Drug Package Recommendation via Interaction-aware Graph Induction*

Zhi Zheng
School of Computer Science and Technology, University of Science and Technology of China
Hefei, Anhui, China
zhengzhi97@mail.ustc.edu.cn

Chao Wang
Tong Xu
wdyx2012@mail.ustc.edu.cn
tongxu@ustc.edu.cn
School of Computer Science and Technology, University of Science and Technology of China
Hefei, Anhui, China

Dazhong Shen
Penggang Qin
sdz@mail.ustc.edu.cn
qinpg@mail.ustc.edu.cn
School of Computer Science and Technology, University of Science and Technology of China
Hefei, Anhui, China

Baoxing Huai
Huawei Technologies
Hangzhou, Zhejiang, China
huaibaoxing@huawei.com

Tongzhu Liu
The First Affiliated Hospital of USTC
Hefei, Anhui, China
liutongzhu2015@126.com

Enhong Chen
School of Computer Science and Technology, University of Science and Technology of China
Hefei, Anhui, China
cheneh@ustc.edu.cn

ABSTRACT

Recent years have witnessed the rapid accumulation of massive electronic medical records (EMRs), which highly support the intelligent medical services such as drug recommendation. However, prior arts mainly follow the traditional recommendation strategies like collaborative filtering, which usually treat individual drugs as mutually independent, while the latent interactions among drugs, e.g., synergistic or antagonistic effect, have been largely ignored. To that end, in this paper, we target at developing a new paradigm for drug package recommendation with considering the interaction effect within drugs, in which the interaction effects could be affected by patient conditions. Specifically, we first design a pre-training method based on neural collaborative filtering to get the initial embedding of patients and drugs. Then, the drug interaction graph will be initialized based on medical records and domain knowledge. Along this line, we propose a new Drug Package Recommendation (DPR) framework with two variants, respectively DPR on Weighted Graph (DPR-WG) and DPR on Attributed Graph (DPR-AG) to solve the problem, in which each the interactions will be described as signed weights or attribute vectors. In detail, a mask layer is utilized to capture the impact of patient condition, and graph neural networks (GNNs) are leveraged for the final graph induction task to embed the package. Extensive experiments on a real-world data set from a first-rate hospital demonstrate the effectiveness of our DPR framework compared with several competitive baseline methods, and further support the heuristic study for the drug package generation task with adequate performance.

*Tong Xu is the corresponding author.

WWW '21, April 19–23, 2021, Ljubljana, Slovenia

ACM ISBN 978-1-4503-8312-7/21/04.
https://doi.org/10.1145/3442381.3449962

CCS CONCEPTS

• **Information systems** → **Data mining**.

KEYWORDS

Drug Recommendation, Package Recommendation, Graph Neural Network

ACM Reference Format:
Zhi Zheng, Chao Wang, Tong Xu, Dazhong Shen, Penggang Qin, Baoxing Huai, Tongzhu Liu, and Enhong Chen. 2021. Drug Package Recommendation via Interaction-aware Graph Induction. In *Proceedings of the Web Conference 2021 (WWW '21), April 19–23, 2021, Ljubljana, Slovenia.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3442381.3449962

1 INTRODUCTION

With the growth of population and the intensification of population aging, people's demand for high-quality medical services continues to rise, and the pressure on the medical workers is increasing. Moreover, certain public health emergencies such as the outbreak of COVID-19, will also have a huge impact on the medical system. Meanwhile, artificial intelligence (AI) technologies have shown enormous potential to reduce human labor. Therefore, if AI technologies could be effectively utilized to realize intelligent diagnosis and drug recommendation clinically, it will greatly improve the overall quality of medical services.

Fortunately, with the popularization of information technology in the medical industry, electronic medical records (EMRs) have been widely used in major hospitals, which powerfully support the downstream intelligent applications like medical image analysis [11, 26], chronic disease management [12, 30], medical text analysis [2, 31], etc. However, due to the limitation of data and technology, drug recommendation based on EMR is still largely unexplored. In terms of data, similar to traditional recommendation system, drug recommendation is sensitive to data quality, but it is hard to get reliable medical data sources. Moreover, most patients have only been recorded once or several times in EMR database, which

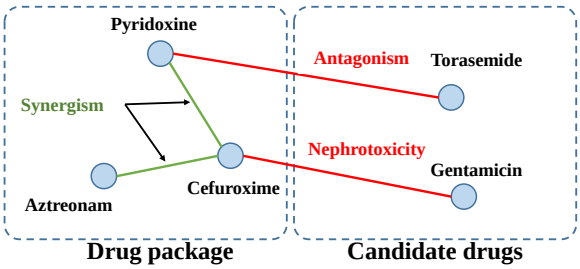


Figure 1: An example for a patient with kidney disease.

makes it hard to utilize conventional personalized recommendation methods based on user preference analysis. In terms of technology, it is very important for the recommender system to consider both drug effect and the interaction between drugs at the same time, and give the patient a suitable drug package, which contains multiple drugs. However, most of existing studies generally rely on traditional methods such as collaborative filtering [43] to solve this problem. Due to the lack of item relation data for interaction analysis, there are limits for these methods to achieve satisfactory performance in practical applications.

In order to address the above challenges, in this paper, we aim to develop a new paradigm for drug package recommendation with the awareness of drug interaction. The rationale behind this is that the interaction between drugs will influence the effect of the drug package, and the impact of drug interaction on drug effect will be further affected by patient conditions. We illustrate this by a patient with kidney disease as shown in Figure 1. The drug package for this patient contains three drugs, respectively pyridoxine, aztreonam and cefuroxime. Cefuroxime is synergistic with the other two drugs, which can improve the effect of the drug package. Torasemide is antagonistic with pyridoxine, so it is not included in the package. Furthermore, the combination of cefuroxime and gentamicin has a synergistic antibacterial effect, but at the same time it may increase nephrotoxicity, so it is not suitable for this patient.

Along this line, we first design a pre-training model to get the embedding of patients and drugs based on neural collaborative filtering (NCF). Then we collect drug interaction data from public online dataset and divide drug pairs into three categories with the help of domain experts, respectively No Interaction, Synergism and Antagonism. After that, we propose to represent drug packages as graphs based on the labeled data. Furthermore, we propose a Drug Package Recommendation (DPR) framework with two variants. The first one, namely DPR on Weight Graph (DPR-WG), regards the effect of drug interaction as graph edge weights, while the second one, DPR on Attributed Graph (DPR-AG), utilizes edge attribute vectors to describe the influence of drug interaction. In both two models, we exploit a mask layer to capture the impact of the patient condition on the drug package representation, and Graph Neural Networks (GNNs) are leveraged for the final graph induction task to embed the package. Finally, extensive experiments on a real-world dataset from a first-rate hospital demonstrate the effectiveness of our DPR framework compared with several competitive baseline methods, and further support the heuristic study for the drug package generation task with adequate performance.

Specifically, the major contributions of this paper can be summarized as follows:

- We develop a new paradigm to represent drug packages as graphs based on drug interaction classification.
- We design a drug package recommendation framework with two variants, which can integrate drug interaction information based on graph induction.
- We propose to utilize a mask layer to capture the impact of patient condition on the drug package representation.
- We conduct extensive experiments on a real-world data set from a first-rate hospital, which clearly validate the effectiveness of our DPR framework and reveal some interesting rules based on the derived insights on patient conditions and drug interaction.

2 RELATED WORK

In this section, we will summarize the related works as following three categories, respectively drug recommendation system, package recommendation system, and graph neural networks.

2.1 Drug Recommendation System

Recommendation systems have been widely used in a variety of applications like social networking and e-commerce. The methods can be broadly classified into two categories, respectively neighborhood-based collaborative filtering methods based on similar users or items [1], and model-based methods, particularly latent factor models that factorize the user-item matrix into user factors and item factors [19]. Recent recommender systems have been further advanced by the significant contribution from deep learning [16, 39, 42], where user preferences and item characteristics can be learned in deep architectures. Based on these technologies, some methods focusing on drug recommendation have been put forward. For example, [44] introduces a LDA-based contextual collaborative model called Medicine-LDA to integrate the multi-source information. [41] constructs a heterogeneous graph which includes patients and drugs, and describes a novel recommendation system based on label propagation. [8] develops a joint model with a recommendation component and an ADR label prediction component to recommend a set of to-avoid drugs. With the increasing emergence of knowledge graph, some researchers have extracted information from medical database like [22] to build up giant medical knowledge graphs. Based on these knowledge graphs, [38] proposes to jointly embed diseases, drugs and patients into a shared lower dimensional space, and decomposes the drug recommendation into a link prediction process. However, these models lack the ability to recommend drugs as a package, and the studies on drug interaction are not thorough enough.

2.2 Package Recommendation System

Most recommendation research concentrates on recommending one item to users at a time. However, in many real world scenarios, the platform needs to show users a set of items, in other words, a package (or a bundle). Several efforts have been made to solve this problem. Some studies turn this problem into optimization problems like 0-1 Knapsack problem, and provide some approximate solutions due to the NP-Hardness [10, 21, 32, 45]. [27] puts

Table 1: Statistics of our dataset.

Discription	Number
The number of records	158,556
The number of drugs	1,428
The number of words in disease document	1,242
The average size of drug packages	18
The number of aligned drugs	565
The number of drug pairs with No Interaction	2,560
The number of drug pairs with Synergism	22,986
The number of drug pairs with Antagonism	6,389

forward a Tourist-Area-Season topic model and proposes a cocktail approach on personalized travel package recommendation. [3] proposes a bundle generation network which decomposes the problem by derterminantal point processes. [33] develops a model which utilizes the trained features of an item recommendation model to learn the personalized ranking over bundles. [7] contributes a neural network solution based on factorized attention network to aggregate the item embeddings in a package. [6] proposes a model based on graph neural network which explicitly models the interaction and affiliation between users, bundles, and items by unifying them into a heterogeneous graph. However, these models neglect the different types of interactions between items, which prevents them from capturing satisfactory performance for drug package recommendation.

2.3 Graph Neural Networks

Recently, many studies on extending deep learning approaches for graph data have emerged. Unlike standard neural networks, GNNs retain a state that can represent information from its neighborhood with arbitrary depth. For example, [18] presents graph convolutional network (GCN) for semi-supervised learning on graph data via an approximation of spectral graph convolutions. [14] presents GraphSAGE to generate node embeddings by sampling and aggregating features from the local neighborhoods of nodes. [35] presents graph attention networks (GATs) which leverage masked self-attentional layers to address the shortcomings of methods based on graph convolutions. [13] further presents that the essence of existing GNNs is to learn a message passing algorithm and an aggregation procedure to compute a function of the entire input graph, and reformulates existing models into a single common framework called Message Passing Neural Networks (MPNNs). With the strong power of learning structure, GNNs have been widely applied in many fields. For example, [24, 40] utilize graph data and graph neural networks for competitive analysis. [28] propose a deep model to integrate structural and temporal social contexts to address the dynamic social-aware recommendation task.

3 PRELIMINARIES

In this section, we first introduce the real-world dataset used in our study, and then propose the problem formulation of drug package recommendation.

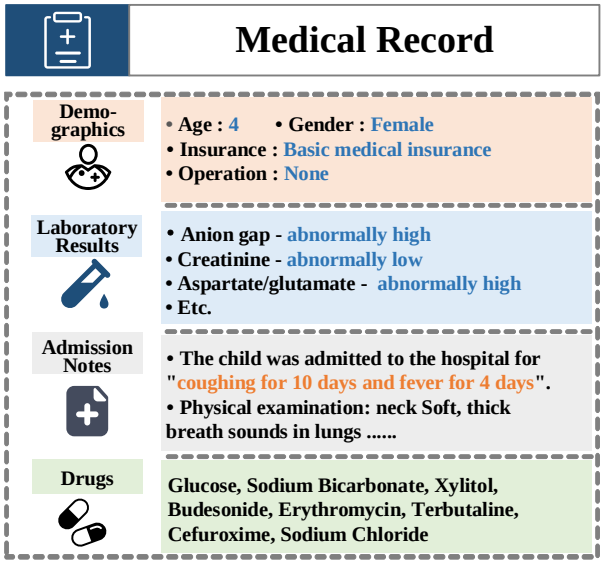


Figure 2: An example of the medical record in our dataset.

3.1 Data Description and Preprocessing

The EMR dataset used in this paper comes from the electronic medical record database of a first-rate hospital in China. As shown in Figure 2, each medical record contains the following information:

- **Demographics**. Demographics are formatted data including basic patient information, such as patient's gender, age, type of medical insurance, whether surgery has been performed, etc. This information provides guidance for doctors to prescribe, for example, some drugs are not suitable for children, while some drugs are only covered by certain medical insurance, etc.
- **Laboratory results**. A laboratory test is a procedure in which the hospital takes a sample of the patient's body fluid or body tissue to get information of the patient's health. The laboratory results are shown as the patient's values and normal values for laboratory items. For example, "glucose value: 77 mg/dL, normal value: 65-99 mg/dL".
- **Admission notes**. An admission note is part of a medical record that documents the patient's status including physical examination findings, reasons why the patient is being admitted for inpatient care to a hospital, and the initial instructions for the patient's care.
- **Drugs**. This information includes all of the drugs used during the patient's hospital stay.

In order to integrate and utilize the above multi-source heterogeneous data, we conduct the following preprocessing steps. First, for the demographics, we convert them into documents, e.g., "Gender : Male, Age : Teenager". Second, for the laboratory results, we divide the results into three levels, respectively normal, abnormally high and abnormally low according to the given normal values. We then extract all abnormal test results (abnormally high and low) from the results and converted them into documents, e.g., "glucose

Table 2: Examples of drug interaction labeling.

Drug A	Drug B	Description	Classification	Direction
Amoxicillin	Oseltamivir	No Interaction	No Interaction	Bidirection
Dipyridamole	Valsartan	Dipyridamole may increase the antihypertensive activities of Valsartan.	Synergism	A to B
Repaglinide	Doxepin	Doxepin may decrease the hypoglycemic activities of Repaglinide.	Antagonism	B to A

value : abnormally high, lipid panel : abnormally high". After that, we merge the demographic documents and laboratory result documents, namely disease documents. Finally, for the admission notes, we remove all the punctuation and meaningless characters, and adjust all of the admission notes in the dataset to the same length by padding and cut-off.

For the purpose of studying the interaction between drugs, we collect data from two large online pharmaceutical knowledge bases, i.e., DrugBank[1] and YaoZhi[2], where users can check drug properties and drug-drug interaction. The drug interaction information in these two databases are stored in text format based on some certain templates. We further classify the templates into three categories with the help of domain experts, respectively No Interaction, Synergism and Antagonism. No Interaction means there is no interaction between two drugs. Synergism means the combination of two drugs can lead to enhanced drug effect, and Antagonism is the opposite. Table 2 shows some examples of different drug interactions. Note that the interaction can be directed, for example, if drug A can increase the effect of drug B, then the direction is from A to B. Moreover, for most of the drug pairs, we cannot confirm whether there is any type of interactions between them, so we leave them as unlabeled. Section 4.2.1 will further discuss how to exploit these labeled and unlabeled data.

Finally, we pick out the EMR records containing more than one drug and we get totally 158,556 EMR records with complete information. More detailed statistics of our data are shown in Table 1.

3.2 Problem Formulation

Based on the above EMR and drug interaction data, here we introduce the problem formulation of drug package recommendation. For facilitating illustration, Table 3 lists some important mathematical notations used throughout this paper.

Suppose there are N patients and M drugs in the training set. Based on the above preprocessing method, for patient i, we can construct the disease document and turn it into one-hot encoding form as $\mathcal{W}_i = \{w_{i,1}, w_{i,2}, \ldots, w_{i,p}\}$, where $w_{i,\cdot}$ is the 0/1 indicator value for a demographic feature or a lab result. In addition, we can formulate the admission note as $\mathcal{T}_i = \{t_{i,1}, t_{i,2}, \ldots, t_{i,q}\}$, where $t_{i,\cdot}$ is a word in the processed admission note. In this way, the patient i can be expressed as a patient description $\mathcal{U}_i = \{\mathcal{W}_i, \mathcal{T}_i\}$. We also have the drug package $\mathcal{P}_i = \{d_{i,1}, d_{i_2}, \ldots, d_{i,s}\}$, where $d_{i,\cdot}$ is a drug that patient i used. Moreover, based on the labeled drug interaction data, we can construct the drug relation matrix $\mathcal{R} \in \mathbb{R}^{M \times M}$, where $\mathcal{R}_{ij}$ represents the interaction between d_i and d_j, namely **0 for No Interaction, 1 for Synergism, 2 for Antagonism and -1 for unknown.** Note that the direction is from d_i to d_j. Along this line,

[1]https://go.drugbank.com/releases/latest
[2]https://db.yaozh.com/interaction

Table 3: Mathematical notations.

Symbol	Description
N, M	The number of patients and the number of drugs;
$\mathcal{P}_i$	The drug package of patient i;
$\mathcal{W}_i$	The disease document of patient i;
$\mathcal{T}_i$	The admission note of patient i;
$\mathcal{U}_i$	The patient discription of patient i;
$\mathcal{G}_i$	The drug package graph of patient i;
$\mathcal{R}$	The drug relation matrix;
Θ	Model Parameters;
d_j	The jth drug in the entire drug set;
$d_{i,\cdot}$	Drug in the drug package of patient i;
$w_{i,\cdot}$	Indicator value in the disease document of patient i;
$t_{i,\cdot}$	Word in the admission note of patient i;
$MLP(\cdot)$	Multilayer Perceptron with ReLU Activation Function.

the problem of drug package recommendation can be formulated as follows:

Definition 1 (Drug Package Recommendation). *Given a set of patient descriptions $\{\mathcal{U}_1, \mathcal{U}_2, \ldots, \mathcal{U}_N\}$ with the corresponding drug packages $\{\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_N\}$, and the drug relation matrix $\mathcal{R}$, the goal of drug package recommendation is to get a personalized scoring function for each patient: $f_u : \mathcal{P} \rightarrow \mathbb{R}$.*

Note that the cold start patients and packages are very common in our drug package recommendation problem. For example, a new patient comes to the hospital or a doctor prescribes a new drug package. This requires the model to score a package based on the patient condition and the effect of drug packages, making the problem radically different from traditional recommendation based on user-item interaction matrix.

4 TECHNICAL DETAILS

In this section, we will introduce the framework of our model in detail. As shown in Figure 3, our framework mainly consists of three components, i.e., pre-training, package graph construction, and drug package recommendation. Specifically, we first design a pre-training method based on neural collaborative filtering to get the initial embedding of patients and drugs. Then, we propose to construct drug package graphs based on the medical records and domain knowledge. Finally, a novel Drug Package Recommendation (DPR) framework with two variants are proposed to solve the drug package recommendation problem.

4.1 Pre-training

A patient's description consists of two heterogeneous parts, and a drug package consists of several drugs. In order to recommend

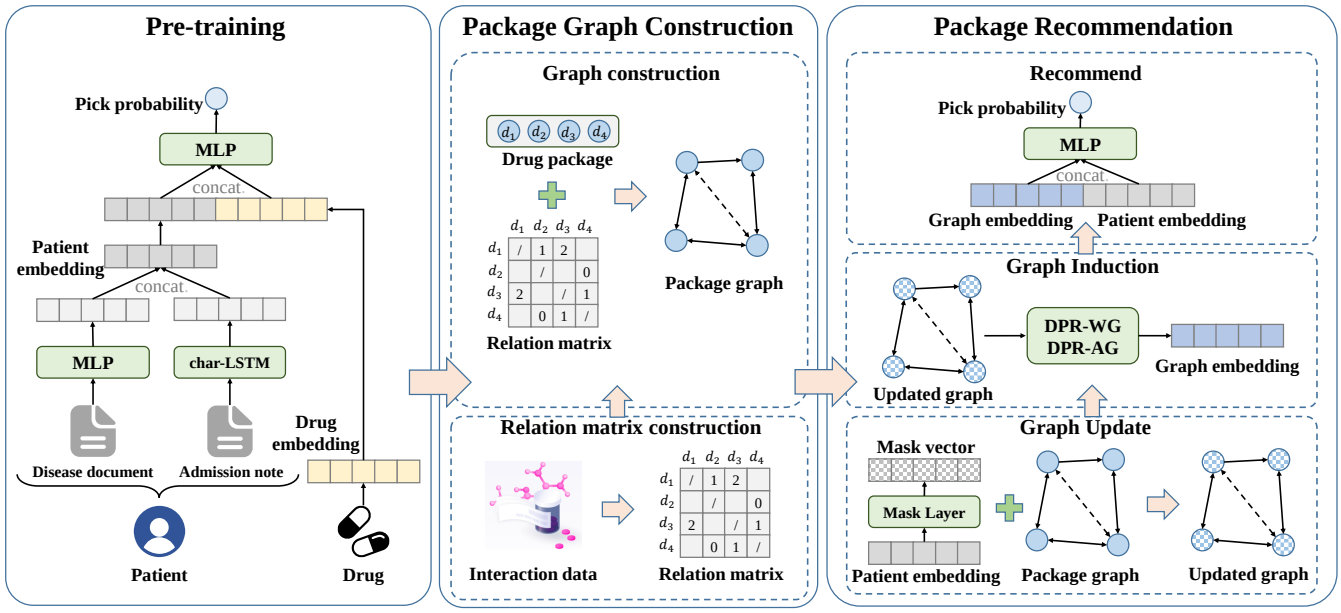


Figure 3: A framework overview of the drug package recommendation system.

drug packages, we first need to get the embeddings of drugs and patients. Therefore, we propose a pre-training method as follows.

First, we propose a hybrid method to get the patient embedding $\mathbf{u}$ based on patient description $\mathcal{U} = \{\mathcal{W}, \mathcal{T}\}$, which can be split into two steps. To be specific, in the first step, we extract the feature of the patient's disease document by MLP as:

$$\mathbf{m}_w = MLP(\mathcal{W}). \tag{1}$$

In the second step, we associate each word t_k in patients' admission notes with a word embedding vector $\mathbf{x}_k$. By this way we can convert $\mathcal{T}$ to a sequence of vectors $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_q)$. Then we input the sequence into char-LSTM [20] as:

$$\begin{aligned}
\mathbf{i}_t &= \sigma(\mathbf{W}_{xi}\mathbf{x}_t + \mathbf{W}_{hi}\mathbf{h}_{t-1} + \mathbf{W}_{ci}\mathbf{c}_{t-1} + \mathbf{b}_i),\\
\mathbf{f}_t &= \sigma\left(\mathbf{W}_{xf}\mathbf{x}_t + \mathbf{W}_{hf}\mathbf{h}_{t-1} + \mathbf{W}_{cf}\mathbf{c}_{t-1} + \mathbf{b}_f\right),\\
\mathbf{c}_t &= \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tanh(\mathbf{W}_{xc}\mathbf{x}_t + \mathbf{W}_{hc}\mathbf{h}_{t-1} + \mathbf{b}_c),\\
\mathbf{o}_t &= \sigma(\mathbf{W}_{xo}\mathbf{x}_t + \mathbf{W}_{ho}\mathbf{h}_{t-1} + \mathbf{W}_{co}\mathbf{c}_t + \mathbf{b}_o),\\
\mathbf{h}_t &= \mathbf{o}_t \odot \tanh(\mathbf{c}_t).
\end{aligned} \tag{2}$$

We get the final time step output $\mathbf{h}_q$ as the embedding of $\mathcal{T}$, and the patient embedding $\mathbf{u}$ is the concatenation of the two parts:

$$\mathbf{u} = \left[\mathbf{m}_w || \mathbf{h}_q\right]. \tag{3}$$

Second, we associate each drug d_j with a randomly initialized embedding $\mathbf{d}_j$ which directly projects drug one-hot ID to the latent space. Finally, We utilize Neural Collaborative Filtering (NCF) framework [16] and Bayesian Personalized Ranking (BPR) loss [34] to train the above embeddings and models. Specifically, for patient i, we get a patient-drug predictive model by feeding patient embedding $\mathbf{u}_i$ and drug embedding $\mathbf{d}_j$ into a matching model:

$$\hat{r}_{ij} = MLP\left(\left[\mathbf{u}_i || \mathbf{d}_j\right]\right), \tag{4}$$

Then we adopt BPR loss as:

$$L = \sum_{i=1}^{N} \sum_{j \in \mathcal{P}_i} \sum_{l \notin \mathcal{P}_i} -\ln \sigma\left(\hat{r}_{ij} - \hat{r}_{il}\right) + \lambda \left\|\Theta\right\|_2^2, \tag{5}$$

where d_j is in drug package and d_l is not. We minimize the loss function forcing the prediction $\hat{r}_{ij}$ to be larger than $\hat{r}_{il}$. $\sigma(\cdot)$ is the sigmoid function, and Θ is the parameter set. L_2 regularization is applied to prevent overfitting.

4.2 Package Graph with Message Passing

Compared with traditional item recommendation, the core problem of drug package recommendation is how to get the representation of drug packages considering the interaction between drugs. Therefore, in this section, we propose to utilize graph models to solve this problem. To be specific, we first present a method to convert the drug packages into package graphs. Then, we formulate the message passing framework which will be further utilized for the graph induction task.

4.2.1 Package Graph Construction. For drug package $\mathcal{P}$, we define a corresponding package graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, where $\mathcal{V}$ is the node set and $\mathcal{E}$ is the edge set. Each specific node $v \in \mathcal{V}$ is associated with corresponding drug embedding $\mathbf{d}$. Each directed edge $e_{vu} \in \mathcal{E}$ also has its attribute, and its form will change with different methods, which will be discussed in later sections.

The topology structure of the package graph $\mathcal{G}$, i.e., whether edge e_{vu} should exist, needs to be defined. Theoretically, since any pair of drugs may have drug interaction, the package graph $\mathcal{G}$ should be a complete graph, where all nodes are connected with each other. However, this will make the time complexity of graph induction increases from $O(n)$ to $O(n^2)$ owing to the pairwise

interaction. Furthermore, we find that the frequency of drug co-occurrence obeys a long-tailed distribution, which means most of the drug pairs have no clear relationship. Therefore, we propose the following two criterions to define the topology of a package graph. For nodes v, u: 1) If $\mathcal{R}_{vu} \neq -1$, which means this drug paired has been labeled in Section 3.1, then edge e_{vu} exists. 2) Calculate the co-occurrence proportion $p_{ij} = num_{ij}/num_i$, where num_i means the number of packages containing drug i, and num_{ij} means the number of packages containing both drug i and drug j. If p_{ij} is bigger than a threshold value, then edge e_{vu} exists.

4.2.2 Message Passing on Package Graph. We propose to exploit the MPNN [13] framework for making use of the package graphs constructed in the last section. MPNN is a general approach to describe GNNs, which inductively learns a node representation by recursively aggregating and transforming the feature vectors of its neighboring nodes. A per-layer update of the MPNN model in our setting involves message passing, message aggregation, and node representation updating, which can be expressed as:

$$\mathbf{m}_{vu}^{(l)} = \text{MESSAGE}(\mathbf{h}_u^{(l-1)}, \mathbf{h}_v^{(l-1)}, \mathbf{e}_{vu}), \tag{6}$$

$$\mathbf{M}_u^{(l)} = \text{AGGREGATION}(\{\mathbf{m}_{vu}^{(l)}, \mathbf{e}_{vu}\} \mid v \in \mathcal{N}(u)\}), \tag{7}$$

$$\mathbf{h}_u^{(l)} = \text{UPDATE}(M_u^{(l)}, \mathbf{h}_u^{(l-1)}), \tag{8}$$

where $\mathbf{m}_{vu}^{(l)}$ is the message vector passing from v to u, $\mathbf{h}_u^{(l)}$ is the representation of node u on the layer l; $\mathbf{e}_{vu}$ is the attribute corresponding to edge e_{vu}. $\mathcal{N}(u)$ is the neighborhood of node u from where it collects information to update its aggregated message $\mathbf{M}_u$. $\mathbf{h}_u^{(0)}$ is initialized by corresponding drug embedding $\mathbf{d_i}$, and we also express it as $\mathbf{d_u}$ for facilitating illustration.

4.3 Drug Package Recommendation

After the formulation of package graphs and massage passing neural networks, we can finish the graph induction task, i.e., get the embedding of the package graph, based on the MPNN framework and further solve the drug package recommendation problem. The key to obtain effective representation of the drug package graph is to utilize the edge attributes to capture the interaction between the drugs. Therefore, we propose the following two ways to formulate the edge attributes in package graphs from two different point of views. First, since the two major interactions in our dataset, respectively Synergism and Antagonism, are opposite to each other, we can simply exploit signed edge weights to describe the drug interaction intensity. Second, if we expect our model to be more generic, we can define the edge attributes as vectors which contain the information about the type of interaction. Along this line, we propose our Drug Package Recommendation (DPR) model with two variants, respectively DPR on Weighted Graph (DPR-WG) and DPR on Attributed Graph (DPR-AG) in the following sections.

4.3.1 DPR on Weighted Graph. In DPR-WG, we present to convert a package graph $\mathcal{G}$ into a weighted graph by assigning real numbers to edge attributes, i.e., $e_{vu} \in \mathbb{R}$. Specifically, for edge e_{vu} in a package graph $\mathcal{G}$, we initialize the edge attribute as:

$$\mathbf{e}_{vu} = \begin{cases} 1 & \mathcal{R}_{vu} = 1, \\ -1 & \mathcal{R}_{vu} = 2, \\ p_{vu} & otherwise. \end{cases}$$

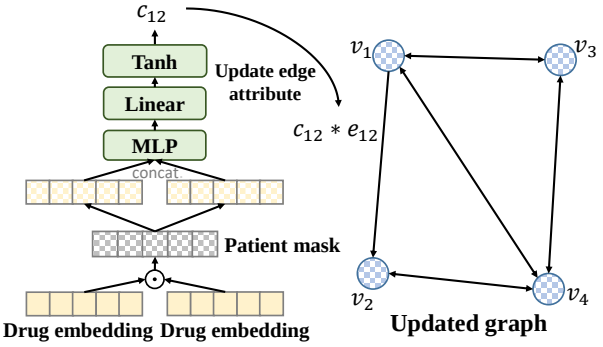


Figure 4: Edge attribute updating progress of DPR-WG.

Note that we set edge attribute for e_{vu} even if $\mathcal{R}_{ij} = 0$ or $\mathcal{R}_{ij} = -1$, since the interaction data may be incomplete or incorrect.

As previously stated, the impact of drug interaction will also be affected by patient condition. Inspried by [37], we propose to utilize a mask layer to extract a mask vector from a patient's embedding $\mathbf{u}$, and get the conditional drug embedding $\hat{\mathbf{d}}_u$ as follows:

$$\hat{\mathbf{d}}_u = \sigma\left(MLP\left(\mathbf{u}\right)\right) \odot \mathbf{d}_u, \tag{9}$$

where the mask layer is formed as $\sigma\left(MLP\left(\cdot\right)\right)$, and the mask vector $\sigma\left(MLP\left(\mathbf{u}\right)\right)$ plays the role of feature selecting on the drug embeddings. $\odot$ represents the element-wise product of two vectors. Then, for edge e_{vu}, we can calculate a contextual impact factor c_{vu} as:

$$c_{vu} = Tanh(a^\top MLP([\hat{\mathbf{d}}_u || \hat{\mathbf{d}}_v])), \tag{10}$$

where $a^\top$ is a row vector which has the same length with the MLP output. The contextual impact factor c_{vu} reflects the impact of the the patient condition on the drug interaction between d_u and d_v. After the above calculation, we can update the edge attribute as $\hat{\mathbf{e}}_{vu} = c_{vu} * \mathbf{e}_{vu}$. Figure 4 shows the updating progress in detail.

Then, we can form the GNN layer using edge weight for filtering as the following steps:

$$\mathbf{m}_{vu}^{(l)} = W_1^{(l-1)} \mathbf{h}_v^{(l-1)}, \tag{11}$$

$$\mathbf{M}_u^{(l)} = \sum_{v \in \mathcal{N}(u)} GRU\left(\hat{\mathbf{e}}_{vu} \mathbf{m}_{vu}^{(l)}, \mathbf{h}_u^{(l-1)}\right), \tag{12}$$

$$\mathbf{h}_u^{(l)} = MLP\left(W_0^{(l-1)} \mathbf{h}_u^{(l-1)} + \mathbf{M}_u^{(l)}\right), \tag{13}$$

where W denotes the model's parameters to be learned, and GRU denotes the gated recurrent neural network [9]. We set the dimension of all the layers equal to the dimension of 0th layer.

Now we can utilize the formed GNN layer for the graph induction task. Note that different from general sparse graphs, a drug package graph is a graph which is dense enough. Therefore, we only need one layer of GNN to extract almost all the information we expected, and there is no need for high-order neighbors, which will be discussed later. For each node v, we have the initial node embedding $\mathbf{d}_v$ and the corresponding hidden representation $\mathbf{h}_v$ from the GNN layer. Following [25], the package graph embedding can be formed as:

$$\mathbf{g} = \sum_{v \in V} \sigma\left(MLP\left([\mathbf{d}_v || \mathbf{h}_v]\right)\right) \odot \left(MLP\left([\mathbf{d}_v || \mathbf{h}_v]\right)\right). \tag{14}$$

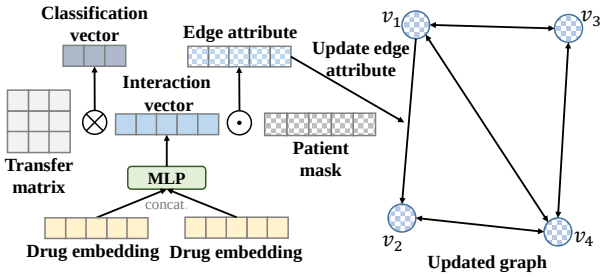


Figure 5: Edge attribute updating progress of DPR-AG.

Again, we utilize NCF framework and BPR loss to train the model. For patient i, we have the patient embedding $\mathbf{u}_i$ and the corresponding package graph embedding $\mathbf{g}_i$. The loss function can be formed as follows, where the MLP model is the final prediction model:

$$L = \sum_{i=1}^{N} \sum_{j \neq i} -\ln \sigma \left(MLP\left([\mathbf{u}_i || \mathbf{g}_i]\right) - MLP\left([\mathbf{u}_i || \mathbf{g}_j]\right)\right) + \lambda \left\|\Theta\right\|_2^2, \tag{15}$$

4.3.2 Drug Package Recommendation on Attributed Graph. In DPR-AG, the package graph $\mathcal{G}$ is formed as an attributed graph, where both nodes and edges have corresponding attribute vectors.

Specifically, for edge e_{vu} and the corresponding drug embedding $\mathbf{d}_v, \mathbf{d}_u$, we first form the edge attribute vector $\mathbf{e}_{vu}$ as the interaction vector between $\mathbf{d}_v$ and $\mathbf{d}_u$, which is calculated by MLP model as:

$$\mathbf{e}_{vu} = MLP\left([\mathbf{d}_v || \mathbf{d}_u]\right). \tag{16}$$

Then, we utilize the mask layer again to update the edge attributes by adding the impact of patient's condition on the interaction vector as follows:

$$\hat{\mathbf{e}}_{vu} = \sigma\left(MLP\left(\mathbf{u}\right)\right) \odot \mathbf{e}_{vu}. \tag{17}$$

The detailed updating progress is shown in Figure 5. Based on the above steps, we can form the GNN layer as the following steps, note that the settings are the same as DPR-WG in the former section:

$$\mathbf{m}_{vu}^{(l)} = W_1^{(l-1)} \hat{\mathbf{e}}_{vu}^{(l-1)}, \tag{18}$$

$$\mathbf{M}_u^{(l)} = \sum_{v \in \mathcal{N}(u)} \mathbf{m}_{vu}^{(l)}, \tag{19}$$

$$\mathbf{h}_u^{(l)} = MLP\left(W_0^{(l-1)} \mathbf{h}_u^{(l-1)} + \mathbf{M}_u^{(l)}\right). \tag{20}$$

After getting the package graph embedding by equation 14, we can form the loss function for DPR-AG. The essential difference between DPR-WG and DPR-AG is that, in DPR-WG, the prior knowledge is leveraged explicitly by initializing the edge weights according to the relation matrix $\mathcal{R}$. On the contrary, we propose to utilize the prior knowledge implicitly in DPR-AG. Specifically, we design a hybrid loss function as:

$$\begin{aligned} L = &\sum_{i=1}^{N} \sum_{j \neq i} -\ln \sigma \left(MLP\left([\mathbf{u}_i || \mathbf{g}_i]\right) - MLP\left([\mathbf{u}_i || \mathbf{g}_j]\right)\right) \\ &- \sum_{i=1}^{N} \sum_{\substack{u,v \in \mathcal{G}_i \\ \mathcal{R}_{uv} \neq -1}} \ln \left(softmax\left(\mathbf{e}_{vu}^{\top} \mathbf{Q}\right)_{\mathcal{R}_{uv}} \right) + \lambda \left\|\Theta\right\|_2^2, \end{aligned} \tag{21}$$

where the MLP model is the final prediction model. $\mathbf{Q} \in \mathbb{R}^{D \times 3}$ is the transfer matrix to transform the edge attribute $\mathbf{e}_{vu}$ into classification probabilities, where D is the dimension of $\mathbf{e}_{vu}$. We add cross entropy loss to the loss function, which aims to force the edge attribute $\mathbf{e}_{vu}$ to contain the interaction type information.

5 EXPERIMENTS

In this section, we evaluate the proposed model with a number of competitive baselines. Meanwhile, many discussions and case studies on drug package recommendation will be presented.

5.1 Experimental Settings

We omit the dataset description in this section since it has been introduced in Section 3.1. Other experimental settings will be described in the following parts.

5.1.1 Baselines and Evaluation Metrics. To evaluate the performance of our models for drug package recommendation, we selected a number of state-of-art methods as baselines. Specifically, we first chose two popular traditional recommendation approaches, and several state-of-art package recommendation models as follows:

- **NCF** [16]: NCF is a state of art deep neural networks on recommendation system, which replacing the inner product in matrix factorization with a neural architecture. This model recommends top K drugs as packages for the patients in test sets based on the patient embeddings, where K is the average size of drug packages.
- **NN**: This method utilizes the pretrained patient embeddings based on NCF, and returns the drug package corresponding to the Nearset Neighbor (NN) by calculating the cosine similarity of patient embeddings.
- **Package2vec**: [36] proposes to utilize Item2vec [4] for enhancing the item embeddings in a package , and we extend Item2vec following [23] to get the embedding of a package. NCF framework and BPR loss are utilized to train the package recommendation model.
- **LDA** [5]: This method utilizes the LDA model to get the embedding of a package and uses the same framework as Package2vec to recommend packages.
- **BR** [33]: BR is a package recommendation method which aggregates item latent vectors to get the package embeddings based on package size and item compatibility.
- **DAM** [7]: DAM is the state-of-art neural network architecture for package recommendation which utilizes factorized attention network to get the embedding of packages.
- **GNN**: This method is a simplified variant of our models, which only uses the package graph structure and ignore the edge attributes.

It is worth noting that the drug package recommendation is much different from general recommendation since there is no fixed users in our task. Therefore, in all of the baseline methods, we exploited the patient embedding model proposed in Section 4.1 to get the representation of patients. Another problem is how to generate packages for patients in test set since most of the models are discriminant. Therefore, we proposed that except for the NCF model which can generate packages itself, all the remaining

Table 4: The performance of each model.

model	Precision	Recall	F1-score
NCF	0.3812	0.5442	0.4200
NN	0.4890	0.4985	0.4732
Package2vec	0.4846	0.5268	0.4857
LDA	0.5014	0.5219	0.4904
BR	0.5068	0.5106	0.4879
DAM	0.5254	0.5107	0.4979
GNN	0.5085	0.5288	0.5009
DPR-WG	0.5133	**0.5488**	0.5137
DPR-AG	**0.5260**	0.5407	**0.5162**

models only pick out the best package from a candidate set, and the candidate set consists of drug packages from 10 most similar patients. The similarity was calculated by the cosine similarity between patient embeddings. Evaluation metrics including Precision, Recall and F1-score were utilized to compare the performance of the models.

5.1.2 Implementation Details. We implemented our model by PyTorch[3] and Pytorch Geometric[4]. The parameters were all initialized using Kaiming [15] initialization. For the pre-training model, we set the output dimension of the MLP, the dimension of char embeddings, and the hidden size of the LSTM as 32, while the dimension of patient embeddings was set as 64. For the construction of package graph, we set the threshold value of co-occurrence proportion as 0.01. For the BPR loss used in this paper, we used negative sampling to train the model and set the negative sampling ratio as 10, which means 10 negative samples for one positive sample. For all the MLP models used in this paper, we set the dimension of hidden layers as 128. In the process of model training, we used the Adam optimizer [17] for parameter optimization. We set learning rate as 0.001 and mini-batch size as 256. The parameters of baselines were set up similarly as our method and were all tuned to be optimal to ensure fair comparisons. For the dataset splitting, we divided our dataset into 80%/10%/10% training/validation/test and we report performance on the test set for the model that performed best on the validation set.

5.2 Discussions

5.2.1 Overall Performance. To demonstrate the effectiveness of our drug package recommendation framework, we compared DPR-WG and DPR-AG with all the baselines, and the results are shown in Table 4. From the results, we can get several observations:

First, the performance of our models surpasses most of the baseline methods on different evaluation metrics. This clearly proves the effectiveness of our DPR framework based on package graph construction and message passing neural networks. Furthermore, our models obtain much higher recall than baselines, which indicates our models are more likely to prevent doctors from neglecting certain factors in practical application.

Second, the performance of NCF model is the worst, since this method based on collaborative filtering prefers to recommend items with higher popularity, and cannot model the drugs as a whole. This clearly verifies the necessity for the studies of package recommendation systems.

Third, the GNN model which only leverages the graph topological structure to exchange information between different drugs cannot achieve comparable result with our model. However, this model surpasses all the other baselines. This verifies the effectiveness of constructing package graphs to capture the interaction between drugs, and futher indicates the effectiveness of our method for the graph induction process.

Last but not least, the results of the models except NCF are close to each other, since patients with similar condition are more likely to use similar drugs.

[3]https://pytorch.org/

[4]https://github.com/rusty1s/pytorch_geometric

Table 5: The results of ablation study.

model	Precision	Recall	F1-score
DPR-WG-Context	0.5126	0.5330	0.5053
DPR-WG-Type	0.5126	0.5377	0.5074
DPR-AG-Mask	0.5152	0.5342	0.5061
DPR-AG-Type	0.5154	0.5317	0.5056
DPR-WG	0.5133	0.5488	0.5137
DPR-AG	0.5260	0.5407	0.5162

5.2.2 Ablation Study. To further validate the effectiveness of each component of our models, we also designed some simplified variants of our models as follows:

- **DPR-WG-Context**: This method is a simplified variant of DPR-WG which only utilizes the edge attributes initialized by the drug interaction matrix and ignores the influence of the patient condition.
- **DPR-WG-Type**: This method is a simplified variant of DPR-WG which only uses the contextual impact factor as edge attributes and ignores the drug interaction type.
- **DPR-AG-Mask**: This method is a simplified variant of DPR-AG which deletes the mask layer in the calculation process.
- **DPR-AG-Type**: This method is a simplified variant of DPR-AG which deletes the cross entropy loss in the loss function. In this way, the edge attributes dose not contain the information of drug interaction type.

The results of ablation study are shown in Table 5 from which we can draw the following conclusions. First, DPR-WG performs better than the two variants. This indicates that both the contextual impact factors and the initial edge weights are significant, which clearly verifies our assumption that patient condition will influence the interaction effect between drugs. Second, DPR-AG also performs better than the two variants, which verifies that both parts of drug interaction type and mask vectors are effectual, and the mask layer we proposed can effectively extract the feature of patient condition.

5.3 Parameter Sensitivity

We investigated the sensitivity of our model parameter in this section. First, we evaluated how the threshold for co-occurrence proportion affected the performance, and the results are shown in Figure 6. From the results, we can find that as the number of edges decreases, the model performance does not change significantly,

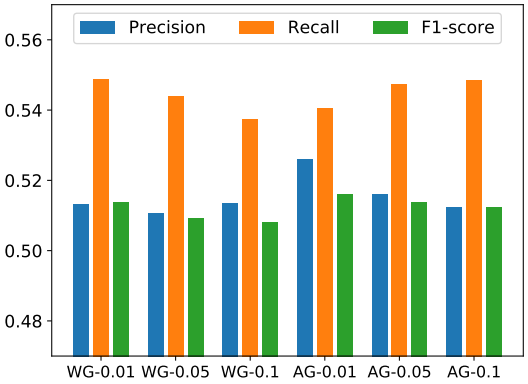
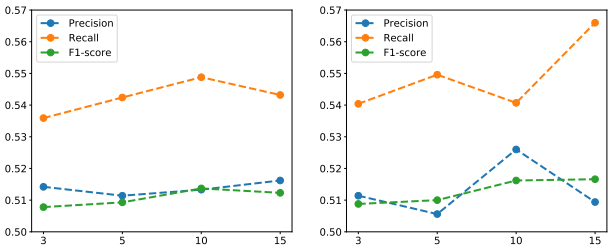


Figure 6: The preformance of our models with different co-occurrence proportion threshold.

Table 6: The preformance of our models with different number of GNN layers.

model	Precision	Recall	F1-score
DPR-WG-1	0.5133	0.5488	0.5137
DPR-WG-2	0.4994	0.5582	0.5100
DPR-AG-1	0.5260	0.5407	0.5162
DPR-AG-2	0.5139	0.5457	0.5128



(a) DPR-WG (b) DPR-AG

Figure 7: The performance of DPR-WG and DPR-AG with different number of negative samples.

and the F1-score shows a downward trend. This indicates the fact that there is no interaction between most of the drug pairs.

Next, we investigated whether utilizing two GNN layers can affect the results. Table 6 shows the results of our two models with one and two GNN layers. The results have not witnessed a performance improvement by adding one more GNN layer. As mentioned before, different from general graphs, we only need one GNN layer to extract almost all the information we expect since the drug package graph is dense enough.

Finally, we verified the impact of the negative sampling ratio. As shown in Figure 7, we can find the performance only fluctuates in a small range, and the model with a small negative sample number also works well in practice. All the above experiments have proved the robustness of the methods proposed in this paper.

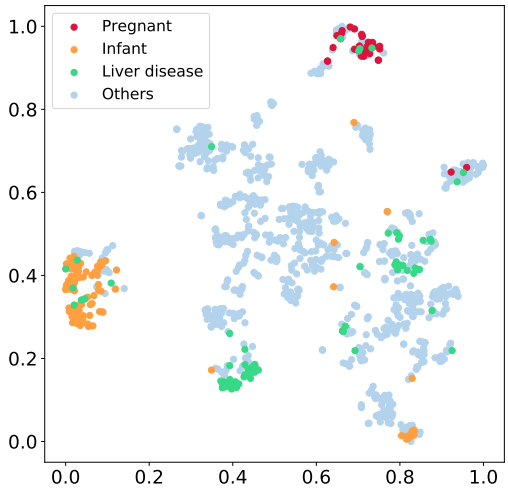


Figure 8: Visualization of mask vectors.

5.4 Case Study

In this part, we present some cases to illustrate the effectiveness of our models and reveal some interesting medical rules based on the derived insights on patient conditions and drug interaction.

5.4.1 Mask Vector Analysis. As mentioned before, we extracted the mask vector $\sigma(MLP(\mathbf{u}))$ of patient u to describe the impact of the patient condition. In order to analyze the effect of the mask vectors, we randomly selected 1,000 patients and their corresponding mask vectors, and projected them into two-dimensional space with t-SNE, which is proposed in [29]. We further selected three representative patient groups with special needs for drugs based on common sense, respectively pregnant women, infants (or young children) and patients with liver disease.

Figure 8 shows the visualization result. We can find that the mask vectors of infants and pregnant women deviate the most from the vectors of other patients, which indicates that these two groups have the most special requirements for drug selecting, and this is consistent with our common sense. Moreover, the mask vectors of patients with liver disease are also relatively deviated from other patients, but the degree of aggregation is lower than previous two groups. This indicates that patients with liver disease have special needs for drugs, but there are also certain personalized needs. We can further study the impact of patient conditions on drug selection by statistical methods such as clusting, which shows a great possibility of our method to help medical researchers.

5.4.2 Contextual Impact Factor Analysis. In Section 4.3.1, we propose to utilize contextual impact factors to reflect the impact of patient condition on drug interaction. In this section we will show how these impact factors play a role for recommending packages.

We picked patient #27667 for detailed analysis. From the EMR database we can know that this patient suffered from gallstones and came to the hospital for cholecystectomy. We input the ground truth drug package into DPR-WG, and got the contextual impact factors between all of the drug pairs. Table 7 shows three samples of drug pairs with different interaction types. From the results we

Table 7: Contextual Impact Factor Analysis for Patient #27667.

Drug 1	Drug 2	Description	Type	Factor
Potassium Chloride	Cefazolin	drug 2 may decrease the excretion rate of drug 1.	Synergism	0.993
Midazolam	Potassium Chloride	drug 1 may decrease the excretion rate of drug 2.	Synergism	-0.264
Ephedrine	Methylprednisolone	drug 1 may increase the excretion rate of drug 2.	Antagonism	-0.309

Table 8: Edge Attribute Analysis for Patient #25256.

Drug 1	Drug 2	Type	$softmax\left(\mathbf{e}_{vu}^{\top}\mathbf{Q}\right)$	$softmax\left(\hat{\mathbf{e}}_{vu}^{\top}\mathbf{Q}\right)$
Warfarin	Ondansetron	Synergism	[0.007, 0.807, 0.184]	[0.015, 0.923, 0.061]
Metformin	Spironolactone	Antagonism	[0.358, 0.163, 0.478]	[0.769, 0.022, 0.208]

can get several observations. First, Cefazolin can bring Potassium Chloride to a higher serum level, and the contextual impact factor for this edge is very high. This shows that our model believes the synergism between these two drugs is necessary for this patient. Second, Midazolam can also bring Potassium Chloride to a higher serum level, but our model gives a small negative factor for this interaction. By understanding more medical knowledge, we know that the combination of these two drugs has a greater risk, and they are even used for euthanasia. Our model predicts the risk of these two drugs. Finally, Ephedrine may decrease the effect of Methylprednisolone, but Methylprednisolone is an adrenal glucocorticoid with strong anti-inflammatory effect, which is very necessary for this patient. So our model gives a small negative factor to adjust the interaction effect between these two drugs. The above examples strongly confirm the effectiveness and interpretability of DPR-WG from different perspectives.

5.4.3 Edge Attribute Analysis. In Section 4.3.2, edge attribute vectors are calculated to describe the interaction between two drugs. The attribute vectors are forced to contain drug interaction category information, and mask vectors are utilized to bring the impact of patient condition. We propose that the mask vector plays a role by feature selecting. If we multiply a contextual edge attribute vector $\mathbf{e}_{vu}$ with the classification transfer matrix $\mathbf{Q}$, we can get a personalized drug interaction classification result, and we will illustrate this intuition in this case study.

We picked patient #25256 for detailed analysis. This patient was a 79-year-old woman with high blood pressure, diabetes and heart disease. We got the corresponding patient mask vector and drug interaction vectors by DPR-AG, and we further got the non-personalized and personalized drug interaction classification results for the drug interaction vectors. Table 8 shows two examples for this. We can find that Warfarin and Ondansetron have a synergistic effect, and the initial drug interaction vector reflects this point. Furthermore, the mask vector enhances this feature, since Warfarin (which can prevent the formation and development of thrombus) is very important for this patient, and it is beneficial to keep this synergistic effect. In addition, Metformin and Spironolactone are marked as antagonistic, but it is not significantly reflected in the drug interaction vector, and the mask vector believes that these two drugs may have no interaction. To explain this, we consulted a doctor and learned that the interaction between these two drugs is relatively moderate, and they are often used together clinically. These examples clearly confirm the interpretability and learning ability of our model.

Table 9: The results of package generation.

model	Non-heuristic	Heuristic
doctor1	39%	61%
doctor2	37%	63%
doctor3	39%	61%
doctor4	45%	55%
doctor5	30%	70%
average	38%	62%

6 PACKAGE GENERATION

Until now, we have considered recommending drug packages that already exist within the EMR database. However, existing packages cannot meet the needs of new patients sometimes. Therefore, we present a heuristic algorithm which combines the existing packages, personalized drug prediction lists and drug interaction matrix to generate new packages. The algorithm is described as follows.

First, we get the drug frequency rank list L which contains drugs in descending order of occurrence frequency in the EMR dataset. Then, we calculate the drug co-occurrence proportion matrix M which is mentioned in Section 4.2. For a new patient, we can get the patient embedding based on the patient's description. With the patient embedding, we can get the candidate set S_1 from similar patients as previously mentioned, and we can get the personalized prediction list l of all drugs by utilizing the NCF model obtained in the pre-training phase, which contains drugs in descending order of predict value. It is worth noting that, as shown in Section 5.2, the top drugs in l can be incorrect. Finally, start with the initial candidate set S_1, we can get new drug packages as:

(1) Form a new candidate set S_2 based on reforming the packages in S_1 by the following ways:
   - Delete the drugs that only appear in a small number of packages in S_1 and rank low in l;
   - Add the drugs that rank low in L and rank high in l, which means these drugs are not recommended just because they have high popularity.

Table 10: Examples for the heuristic method.

Patient ID	Ground Truth	Non-heuristic	Heuristic
#28062	Glucose, Isoniazid, Silybin, Kanamycin, Rifampin, Levofloxacin, Aspirin, Clindamycin, Pyridoxine	Glucose, Levofloxacin Clindamycin, Moxifloxacin	Glucose, Levofloxacin Clindamycin, Moxifloxacin **Isoniazid**, **Silybin**, **Pyridoxine**
#28199	Carboprost methyl ester, Cefuroxime Sodium, Aminomethylbenzene, Hydroxyethyl starch, Peptide hormones,Lidocaine	Lidocaine, Glucose, Cefuroxime, Aminomethylbenzene, Carboprost methyl ester, Carprost tromethamine, Peptide hormones, Reserpine, Oxytocin	Carboprost methyl ester, Aminomethylbenzene, Misoprostol, **Hydroxyethyl starch**, **Peptide hormones**, **Lidocaine**

(2) Generate candidate set S_3 by modifying the drugs in S_2 using more radical strategies as:
- If drug d ranks high in l and has synergism relationship with a drug in package p, then add drug d to package p;
- If drug d ranks high in l and has high co-occurrence proportion with a drug in package p, then add drug d to package p;
- If drug d_1 and d_2 in package p have antagonism relationship and low co-occurrence proportion, then delete the drug with lower lank in l;

(3) returns final candidate set $S = S_1 \cup S_2 \cup S_3$.

We verified the effectiveness of our heuristic algorithm on DPR-WG, where the non-heuristic model selected the best package from the initial candidate set S_1, and the heuristic model selected best package from S. Due to the hidden security risks of directly using the generated package, we randomly selected some test samples and handed them to five doctors to mark the packages they preferred.

The results are shown in Table 9, where the percentages reflect the ratio of the doctors' choice. From the results, we can find that utilizing the drug packages generated by the heuristic algorithm can significantly improve the performance of drug package recommendation. Furthermore, we picked two examples to illustrate the effect of the heuristic method. The examples are shown in Table 10, where patient #28062 is a patient with tuberculosis, and patient #28199 is a pregnant woman. We can find that both the adding and deleting strategies are effective. For the adding strategy, Isoniazid was added since the rank promotion between list L and l, and Pyridoxine was added because of the synergism interaction with Levofloxacin in the first example. For the deleting strategy, several incorrect drugs were deleted in the second example. All the results confirm the effectiveness of our package generation method.

7 CONCLUSION

In this paper, we studied the problem of drug package recommendation. Specifically, we first designed a pre-training method based on neural collaborative filtering to get the initial embedding of patients and drugs. Then, the drug interaction graph was initialized based on medical records and domain knowledge. Furthermore, we proposed a new drug package recommendation framework with two variants, respectively DPR-WG and DPR-AG to solve the problem, in which each the interactions was described as signed weights or attribute vectors. Finally, extensive experiments on a real-world data set from a first-rate hospital demonstrated the effectiveness of our DPR framework compared with several competitive baseline methods, and further supported the heuristic study for the drug package generation task with adequate performance.

ACKNOWLEDGMENTS

This research was partially supported by grants from the National Key Research and Development Program of China (Grant No.2018YF B1402600), the National Natural Science Foundation of China (Grant No.62072423), and the Key Research and Development Program of Anhui Province (No.1804b06020377).

REFERENCES

[1] Gediminas Adomavicius and Alexander Tuzhilin. 2005. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE transactions on knowledge and data engineering* 17, 6 (2005), 734–749.
[2] Naveed Afzal, Sunghwan Sohn, Sara Abram, Christopher G Scott, Rajeev Chaudhry, Hongfang Liu, Iftikhar J Kullo, and Adelaide M Arruda-Olson. 2017. Mining peripheral arterial disease cases from narrative clinical notes using natural language processing. *Journal of vascular surgery* 65, 6 (2017), 1753–1761.
[3] Jinze Bai, Chang Zhou, Junshuai Song, Xiaoru Qu, Weiting An, Zhao Li, and Jun Gao. 2019. Personalized Bundle List Recommendation. In *The World Wide Web Conference* (San Francisco, CA, USA) *(WWW '19)*. Association for Computing Machinery, New York, NY, USA, 60–71.
[4] Oren Barkan and Noam Koenigstein. 2016. Item2vec: neural item embedding for collaborative filtering. In *2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE, 1–6.
[5] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research* 3, Jan (2003), 993–1022.
[6] Jianxin Chang, Chen Gao, Xiangnan He, Depeng Jin, and Yong Li. 2020. Bundle Recommendation with Graph Convolutional Networks. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1673–1676.
[7] Liang Chen, Yang Liu, Xiangnan He, Lianli Gao, and Zibin Zheng. 2019. Matching User with Item Set: Collaborative Bundle Recommendation with Deep Attention Network. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*. International Joint Conferences on Artificial Intelligence Organization, 2095–2101.
[8] Wen-Hao Chiang, Li Shen, Lang Li, and Xia Ning. 2018. Drug recommendation toward safe polypharmacy. *arXiv preprint arXiv:1803.03185* (2018).
[9] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555* (2014).
[10] Ting Deng, Wenfei Fan, and Floris Geerts. 2013. On the complexity of package recommendation problems. *SIAM J. Comput.* 42, 5 (2013), 1940–1986.
[11] Steven E Dilsizian and Eliot L Siegel. 2014. Artificial intelligence in medicine and cardiac imaging: harnessing big data and advanced computing to provide personalized medical diagnosis and treatment. *Current cardiology reports* 16, 1 (2014), 441.
[12] Francisca García-Lizana and Antonio Sarría-Santamera. 2007. New technologies for chronic disease management and control: a systematic review. *Journal of telemedicine and telecare* 13, 2 (2007), 62–68.

[13] Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. 2017. Neural Message Passing for Quantum Chemistry. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70* (Sydney, NSW, Australia) *(ICML'17)*. JMLR.org, 1263–1272.

[14] William L. Hamilton, Rex Ying, and Jure Leskovec. 2017. Inductive Representation Learning on Large Graphs. In *Proceedings of the 31st International Conference on Neural Information Processing Systems* (Long Beach, California, USA) *(NIPS'17)*. Curran Associates Inc., Red Hook, NY, USA, 1025–1035.

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*. 1026–1034.

[16] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural Collaborative Filtering. In *Proceedings of the 26th International Conference on World Wide Web* (Perth, Australia) *(WWW '17)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 173–182.

[17] Diederik P. Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Yoshua Bengio and Yann LeCun (Eds.). http://arxiv.org/abs/1412.6980

[18] Thomas N. Kipf and Max Welling. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net.

[19] Yehuda Koren, Robert Bell, and Chris Volinsky. 2009. Matrix factorization techniques for recommender systems. *Computer* 42, 8 (2009), 30–37.

[20] Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer. 2016. Neural Architectures for Named Entity Recognition. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, San Diego, California, 260–270. https://doi.org/10.18653/v1/N16-1030

[21] Theodoros Lappas, Kun Liu, and Evimaria Terzi. 2009. Finding a team of experts in social networks. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. 467–476.

[22] Vivian Law, Craig Knox, Yannick Djoumbou, Tim Jewison, An Chi Guo, Yifeng Liu, Adam Maciejewski, David Arndt, Michael Wilson, Vanessa Neveu, et al. 2014. DrugBank 4.0: shedding new light on drug metabolism. *Nucleic acids research* 42, D1 (2014), D1091–D1097.

[23] Quoc Le and Tomas Mikolov. 2014. Distributed Representations of Sentences and Documents. In *Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32* (Beijing, China) *(ICML'14)*. JMLR.org, II–1188–II–1196.

[24] Shuangli Li, Jingbo Zhou, Tong Xu, Hao Liu, Xinjiang Lu, and Hui Xiong. 2020. *Competitive Analysis for Points of Interest*. Association for Computing Machinery, New York, NY, USA, 1265–1274.

[25] Yujia Li, Daniel Tarlow, Marc Brockschmidt, and Richard S. Zemel. 2016. Gated Graph Sequence Neural Networks. In *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*, Yoshua Bengio and Yann LeCun (Eds.).

[26] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I Sánchez. 2017. A survey on deep learning in medical image analysis. *Medical image analysis* 42 (2017), 60–88.

[27] Qi Liu, Yong Ge, Zhongmou Li, Enhong Chen, and Hui Xiong. 2011. Personalized Travel Package Recommendation. In *11th IEEE International Conference on Data Mining, ICDM 2011, Vancouver, BC, Canada, December 11-14, 2011*, Diane J. Cook, Jian Pei, Wei Wang, Osmar R. Zaïane, and Xindong Wu (Eds.). IEEE Computer Society, 407–416.

[28] Yang Liu, Zhi Li, Wei Huang, Tong Xu, and En-Hong Chen. 2020. Exploiting Structural and Temporal Influence for Dynamic Social-Aware Recommendation. *Journal of Computer Science and Technology* 35 (2020), 281–294.

[29] Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of machine learning research* 9, Nov (2008), 2579–2605.

[30] Suzanne Martin, Greg Kelly, W George Kernohan, Bernadette McCreight, and Christopher Nugent. 2008. Smart home technologies for health and social care support. *Cochrane database of systematic reviews* 4 (2008).

[31] Tamara P Miller, Yimei Li, Kelly D Getz, Jesse Dudley, Evanette Burrows, Jeffrey Pennington, Azada Ibrahimova, Brian T Fisher, Rochelle Bagatell, Alix E Seif, et al. 2017. Using electronic medical record data to report laboratory adverse events. *British journal of haematology* 177, 2 (2017), 283–286.

[32] Aditya G. Parameswaran and Hector Garcia-Molina. 2009. Recommendations with prerequisites. In *Proceedings of the 2009 ACM Conference on Recommender Systems, RecSys 2009, New York, NY, USA, October 23-25, 2009*, Lawrence D. Bergman, Alexander Tuzhilin, Robin D. Burke, Alexander Felfernig, and Lars Schmidt-Thieme (Eds.). ACM, 353–356.

[33] Apurva Pathak, Kshitiz Gupta, and Julian J. McAuley. 2017. Generating and Personalizing Bundle Recommendations on *Steam*. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, Shinjuku, Tokyo, Japan, August 7-11, 2017*, Noriko Kando, Tetsuya Sakai, Hideo Joho, Hang Li, Arjen P. de Vries, and Ryen W. White (Eds.). ACM, 1073–1076.

[34] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian Personalized Ranking from Implicit Feedback. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence* (Montreal, Quebec, Canada) *(UAI '09)*. AUAI Press, Arlington, Virginia, USA, 452–461.

[35] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. Graph Attention Networks. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net.

[36] Mengting Wan, Di Wang, Jie Liu, Paul Bennett, and Julian McAuley. 2018. Representing and recommending shopping baskets with complementarity, compatibility and loyalty. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. 1133–1142.

[37] Hao Wang, Tong Xu, Qi Liu, Defu Lian, Enhong Chen, Dongfang Du, Han Wu, and Wen Su. 2019. MCNE: An end-to-end framework for learning multiple conditional network representations of social network. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 1064–1072.

[38] Meng Wang, Mengyue Liu, Jun Liu, Sen Wang, Guodong Long, and Buyue Qian. 2017. Safe Medicine Recommendation via Medical Knowledge Graph Embedding. *CoRR* abs/1710.05980 (2017). arXiv:1710.05980

[39] Hong-Jian Xue, Xinyu Dai, Jianbing Zhang, Shujian Huang, and Jiajun Chen. 2017. Deep Matrix Factorization Models for Recommender Systems. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*, Carles Sierra (Ed.). ijcai.org, 3203–3209.

[40] Le Zhang, Tong Xu, Hengshu Zhu, Chuan Qin, Qingxin Meng, Hui Xiong, and Enhong Chen. 2020. Large-Scale Talent Flow Embedding for Company Competitive Analysis. In *WWW '20: The Web Conference 2020, Taipei, Taiwan, April 20-24, 2020*, Yennun Huang, Irwin King, Tie-Yan Liu, and Maarten van Steen (Eds.). ACM / IW3C2, 2354–2364.

[41] Ping Zhang, Fei Wang, Jianying Hu, and Robert Sorrentino. 2014. Towards personalized medicine: leveraging patient similarity and drug similarity analytics. *AMIA Summits on Translational Science Proceedings* 2014 (2014), 132.

[42] Shuai Zhang, Lina Yao, Aixin Sun, and Yi Tay. 2019. Deep learning based recommender system: A survey and new perspectives. *ACM Computing Surveys (CSUR)* 52, 1 (2019), 1–38.

[43] Yin Zhang, Daqiang Zhang, Mohammad Mehedi Hassan, Atif Alamri, and Limei Peng. 2015. CADRE: Cloud-assisted drug recommendation service for online pharmacies. *Mobile Networks and Applications* 20, 3 (2015), 348–355.

[44] Zhi Zheng, Tong Xu, Chuan Qin, Xiangwen Liao, Yi Zheng, Tongzhu Liu, and Guixian Tong. 2020. Multi-Source contextual collaborative recommendation for medicine. *Journal of computer research and development* 57, 8 (2020), 1741–1754.

[45] Tao Zhu, Patrick Harrington, Junjun Li, and Lei Tang. 2014. Bundle recommendation in ecommerce. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*. 657–666.