the learned knowledge to small hospitals and clinics, so that their prescriptions are more in line with professional standards. In order to achieve these goals, we need to address the following issues:

Due to the large individual differences in patients, such as children, adults, and the elderly, such as their liver and kidney functions, nervous system and other physiological characteristics, so the same diagnosis may lead to different treatment regimens. Therefore, in order to suit the remedy to the case with greater precision, we need to consider the individualized use of medicines.

Since the relationship between disease and symptoms is not a simple one-to-one relationship, the occurrence of a single disease may cause the simultaneous occurrence of multiple symptoms, so doctors must treat patients through the combination of multiple drugs. Similarly, the presence of two or more chronic or acute diseases at the same time in individual patients is becoming more common. Therefore, it is a challenge for us to address the complex relationship between disease and drug use.

The increase in drug species represents that the compatibility relationships between drugs are more complicated. In addition, consider that there would be more drug overuse and abuse in the case of "big prescription for minor ailments", while polypharmacy may increase drug side effects, even more adverse drug-drug interactions [6]. Therefore, drug-drug interaction should be taken into account when evaluating the appropriateness of rational drug use to reduce the adverse reactions of prescription combined drug use.

In order to address the above issues, in this paper, we propose a regulatory framework of rational drug use based on medical consortium and big data through mining the clinical experience of prescription big data and medical knowledge of drug instructions. To be specific, We first extract information from the big data of prescription of tertiary hospitals and medical text data, and establish the medication knowledge graph based on the extracted information. Second, based on medication knowledge graph, we extract physiological feature, diagnostic feature and drug interaction feature through feature enhancement. Finally, we construct the sequence generation model to solve the complex relationship among patient, disease and drug, then evaluate the appropriate medication of prescribing by doctors in small hospitals by using the model learned from tertiary hospital.

In conclusion, the contribution of this work is as follows:

1. To the best of our knowledge, this is the first work on a data-driven evaluation of rational appropriate medication use. By exploiting the knowledge in prescription big data from a tertiary hospital, and combining it with medical text knowledge, we are able to evaluate the medication appropriateness in clinical practice of small hospitals based on drug list and medication course of treatment.

2. We propose a data-driven experience extraction of clinical rational drug use and appropriate medication evaluation framework based on sequence generation model. First, we establish the medication knowledge graph based on the information extracted from the historical prescription big data of tertiary hospitals and medical text data. Second, to achieve accurate individualized medicine, we extract the patients’ physiological, diagnostic and drug interaction information respectively, and use the corresponding feature enhancement method to convert the three types of information into physiolog-
ical, diagnostic and drug interaction feature. Finally, we employ the sequence generation model to learn the complex relationship between people, diseases and drug, and use the model learned from tertiary hospital to provide an evaluation of the medication appropriateness of prescribing by doctors in small hospitals from the two aspects of drug list and medication course of treatment.

3. We evaluate the proposed framework with two medical record data sets: Medical Information Mart for Intensive Care III (MIMIC_III) and real-world prescription big data collected from tertiary hospitals. Results show that our method has a more accurate medication regimen prediction ability, and consistently outperforms other baselines. In addition, it has achieved high accuracy in the appropriate medication detection of prescriptions in small hospitals.

The remainder of this paper is organized as follows. We first survey the related work in Section 2, and then present the overview of the proposed framework in Section 3. In Section 4, we detail the medication knowledge graph construction. We describe the drug recommendation model based on the medication knowledge graph in Section 5. The results of the experiments are shown in Section 6. Finally, we conclude the paper in Section 7.

2 Related Works

In this section, we briefly survey the existing literature from the following two perspectives: (1) appropriateness of rational drug use, and (2) medical big data analytics.

2.1 Appropriateness of Rational Drug Use

Rational drug use has become an important topic in clinical practice. However, due to the widespread fraud of large prescriptions for minor ailments in the gray area of drug supervision, the phenomenon of unreasonable drug use is increasing, including over-prescribing and inadequate prescribing [7]. Improper use of drugs may cause serious consequences. In addition, inappropriate drug use will brings more risks of drug interactions [8]. Therefore, while emphasizing the safety, effectiveness and economy of rational medication, we should also pay attention to appropriate medication issues.

In order to make drug use more appropriate and rational, various drug-related clinical decision support systems have been proposed by researchers and applied in hospitals [9]. A simple implementation of the CDS system is to provide a complete medication guidelines and a list of parameters for each drug to doctors for reference [10]. The National Medical Insurance Administration also proposed the DRGs payment system, which is used to implement single-disease payment for unified drug use according to patients’ conditions [11]. A more intelligent approach to drug delivery would be to offer the physicians a specific medication recommendations tailored to the patient’s specific conditions [10]. A complete evaluation of clinical decision support systems for drug management can be found at [12].

At the heart of these CDS systems are various dosing rules, which are compiled from existing medicine information databases and documentation. However, due to the constant change of medication information because of ongoing clinical experience and research, it is difficult to ensure the timeliness and correctness of medication rules, bringing potential risks for patient medication safety [12, 13]. For example, nearly three-quarters of CDS alerts were overwritten in clinical practice, and 40% of overwrites are inappropriate, which is reported in [14]. And DRGs is essentially prescribing drugs according to uniform rules, which doesn’t take into account personal medication. To address these issues, in this paper, we propose to facilitate medication prediction directly from historical medication data and medical text data instead of using uniform medication information and rules.

2.2 Medical Big Data Analytics

With the advent of the big data society, the 4V characteristics of big data are well known: volume, variety, velocity, and veracity [15]. Big data includes large-scale medical, environmental, financial, geographic, and social media information and continues to grow [16]. In the medical and healthcare field, large volumes of data are generated, which present a broad prospect in clinical applications [17, 18]. According to statistics [19], the global volume of big data on health care is expected to double every two years, and by 2020, it will increase by 50 times compared with that of 2011. Medical big data analytics aims to exploit the knowledge in data to provide predictive modeling and clinical decision support, disease or safety surveillance for public healthcare, etc. [16]. Big data analytics in the medical frequently exploits analytic methods developed in data mining, including classification, clustering, and regression [15].

The advantages of big data technology and the experiences of clinicians should be integrated to provide reliable
diagnoses and conclusions [20]. For example, in the safety of rational medicines, Shao, Y. T. et al. [21] construct a probabilistic probability model of massive prescription data based on a knowledge graph to evaluate the risk of drug combination by the graph search algorithm. In rational use of medicines, Shang, Junyuan et al. [22] jointly model the longitudinal patient records as an EHR graph and drug knowledge base as a DDI graph, by generating sequence models which trained end-to-end to provide effective and safe medication recommendations. Based on the experimental results on real-world EHR, GAMENet outperformed all baselines in drug-drug interaction(DDI) rate reduction [22].

However, to the best of our knowledge, there are still few studies that work in big data analysis of appropriate drug use that combine knowledge graph and sequence generation model techniques. One of the relevant ideas to our work was to use big data in health care to make rational medication treatment predictions [22]. In that work, the authors propose the Graph Augmented Memory Networks (GAMENet) and trained end-to-end to provide a safe and personalized recommendation of medication combination. Another related idea of our work was data-migration [23], where the authors propose and evaluate a method for anomaly detection in healthcare in a real scenario based on a provider-consumer model. It assigns anomaly scores to the cities (consumers) as a function of their demand, and transfers the scores from cities to hospitals (providers). In this work, we use large-scale prescription data sets from tertiary hospitals to exploit the knowledge of medication in clinical practices, and integrate the medical knowledge in drug instructions to provide a rational evaluation of appropriate drug use for small hospitals and clinics.

3 Framework Overview

We propose a framework for experience extraction of clinical rational drug use and appropriate index evaluation, as illustrated in Figure 1. In the medication knowledge graph construction stage, we first extract drug triads based on the data source from historical prescriptions from Tertiary hospitals and auxiliary medical text data, then establish the medication knowledge graph of patient-disease-medicine as a relationship based on the extracted information. In the modeling phase, we first employ Gaussian Mixture Model (GMM) [24] to group patient population representation as physiological features, based on the four physiological information of gender, age, height, and weight. Second, we transform patients’ diagnostic information into word vectors as diagnostic features through pre-training word vector Bidirectional Encoder Representations from Transformers (Bert) [25] to enhance the semantic representation between
diagnoses. Third, in order to reduce adverse drug interactions caused by drug combinations, we employ a graph convolution network (GCN) [26] to transform drug interaction information into drug interaction features. Finally, we exploit the medication regimen from historical prescription data to train a sequence generation model. In the analysis stage, given a new prescription from small hospitals or clinics, we use the trained model to predict the rational medication regimen for the prescription and provide an evaluation of rational appropriate drug use in terms of drug list and medication course of treatment to physicians and pharmacists in small hospitals. We elaborate on the details of the key components in the following sections.

4 Medication Knowledge Graph Construction

In this section, our objective is to construct a medication knowledge graph to model medication rules for co-prescribed in big data. However, relying only on historical prescription data is not enough to simulate the comprehensive medication rules, because adverse drug reactions may not be reflected in clinical practice. Therefore, we also incorporate the drug interaction information extracted from the drug instructions as a supplement. First, due to the large amount of non-(semi-)structured data in historical prescription big data, we need to transform these data into structured triplet data. Second, we build the clinical experience edges and medical knowledge edges based on structured data of historical prescriptions and drug instruction. Finally, we construct our medication knowledge graph according to two kinds of edges. We elaborate on the details as follows.

4.1 Information Extraction

In this step, in order to extract drug entities and relationships, we modeled the problem as an information extraction task in Natural Language Processing (NLP) and solved it by information extraction technology. First, for historical prescription data, we transformed semi-structured disease-drug-diagnosis information into structured clinical triples to achieve a complete delineation of clinical experience. Then, for auxiliary medical text data, we extracted medical knowledge triples to supplement the medication knowledge graph. The information extract details are elaborated as follows.

4.1.1 Clinical Experience Extraction

In this step, we extracted clinical experience based on the collected prescription big data. The historical prescription data mainly includes the prescription number, the patient’s age, height, weight, and other personal signs, diagnosis of disease, drugs and their course of treatment, and other information. In order to better show clinical medication experience, we established explicit attributes of entities and implicit triple relationships between entities according to medication knowledge.

Specifically, we first extract different entities in the prescription, including patients, diseases, and drugs. Then, for the patient entity, We regard physiological characteristics such as gender, age, height, and weight as the attributes of the patient entity. And if there is a diagnosis on the prescription that is associated with a pregnant woman, such as at 14 weeks of gestation, the patient will be given the role of the pregnant woman. Furthermore, we construct implicit relationships between different entities based on prescriptions, such as the relationship between the patient and the drug, the relationship between the patient and the disease, and the relationship between the diagnosis and the disease. Finally, we iterate over each prescription and use a triple to represent all the entities in the prescription and their relationships, such as "Influenza - Prescribe - Ribavirin Spray", etc.

4.1.2 Medical Knowledge Extraction

In this step, we extract medical knowledge based on the collected data set of drug instructions. Since there is an implicit regional structure in each drug instruction, as shown on the left in Figure 2, we first divide a part of the collected drug instruction data into blocks to extract the required structured information such as drug name, main ingredi-
ents, indications, contraindications, adverse reactions, precautions, and drug interactions, etc. Second, we manually label the pre-processed dataset based on the open source labeling tool: YEDDA, as shown in Figure 2. And the labeled entity includes but is not limited to the drug names, diseases, ingredients, indications, adverse reactions, and contraindications, etc. Third, we also marked another dataset in the format of (text, entity, relationship, entity) on the module data of annotated notes and drug interaction to extract drug interactions. According to the harm degree of drug interaction to the human body, the relationship fields of drug interaction are divided into 4 categories, which are beneficial, no effect, unknown, and harmful. Finally, we model the medical knowledge extraction problem as named entity recognition and relation extraction tasks in natural language processing to extract medical triplet information, as shown in Figure 3.

Specifically, we first train the Bert-BiLSTM-CRF model to recognize medical entities, including drug, ingredient, disease, indication, and contraindication. The Bert-BiLSTM-CRF model was proved to outperformed all other models in NLP of Chinese electronic health documents [27]. Second, to extract the relationships between entities, such as drug interactions, we construct the relation extracting model(RE model): BertModel + Dropout + Linear. Finally, we employ the trained entity recognition model to extract medical entities. And for drug interaction data, we identify the relationships based on the extracted entities. There are two approaches to forming medical triplet data. The first method is to take the drug name and other entities as the first entity and the second entity respectively, and label as the relation, such as "Ribavirin spray - Ingredient - Ribavirin", etc. The second method is to use the triplet data extracted from RE model, such as "Cefoperazone sodium for injection - Contraindication - Amikacin", etc.

4.2 Graph Node and Edge Construction

In this step, in order to construct the medication knowledge graph $G = (E, R)$, we construct the clinical experience edge set $R_a$ and the medical knowledge edge set $R_b$, where $E$ represents the entities in the medication knowledge graph and $R$ represents the relations between entities. Figure 4 illustrates the structure of our medication knowledge graph, in which nodes represent the entities, including drugs, diseases, patients, and prescriptions, and the edges represent the relationships between nodes. In graph theory, the triple $Q$ is defined as the set of $(e_i, r, e_j)$, where $e_i$ and $e_j$ denote two different entities, and $r$ denotes the relation between node $e_i$ and node $e_j$. As shown in Figure 4, the black edge sets represent the clinical experience edges $R_a$, and the red edge sets represent the medical knowledge edges $R_b$. The construction details are elaborated as follows.

4.2.1 Clinical Experience Edges

We first a construct clinical experience edge sets $R_a$ based on the triples extracted from prescription big data. Specifically, the edge sets $R_a$ of graph $G$ are defined as follows: for the triples extracted from the prescription record, we set up two nodes and assign a directed edge between the corresponding nodes in graph $G$. The triples include "Patient-Have-Prescription", "Prescription-Diagnose-Disease", "Prescription-Prescribe-Drug", "Patient-Have-Disease" and "Patient-Use-Drug".
4.2.2 Medical Knowledge Edges

The second step is to construct medical knowledge edges $R_b$ based on the triples extracted from the drug instructions. Specifically, the edge sets $R_b$ of graph $G$ are defined as follows: for the triples extracted from drug instructions, we set up two nodes and assign a directed edge between the corresponding nodes in graph $G$. The triples include "Drug-Ingredient-Drug", "Drug-Indication-Disease", "Drug-Contraindication-Patient", "Drug-Contraindication-Disease", "Drug-Interaction-Drug" and "Drug-Contraindication-Drug".

5 Drug Recommendation Model Based On Knowledge Graph

In this section, our objective is to model the complex relationships among patients, diagnoses, and drugs based on the medication knowledge graph constructed in the previous phase. Since there are many prescription features in the medication knowledge graph, we first extract the features of patients, diagnoses, and drugs. Then we employ the sequential generation model to model the sequential decision-making process of drug regimens. We specify the specific work as follows.

5.1 Feature Extraction From Graph

In clinical practice, most pediatric medicines are dosed according to the patient’s age [28], body age, or body weight ($mg/kg$) [29]. Moreover, the treatment also varies depending on the patient’s symptoms and indications, therefore diagnostic information is helpful when developing medication regimens [30]. Since there are more combination drugs in the supernormal prescription drugs, which are more likely to cause adverse drug reactions (ADRs). Therefore, drug interactions should also be considered in the rational and appropriate use of drugs. Based upon this prior knowledge, we extract the corresponding physiological, diagnostic features, and drug interaction feature from the medical knowledge graph constructed in the previous phase. We elaborate on the details as follows.

5.1.1 Physiological Feature Extraction

In this step, our objective is to extract patients’ physiological information from historical prescription big data as the features of rational drug use. Physiology metrics of patients, such as sex, age group, body weight, and body height, are usually the most important considerations in clinical medication calculation. Although the combination of these factors area provides the greatest accuracy in calculating medication regiments, simple digital groupings of different age groups, heights or weights could lead to excessive discrete of the population sample. Therefore, we group patient populations as a physiological feature by modeling physiological information.

Due to differences in gender, height, weight, and other physiological characteristics, the distribution of prescription data may be composed of N Gauss. For example, patients with chronic gastritis. By plotting the distribution of patients of different ages in male and female genders, as shown in Figure 2, we can observe that there are roughly two component kernels in the distribution. The component kernel was contributed by male patients aged around 50 and female patients aged around 60 respectively. Therefore We can use a mixed Gaussian distribution to fit all patient prescription data. We use BMI (Body Mass Index, healthy or not) to represent height and weight for each prescription data. First, we estimate the optimal number of component cores $n$ for the patient’s prescription data by AIC [31] and BIC [32]. The number of component cores is optimal when AIC and BIC are as small as possible. Then we calculate the probability distri-
bution of the patient in each component kernel. Finally, we take the group with the highest probability as physiological characteristics, as shown in Figure 6. The probability density function of the mixed Gaussian distribution is as follows:

$$p(x) = \sum_{k=1}^{K} \pi_k N(x|\mu_k, \Sigma_k)$$  \hspace{1cm} (1)

Where $X$ is the age distribution of prescriptions. $K$ is the number of sub-Gaussian models in Gaussian mixture models. $\pi_k$ is the mixture coefficient, which is the probability that each observation data belongs to the $k$th submodel. The $N(x|\mu_k, \Sigma_k)$ is the distribution function. The $\mu_k$ is the expectation and the $\Sigma_k$ is the covariance of the $k$th component in the mixed model. And the above variables are satisfied by the following equation:

$$\sum_{k=1}^{K} \pi_k = 1, \hspace{1cm} (2)$$

$$0 < \pi_k < 1 \hspace{1cm} (3)$$

With the EM algorithm (Expectation-Maximization algorithm) [33], we can iteratively calculate the parameters in GMM: $(\pi_k, \mu_k, \Sigma_k)$. In short, the EM algorithm has two steps. The first step is E (expectation), which updates the implicit variable. The second step is M (maximization), which is used to update the parameters of each Gaussian distribution in GMM. Then the above two steps are repeated until the iteration termination condition is reached.

5.1.2 Diagnostic Feature Extraction

In this step, our objective is to extract diagnostic sequence information as diagnostic features of the patient. A simpler approach to represent diagnostic information is to digitize the diagnosis. But because the diagnostic text representation is rich in semantic information of words, such as text similarity. Therefore, considering that word vectors are rich in more semantic features, we choose to represent diagnostic features by transforming text into word vectors through BERT pre-trained word model [25], as shown in Figure 7. We elaborate on the details as follows.

When sending a diagnostic text into BERT, it will encode the text into an input vector whose length is always 512. For an input vector, it is composed of three embedding features: 1) WordPiece [34], 2) Position Embedding, 3) Segment Embedding. What’s more, as shown in Figure 7, its framework consists of the multi-layer transformers proposed in [35], transformers realize a series of encoding and decoding to transform an input text into a possible predicted result. Finally, the diagnostic text is converted to tokens by Bert, and a word vector at the corresponding position of each token is printed. We take out the results of the penultimate hidden layer and use the results of all vector mean pooling as diagnostic features.

5.1.3 Drug Interaction Feature Extraction

In this step, our objective is to establish drug interaction relationships from the medication knowledge graph as another feature. It can be found that the main form of drug interaction relation stored in the medication knowledge graph is pair, and it is more effective to use the graph structure to represent the drug interaction relation. Therefore, we used such a pair relationship to generate the drug interaction matrix $A$ to represent the drug interaction graph. However, the interaction diagram exists as a two-dimensional matrix, while the other two features exist as one-dimensional vectors. Therefore, we need to transform the characteristics of drug interaction so that it can be spliced with the other two characteristics as the input of a reasonable drug recommendation model. We elaborate on the details as follows.

The size of the drug interaction matrix is $N \times N$, where $N$ is the size of the drug set. For the MIMIC_III data set, we first select two drugs in the drug set randomly, assuming that the coded values are $i, j$. Then we map its ATC4 code to the CID classification. Finally, we used the CID code as the keyword to search the drug interaction database for adverse drug interaction risks in these two categories. We iterate through all drug pairs in the database to generate an adjacency matrix of the final drug interaction diagram, which reflects the currently known drug combination contraindications and can reduce the number of treatment options that produce adverse drug reactions when a drug is recommended. For the medical
record data of Fujian Province, we also randomly select two drugs in the drug set and then search in the medication knowledge graph with the keyword of drug composition for whether these two drugs have adverse drug interaction risks. If there is, the corresponding element of the marker matrix is 1, otherwise, it is 0. The simplest way to convert a two-dimensional matrix to a one-dimensional vector is to compress the matrix. However, such compression destroys the connection between nodes, making it impossible for the model to learn the complete drug interaction relationship.

Therefore, we employ the graph neural network method to embed the 2-D drug interaction characteristics into the 1-d space. The specific transformation process is described as follows:

Like the convolutional neural network in image vision, the Graph Convolution Network (GCN) [26] is used for feature extraction. Therefore, we use the basic graph convolution network to construct a simple graph neural network and map the graph node representation to the low-dimensional vector space while preserving the topology and node information of the graph. A graph neural network with two GCN layers is established in this paper, where A is the graph structure and X is the matrix representation of the graph. The GCN layer compresses the hidden representation of each node by aggregating the feature information from the node neighborhood, and after the feature aggregation, a nonlinear permutation such as ReLU is applied to the generated output. Through the stacking of multiple layers of GCN, the final hidden representation of each node in the diagram obtains information from subsequent neighborhoods. Finally, we connect it to a fully connected network to obtain one-dimensional vector output. Through the transformation of the graph neural network, obtain the one-dimensional representation of the characteristics of drug interactions.

5.2 Sequence Generation Model

In this step, our objective is to predict the rational medication regimen based on the extracted features. One intuitive method is to concatenate the physiology and indication features into a vector and build a regression or classification model to predict the rational medication regimen. However, due to the considerable variety of the two categories of features, such a direct concatenation of the two heterogeneous features does not perform well, especially when some features play a dominant role in specific medication conditions [36]. To address these challenges, we use the sequence generation model to transform the problem into a sequence decision process of drug regimen, including the medication list and treatment of drug use. We elaborate on the details as follows.

First, the symbols are defined, with $X$ representing the diagnostic space and $Y$ representing the drug treatment space. $R = \{(X_1,Y_1),(X_2,Y_2),...,(X_k,Y_k)\}$ is a set of prescription records, and $X_k \subseteq X$ is a diagnostic sequence, and $Y_k = \{X^1_k,X^2_k,...,X^n_k\}$, $Y_k \subseteq Y$ is a sequence of medication regimens, $Y_k = \{y^1_k,y^2_k,...,y^n_k\}$. $|X_k|$ and $|Y_k|$ are the sequence lengths of $X_k$ and $Y_k$, respectively. There is no explicit mapping of the corresponding elements between diagnostic sequence $X_k$ and drug sequence $Y_k$. In order to avoid confusion, if there is no ambiguity, we omit $k$ in the symbol.

The purpose of drug prediction is to select the best sequence of medication regimen $Y_k$ among all drugs $Y$ based on the diagnostic sequence $X$. Therefore, the model in this paper should have the ability to learn to map any diagnostic sequence to a corresponding medication regimen sequence, which requires the model in this paper to learn not only the relationship between drugs and diagnosis but also the relationship between drugs and drugs. Therefore we used the popular Transformer approach in natural language processing to generate drug sequences. We’ll take a brief look at one of the key mechanisms in the Transformer model and the overall architecture.

5.2.1 Attention Mechanism

The Transformer is based on the attention mechanism, so before introducing the overall framework of the Transformer model, we first introduce the role of the attention mechanism. The sequence generation model Transformer is composed of encoder modules and decoding modules. As shown in Figure 8(a), the encoder compresses the information expressed by the input sequence into a fixed-length semantic vector, and then the decoder decodes the information based on this semantic vector and generates the target sequence one by one. It means that elements at any point in the input sequence are equally important to the current target element. This method of learning input sequence can not express the position information of the sequence, and if the sequence is too long, the decoding effect of the fixed semantic vector will be poor, because it’s easy to lose the information contained in the sequence. In order to solve the above problems, Luong et al. proposed the attention mechanism in 2015 [37]. As shown in Figure 8(b), the mechanism generates an independent semantic vector for each element in the output sequence, which could express the different importance of each input.
sequence to the decoded target element. The semantic vector \( C_i \) is the weighted sum of the elements in the input sequence:

\[
C_i = \sum_{j=1}^{L_i} \alpha_{ij} h_j \tag{4}
\]

\[
\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{L_i} \exp(e_{ik})} \tag{5}
\]

\[
e_{ij} = a(s_{i-1}, h_j) \tag{6}
\]

Where \( L_i \) is the length of the input sequence, \( \alpha \) is the distribution of attention, \( \alpha_{ij} \) represents the importance of the element in the input sequence to the element that determines the output sequence, and \( h_j \) represents the implicit state of the \( j^{th} \) element in the input sequence. \( \alpha \) is a similarity measure, which is calculated according to the correlation between the \( j^{th} \) element in the input sequence and the \( i^{th} \) element in the output sequence. As shown in 6, \( e_{ij} \) is obtained from the output \( S_{i-1} \) of the hidden layer at the time of \( i - 1 \) in the decoder and the correlation degree of the hidden state \( h_j \) corresponding to each element in the encoder. There are many methods to calculate the similarity between \( s_{i-1} \) and \( h_j \). In this paper, we adopt the dot product method as shown in 7 to calculate the similarity.

\[
a(h_i, \overline{h}_s) = h_i^T \overline{h}_s \tag{7}
\]

The above attention mechanism focuses on the relationship between the input sequence and output sequence, which can help us obtain more information when we model the relationship between drugs and prescriptions. However, the relationship between elements in the input sequence and output sequence is not taken into account. In order to solve this problem, another new attention mechanism self-attention [38] has been proposed and widely used. Compared with the use of various cyclic neural networks that require longer information accumulation, the model with the introduction of the self-attention mechanism can not only obtain the dependency relationship between two elements that are far apart but also obtain the dependency relationship between the internal elements of the sequence more easily. In addition, the self-attention mechanism also improves the parallel computing capability of the model, greatly reducing the training time of the model. The Transformer model used in this paper is also based on the self-attention mechanism.

In the self-attention mechanism, the input sequence will be represented in the form of key-value pairs, and then the input sequence with \( N \) elements will be represented as \((K, V) = [(k_1, v_1), (k_2, v_2), ..., (k_N, v_N)]\), Where the key value is used to compute the attention distribution \( \alpha \) and the value value is used to compute the semantic vector. The output sequence will be represented as \( N \) queries, so the semantic vector computation problem can be considered as an addressing operation. We use query to find the \( \text{key} = \text{query} \) element in the input sequence, and the value obtained is the semantic vector or called attention. In particular, the self-attention mechanism can be thought of as soft addressing. Instead of looking only for elements whose key value is equal to the query value, a weighted sum is applied to all values to calculate the final attention. The weight of each value is determined by calculating the similarity between the query and each key. Therefore, the formula for calculating attention is shown as follows.

\[
Attention(Q, K, V) = \text{softmax}(\frac{QK^T}{\sqrt{d_k}})V \tag{8}
\]

Where \( Q \in \mathbb{R}^{m \times d_q} \), \( K \in \mathbb{R}^{m \times d_k} \), \( V \in \mathbb{R}^{m \times d_v} \), so Attention is a \( n \times d_v \) matrix. Different from the general attention mechanism, the self-attention mechanism also performs a division operation when calculating the attention distribution coefficient, in order to avoid the similarity value calculated by the inner product method being too large, which results in 0 or 1 is generated when the softmax function is used for normalization, losing the meaning of soft addressing.

5.2.2 Overall Framework of Transformer

The Transformer model is a machine translation model proposed by Google in 2017 [35]. As shown in Figure 9, the Transformer model is mainly composed of Encoder modules and Decoder modules. Each module is composed of several Encoder layers and Decoder layers, and the number of stacked layers can be adjusted according to the difficulty of the tasks. This model abandons the traditional RNN CNN architecture and adopts the structure of a full self-attention stack, which achieves excellent performance in natural language processing tasks. In this paper, We build a
two-layer Transformer model to carry out the task of rational drug recommendation.

There are three kinds of attention in Transformer, namely, self-attention in Encoder, self-attention in Decoder, and attention between Encoder and Decoder. In order to capture all the spatial information in the input sequence, the attention calculation method in Transformer is improved on the basis of the previous introduction by introducing the concept of Multi-Head, which mainly projects Query, Key, and Value to different spaces $h$ times by linear transformation, and then obtains $h$ self-attention matrices by calculation. Since the feed-forward layer can only receive one matrix, we finally spliced $h$ matrices and multiplied by a weight matrix to generate the final attention matrix. When self-attention between input and output sequences is calculated, set $Q = K = V$, and the specific calculation process is shown as follows:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, ..., \text{head}_h)W^0$$ (9)

$$\text{head}_i = \text{Attention}(QW^i, KW^i, VW^i)$$ (10)

As we can see from Figure 9, the Decoder structure is similar to Encoder, with the difference that Decoder has two attention layers. The Decoder’s goal in decoding is to give a probability distribution for the first output element. Therefore, we need to calculate the attention value of the last Decoder layer in the Decoder module. The input to this attention layer consists of two parts, the output of the last Encoder module and the output of the first Decoder module in the Decoder module. Therefore, in calculating the attention of the Decoder layer, the value $K, V$ in 9 comes from Encoder and $Q$ from Decoder. Decoder decoding is different from Encoder in that it can compute in parallel because it needs to use the output of the previous Decoder layer as a query, so the Decoder is used one by one to generate the elements of the output sequence. In the model training process, the Decoder layer uses real values, so the mask method should be used to calculate self-attention between output sequences to ensure that the current model can not get more information than the current location.

Based on the Transformer model, as shown in Figure 9, we join the features together as the input sequence. And after a multistep process of encoding and decoding, we select the element with the highest probability in the probability distribution as the prediction result for the current position, until the end of the generated identifier or reached. The resulting sequence is used as the model’s recommendation for the current patient. Next, we will verify the effectiveness of the Transformer model in the prediction of rational drug use through several comparative experiments.

### 6 Evaluation

In this section, we evaluate our method with medical record data set collected from MIMIC_III dataset and real-world, anonymized prescription big data collected from tertiary hospitals. We first introduce the experiment settings and then present the evaluation results. Finally, we display our analysis results on the visualization platform.

#### 6.1 Experiment Settings

##### 6.1.1 Dataset

After data cleansing, we obtain a dataset containing 1,084,594 prescriptions with 23,225 patients, 2,393 medicines, and 5,591 diagnoses from MIMIC_III database, and another dataset containing 230,390 prescriptions with 19,146 patients, 3,782 diagnoses, and 1,198 medicines from a tertiary hospital. The summary of the dataset is shown in Table 1.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>MIMIC_III</th>
<th>FUJIAN</th>
</tr>
</thead>
<tbody>
<tr>
<td># Prescriptions</td>
<td>1,084,594</td>
<td>230,390</td>
</tr>
<tr>
<td># Medicines</td>
<td>2,393</td>
<td>1,198</td>
</tr>
<tr>
<td># Diagnoses</td>
<td>5,591</td>
<td>3,782</td>
</tr>
<tr>
<td># Patients</td>
<td>23,225</td>
<td>19,146</td>
</tr>
</tbody>
</table>

#### 6.1.2 Evaluation Plan

Firstly, we randomly select 80% of the prescriptions collect from the constructed medication knowledge graph for...
training and the left 20% for evaluation. Then we collected 100 problematic prescriptions with inappropriate drug use from small hospitals as a test dataset to evaluate our trained model. Specifically, for each prescription, we use our model to classify whether these prescriptions are of inappropriate use of drugs. We evaluated our model by measuring the proportion of correctly classified prescriptions in terms of both medication sequence list and medication treatment and using rational medication regimens to represent the predicted results of both.

6.1.3 Evaluation Metrics

To measure the accuracy of the proposed model, we used Jaccard Similarity Score (Jaccard), Precision, Recall, and Average F1 (F1). Jaccard is defined as the size of the intersection divided by the size of the union of ground truth medication regimen $Y_t^{(k)}$ and predicted medication regimen $\hat{Y}_t^{(k)}$:

$$
Jaccard = \frac{1}{\sum_k N} \sum_k \frac{|Y_t^{(k)} \cap \hat{Y}_t^{(k)}|}{|Y_t^{(k)} \cup \hat{Y}_t^{(k)}|} \tag{11}
$$

where $N$ is the number of patients in test set.

$$
Precision = \frac{1}{\sum_k N} \sum_k \frac{|Y_t^{(k)} \cap \hat{Y}_t^{(k)}|}{|\hat{Y}_t^{(k)}|} \tag{12}
$$

$$
Recall = \frac{1}{\sum_k N} \sum_k \frac{|Y_t^{(k)} \cap \hat{Y}_t^{(k)}|}{|Y_t^{(k)}|} \tag{13}
$$

$$
F1 = \frac{2 \times precision \times recall}{precision + recall} \tag{14}
$$

When considering the accuracy of drug prediction, we also need to measure the safety of drug prediction. So to measure medication safety, we define DDI Rate to judge the probability of drug interactions in the predicted drug sequence:

$$
DDIRate = \frac{\sum_k N \sum_{i,j} [((c_i, c_j) \in \hat{Y}_t^{(k)}) \land (c_i, c_j) \in e_d]}{|\sum_k N \sum_{i,j} 1|} \tag{15}
$$

where the set will count each medication pair $(c_i, c_j)$ in the recommendation set if the pair belongs to the drug interaction adjacency matrix constructed in 6.1. Here $N$ is the size of the test dataset.

6.1.4 Baseline Methods

We compared our method with various baseline methods with regard to medication regimen prediction and medication regimen adequate evaluation. For medication regimen prediction, we compared our model with several baseline methods as follows.

1. **Bi-LSTM**: this baseline is a sequence-sequence model. At the encoding end, BI-LSTM is used to learn the diagnostic information at the input end, and at the decoding end, ordinary LSTM is used to predict drugs.

2. **GAMENet**: this baseline is a memory-enhancing neural network model that inherits a drug interaction knowledge graph through the graph convolutional network storage module to provide safe and personalized drug combination recommendations.

For appropriate medication evaluation, we compare our model with the following baselines.

1. **Empirical**: This method is based only on the medical experience of professional doctors in tertiary hospitals, without considering the drug contraindication information from existing drug instructions.

6.2 Evaluation Results

6.2.1 Medication Regimen Prediction Evaluation Results

Table 2 shows the rational drug use prediction results on the MIMIC_III dataset using our proposed method as well as the baselines. Results show our proposed method has the highest score among all baselines with respect to Jaccard, Precision, Recall, and F1. The model we used benefited from the advantages of its structure, which could obtain the relationship between patients’ multiple diagnoses, making it closer to the real doctor’s prescription when making drug predictions.

Table 3 shows the rational drug use regimen prediction results on the outpatient medical record data set using our proposed method as well as the baselines. Results show our proposed method achieves the best performance compared to other baseline methods.

Since this dataset is different from MIMIC_III, no authoritative drug classification has been performed, and drugs with therapeutic equivalence exist in this dataset as multiple drugs. Therefore, we chose to provide three alternative elements for each element generated by the sequence model. It is deemed to be the correct prediction when the actual use of the drug appears in the three alternatives. The calculation formula of its evaluation index is as follows:
Table 2  The Rational Medication Regimen Prediction Results of MIMIC_III Dataset

<table>
<thead>
<tr>
<th>Methods</th>
<th>Jaccard</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
<th>DDI Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bi-LSTM</td>
<td>0.5115</td>
<td>0.6705</td>
<td>0.5697</td>
<td>0.6160</td>
<td>0.1624</td>
</tr>
<tr>
<td>GAMENet</td>
<td>0.6517</td>
<td>0.7501</td>
<td>0.6535</td>
<td>0.6985</td>
<td>0.1324</td>
</tr>
<tr>
<td>Ours</td>
<td>0.8685</td>
<td>0.9555</td>
<td>0.8927</td>
<td>0.9173</td>
<td>0.0867</td>
</tr>
</tbody>
</table>

Table 3  The Prediction Results of Outpatient Medical Record Dataset

<table>
<thead>
<tr>
<th>Methods</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bi-LSTM</td>
<td>0.4355</td>
</tr>
<tr>
<td>GAMENet</td>
<td>0.6355</td>
</tr>
<tr>
<td>Ours</td>
<td>0.7769</td>
</tr>
</tbody>
</table>

\[ \text{Precision} = \frac{1}{\sum_{k}^{N} \sum_{l}^{M} \frac{|Y_{l}^{(k)} \cap \hat{Y}_{l}^{(k)}|}{M}} \tag{16} \]

where \( M = \min(|Y_{l}^{(k)}|, |\hat{Y}_{l}^{(k)}|) \).

Table 4 clearly shows the predicted results of rational drug use of the model in the Fujian province data set. The actual prescriptions shown on the left of the table below are two drugs taken by a boy for allergic rhinitis, mycoplasma infection, and bronchitis. On the right are the recommendations for drug therapy provided by our model. It can be found that the model in this paper can accurately cover the real prescription after providing three alternatives, and most of the other alternatives provided are also drugs for the treatment of respiratory diseases such as rhinitis. It indicates that the model in this paper can get the drug recommendations of actual doctors according to patient diagnosis and other characteristics.

Table 4  Drug Recommendation Cases in Fujian Province Data Set

<table>
<thead>
<tr>
<th>Real Prescription</th>
<th>Predict Prescription</th>
</tr>
</thead>
<tbody>
<tr>
<td>GanAn mixture</td>
<td>GanAn mixture, Josamycinpropionate Granules, Mometasone Furoate Aque</td>
</tr>
<tr>
<td>Loratadine Tablets</td>
<td>Montelukast Sodium Oral Granules, Loratadine Tablets, Ambroterol oral solution</td>
</tr>
</tbody>
</table>

Table 5  The Influence of Physiological Features

<table>
<thead>
<tr>
<th>Jaccard</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Physical Features</td>
<td>0.8556</td>
<td>0.9546</td>
<td>0.8727</td>
</tr>
<tr>
<td>Have Physical Features</td>
<td>0.8627</td>
<td>0.9564</td>
<td>0.8855</td>
</tr>
</tbody>
</table>

Table 6  The Influence of DDI Features

<table>
<thead>
<tr>
<th>DDI Rate</th>
<th>Jaccard</th>
</tr>
</thead>
<tbody>
<tr>
<td>No DDI Features</td>
<td>0.0906</td>
</tr>
<tr>
<td>Have DDI Features</td>
<td>0.0837</td>
</tr>
</tbody>
</table>

6.2.2 Medication Regimen Appropriateness Evaluation Results

In this paper, we use the trained models as classifiers to judge the appropriateness of prescriptions in small hospitals. Table 7 shows the average accuracy scores of medicine using appropriate evaluation using the proposed method as well as the baselines. We can see that the proposed method achieves the best performance with regard to evaluation accuracy scores. Specifically, the baseline method \textit{Empirical} baseline attempts to evaluate the appropriate use of drugs based only on the medication experience of professional doctors in tertiary hospitals and did not take into account the contraindication information of the drugs in the available drug instructions, which results in some combination drugs with adverse reactions being misjudged. In summary, the proposed method integrates the two heterogeneous information to model sequential patterns and therefore improves the accuracy of evaluation. The calculation formula of its evaluation index is as follows:

Table 7  The Appropriate Evaluation of Prescription in Small Hospitals

<table>
<thead>
<tr>
<th>Methods</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empirical</td>
<td>83%</td>
</tr>
<tr>
<td>Ours</td>
<td>89%</td>
</tr>
</tbody>
</table>

Table 8  The Rational Medication Regimen Prediction Results of MIMIC_III Dataset

<table>
<thead>
<tr>
<th>Methods</th>
<th>Jaccard</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
<th>DDI Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bi-LSTM</td>
<td>0.5115</td>
<td>0.6705</td>
<td>0.5697</td>
<td>0.6160</td>
<td>0.1624</td>
</tr>
<tr>
<td>GAMENet</td>
<td>0.6517</td>
<td>0.7501</td>
<td>0.6535</td>
<td>0.6985</td>
<td>0.1324</td>
</tr>
<tr>
<td>Ours</td>
<td>0.8685</td>
<td>0.9555</td>
<td>0.8927</td>
<td>0.9173</td>
<td>0.0867</td>
</tr>
</tbody>
</table>

Table 9  The Prediction Results of Outpatient Medical Record Dataset

<table>
<thead>
<tr>
<th>Methods</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bi-LSTM</td>
<td>0.4355</td>
</tr>
<tr>
<td>GAMENet</td>
<td>0.6355</td>
</tr>
<tr>
<td>Ours</td>
<td>0.7769</td>
</tr>
</tbody>
</table>

\[ \text{Precision} = \frac{1}{\sum_{k}^{N} \sum_{l}^{M} \frac{|Y_{l}^{(k)} \cap \hat{Y}_{l}^{(k)}|}{M}} \tag{16} \]

where \( M = \min(|Y_{l}^{(k)}|, |\hat{Y}_{l}^{(k)}|) \).

Table 10  Drug Recommendation Cases in Fujian Province Data Set

<table>
<thead>
<tr>
<th>Real Prescription</th>
<th>Predict Prescription</th>
</tr>
</thead>
<tbody>
<tr>
<td>GanAn mixture</td>
<td>GanAn mixture, Josamycinpropionate Granules, Mometasone Furoate Aque</td>
</tr>
<tr>
<td>Loratadine Tablets</td>
<td>Montelukast Sodium Oral Granules, Loratadine Tablets, Ambroterol oral solution</td>
</tr>
</tbody>
</table>

In addition, we conducted two comparative experiments to determine whether the addition of patients’ physiological features and DDI features could improve the effectiveness of our model. Table 5 shows the accuracy rate of drug recommendation is improved after the introduction of physiological characteristics in our model. Table 6 shows that after the introduction of DDI features in our model, although the accuracy of model prediction is slightly sacrificed, the probability of adverse drug interactions caused by recommended drugs is reduced. Therefore, we can adjust the weight proportion of DDI characteristics to meet the application requirements.
6.3 Clinical Appropriate Medication Evaluation System

In order to demonstrate the work of this paper more clearly, we have built a platform for the appropriateness of rational drug use and applied it in a small hospital to evaluate the appropriateness of doctors’ prescriptions. As shown in Figure 10, this platform is mainly divided into two parts, among which the left view is divided into patient-disease-drug three subgraphs, and the right view shows the evaluation results. In the left frame, you can firstly fill in patient information, diagnostic information, and drug information. Then click the button to evaluate for appropriateness of prescribing. Finally, the predicted medication results are displayed on the right side of the frame, with a tabulated comparison of the doctor’s medication regimen and the predicted medication regimen. At the same time, the entity relationship involved with the disease in the medication graph would be visualized below the evaluation results, which would be convenient for doctors and pharmacists to further review and modify prescriptions.

4. Case Study

We conduct a case study of one prescription randomly selected from 100 problematic prescriptions of inappropriate medication in small hospitals. As shown in Figure 10, in the input prescription, the patient is male, height 106cm, weight 17.5kg, age 4 years old. The patient’s diagnosed symptoms were acute bronchitis, bronchitis, and acute upper respiratory tract infection. The prescribed medicines were Amoxicillin and Clavulanate Potassium for Oral Suspension,Combivent, Pulmicort Respules and the corresponding courses of administration for the three drugs were 3 days, 1 day, and 1 day, respectively. Based on our proposed framework, the predicted medication regimen was Amoxicillin and Clavulanate Potassium for Oral Suspension for 3 days and Combivent for 1 day. After evaluation, the system would provide default color labels and red labels to represent consistent medication regimens and inconsistent medication regimens respectively. The red label in the picture indicated whether Budesonide Suspension for Inhalation(PULMICORT RESPULES) are unnecessary drugs.

7 Conclusions

In this paper, we investigate one of the key problems in rational medication, i.e., the evaluation of appropriate medication use. We propose a framework of rational appropriate drug
use based on medical association and big data to accurately predict medication regimens by leveraging prescription big data and medical text data. More specifically, we first employ Gaussian Mixture Model (GMM) to group patient population representation as physiological features. Second, we extract the patient’s diagnostic information and transform it into a word vector by pre-training word vector Bert(Bidirectional Encoder Representations from Transformers) as the diagnostic features. Third, in order to reduce adverse drug interactions caused by drug combinations, we employ a graph convolution network to transform drug interaction information into drug interaction features. Finally, we use the sequence generation model to give the a rational medication regimen. We evaluate the proposed framework with two medical record datasets: Medical Information Mart for Intensive Care III(MIMIC-III) and real-world prescription big data collected from tertiary hospitals. Results show that our method has a more accurate medication regimen prediction ability, and consistently outperforms other baselines. Results show that our method is more effective than other baseline methods in the accuracy of medication regimen prediction of rational medication. In addition, it has achieved high accuracy in the appropriate medication detection of prescriptions in small hospitals.

One of the limitations of this work is the feature selection. There might be other indication or physiology features that could be found to be associated with medication regimens and be used as predictive features for example. For example, for adolescents, the probability of developing corresponding diseases during adolescence can also be considered in the prediction model to improve the prediction accuracy in teenagers. Presently, We are currently working with hospitals to retrieve richer information related to prescription dataset, such as the Picture Archiving and Communication Systems (PACS) and inspection results from the Lab Information Systems (LIS), which we believe will provide useful and important features for drug regimen prediction.

In the future, we plan to extend our work in the following directions. First, we plan to involve more data sources from other hospital information systems, especially data from clinical laboratories, to investigate more relevant factors of doctor medication. Second, we plan to investigate the reasons for the wrong medication sequence list, including overtreatment or undertreatment by model overfitting, and then leverage the knowledge to improve our predictive models. Third, we plan to integrate our method with the existing clinical decision support systems to provide dosing recommendations for doctors and pharmacists in small hospitals.

References

1. Promoting rational use of medicines. WHO Regional Office for South-East Asia, 2011.
15. J. Fan, F. Han, and H. Liu, “Challenges of big data analysis,” National


Shiwang Huang received the B.Eng. degree in computer science and technology from Nanjing University of Posts and Telecommunications, China in 2019. He is currently pursuing the master’s degree in Department of Computer Science, Xiamen University.

Xiaohai Cai received the B.Eng. degree in computer science and technology from Jiangsu University, China in 2018. He is currently pursuing the master’s degree in Department of Computer Science, Xiamen University.

Binbin Zhou received her M.Phil degree in computer science from Hongkong Polytechnic University, in 2011. She is currently pursuing a Ph.D. degree in the Department of Computer Science, Zhejiang University. Her research interests are mainly in urban computing, big data applications, and ubiquitous computing.

Zhiming Lin received the B.Eng. degree in computer science and technology from Fujian Normal University, China in 2019. He is currently pursuing the master’s degree in Department of Computer Science, Xiamen University.

Yunting Shao received the master’s degree in Department of Computer Science, Xiamen University, in 2020. Her research interests include big data analytics and rational medication.

Longbiao CHEN is an assistant professor with Fujian Key Laboratory of Sensing and Computing for Smart City, Xiamen University. He obtained his Ph.D. degree in computer science from Sorbonne University, France in 2018 and Zhejiang University, China in 2016. His research interests include Ubiquitous Computing, Urban Computing, and Big Data Analytics.

Chenhui Yang received his Ph.D. degree from Zhejiang University, China. He is a professor in the School of Information Science and Technology at Xiamen University, China. His research interests include computer vision and computer graphics.

Linghong Hong received her M.S. degree in pharmaceutical science from Zhejiang University, China in 2017. She is the pharmacist in charge of Xi'ang'an Hospital of Xiamen University. Her research interests include clinical pharmacy, intravenous admixture, rational use of medicines, and medical big data analytics.