Personalized Prescription for Comorbidity

Lu Wang[1], Wei Zhang[1], Xiaofeng He[1(✉)], and Hongyuan Zha[2]

[1] School of Computer Science and Software Engineering,
East China Normal University, Shanghai, China
joywanglulu@163.com, zhangwei.thu2011@gmail.com, xfhe@sei.ecnu.edu.cn
[2] Georgia Institute of Technology, Atlanta, USA
zha@cc.gatech.edu

Abstract. Personalized medicine (PM) aiming at tailoring medical treatment to individual patient is critical in guiding precision prescription. An important challenge for PM is comorbidity due to the complex interrelation of diseases, medications and individual characteristics of the patient. To address this, we study the problem of PM for comorbidity and propose a neural network framework Deep Personalized Prescription for Comorbidity (PPC). PPC exploits multi-source information from massive electronic medical records (EMRs), such as demographic information and laboratory indicators, to support personalized prescription. Patient-level, disease-level and drug-level representations are simultaneously learned and fused with a trilinear method to achieve personalized prescription for comorbidity. Experiments on a publicly real world EMRs dataset demonstrate PPC outperforms state-of-the-art works.

Keywords: Personalized prescription · Deep learning
Multi-source fusion · Comorbidity

1 Introduction

Restricted by the traditional care delivery models, many doctors still prescribe therapies based on their own experience and population averages, which causes inefficient care for significant portions of patients [1]. As reported from the literature, 75% patients on average take ineffective cancer drugs and 70% patients take ineffective Alzheimer's drugs [2]. Personalized medicine (PM) which tailors the medical treatment to individual patient is promising to guide precision prescription [3]. An extremely important challenge for PM is comorbidity. Comorbidity stands for two or more complex disease conditions in the same patient and has complex interrelation of diseases, medications and individual characteristics of the patient [4,5]. Some researches show comorbidity is reported in 35% to 80% of all ill people [6,7]. In the United States, about 80% of medicare costs are caused by patients with 4 or more chronic diseases [8]. Recently with the availability of massive electronic medical records (EMRs), exploring the healthcare data has great potential to support intelligent personalized prescription for comorbidity.

J. Pei et al. (Eds.): DASFAA 2018, LNCS 10828, pp. 3–19, 2018.
https://doi.org/10.1007/978-3-319-91458-9_1

Researches about prescription based on EMRs are mainly divided into pattern-based and model-based approaches. Pattern-based methods recommend prescriptions by measuring the similarities among records of patients [9,10]. These methods are challenging to learn the relation of patients' information (e.g., disease, demographic information, lab information, etc) and medications. Model-based methods include decision-theoretic methods [11] and statistical methods [12]. But these methods only focus on one specific disease. Recently, two deep models are proposed to learn a nonlinear mapping from multiple diseases to multiple drugs based on EMRs [13,14], and achieve significant improvements. Without considering patient-specific information, these deep methods recommend constant-treatment for patients with same diseases. However, it is not in line with real situations. As shown in Table 1, the two patients are with the same diseases. Due to the different physiologic states, they take different treatments.

Table 1. The difference and intersection treatments of two patients with same diseases.

Diagnosis	Intersection treatments	Difference treatments
Pure hypercholesterolemia, Intermediate coronary syndrome, Hypertension NOS, Coronary atherosclerosis of native coronary artery	Meperidine, Neostigmine, Phenylephrine HCl, Ranitidine, Oxycodone-Acetaminophen, Metoclopramide, Calcium Gluconate, Glycopyrrolate, Magnesium Sulfate, Milk of Magnesia, Nitroglycerin, Aspirin EC, Acetaminophen, Sucralfate, Bisacodyl, Docusate Sodium, Potassium Chloride, Furosemide, Morphine Sulfate, Aspirin, Metoprolol	Propofol, Vancomycin HCl, Ibuprofen, Midazolam HCl, Chlorpheniramine Maleate, Hydrochlorothiazide, Hespan, Nitroprusside Sodium, Ondansetron, Diphenhydramine HCl
		CefazoLIN, Insulin Human Regular, Propofol, Docusate Sodium, Dextrose 50%, Insulin, Simvastatin, Sodium Chloride 0.9% Flush

There are two important issues remained in the aforementioned methods. (1) Non-personalized medicine. Existing methods for comorbidity ignore massive individual characteristics of the patient, such as demographic and laboratory information, which fail to recommend patient-specific prescription. (2) Lack of medical knowledge. Medical knowledge can guide us to learn a more effective and interpretable model. Furthermore, learning different "weights" of multiple diseases for comorbidity patients is also a difficult issue [15].

To tackle these issues, we integrate multi-source patient-specific information to learn patient-level representation. The representations and severities of multiple diseases are learned by employing medical knowledge and attention mechanism. The main contributions of this paper can be summarized as follows:

– To obtain the interdependencies among diseases, medications and individual characteristics of the patient, we design a deep learning model to integrate

multi-source information to learn the patient-level, disease-level and drug-level representations simultaneously, and fuse them with a trilinear method. (for comorbidity challenge)
- Patient-level representation is learned based on multiple patient-specific information, such as demographic and laboratory information (for issue 1). Disease-level representation is obtained by medical ontologies, where an attention mechanism is used to learn the different severities of multiple diseases (for issue 2).
- We evaluate our method over a real world EMRs MIMIC-3 and show that it outperforms state-of-the-art approaches for prescription.

The rest of this paper is organized as follows. We summarize the related work in Sect. 2. The proposed method is presented in Sect. 3. Experimental results and analysis are introduced in Sect. 4. We conclude our work in Sect. 5.

2 Related Work

Computational methods that leverage EMRs to support healthcare begin to draw attention in recent years. To learn good representations of diagnosis and prescription, several models from the fields, such as image processing and machine translation, are also leveraged to represent medical ontology.

Diagnosis is first handled by neural networks in 1989 [16]. Recently, deep models such as multi-layer perceptron (MLP) and recurrent neural networks (RNN) are applied to diagnose life-threatening diseases. Lipton et al. are the first to apply long short-term memory (LSTM) [17] to multi-label diagnoses, which takes the clinical variables as input to predict the diagnosis in intensive care unit setting [18]. A gated recurrent unit (GRU) [19] model is used to early detect heart failure with the row value of patients' records [20]. However, for the distinct tasks and different input, these methods can not be directly applied to prescription.

Prescription settled by pattern-based methods is to identify the treatments based on the similarities among records of patients [9,10,21]. As for model-based studies, Cheerla and Gevaert [12] use SVM to recommend proper treatments for pan-cancer patients with microRNA. Concurrently, Bajor and Lasko use a GRU model to predict the total medications for multiple diagnosis records of a patient to check the EMRs records [13]. However, the disease representations learned by Bajor et al. are not well aligned to the medical knowledge [13]. Zhang et al. also design a deep learning model LEAP to predict safe prescription with the input of multiple diseases [14]. Bajor's method and LEAP are established as state-of-the-art approaches, but they ignore the patient-specific information. These approaches are not effective for personalized prescription in comorbidity for: (1) due to the complex and abstruse correlation among multiple diseases, it is hard to measure their similarities; (2) ignoring the individual information of patients, the methods may recommend the same medications for patients with the same disease. As shown in Fig. 1, it is not in line with the real situation.

Neural Attention Model is designed for solving neural machine translation tasks which cause a bottleneck by using a fixed-length vector to represent a sentence [22]. To predict a target word, attention model automatically focuses on the related words in the source sentence. Recently, it is applied to image processing [23], dialog systems [24], machine translation [22] and popularity prediction [25]. Retain [26] is the pioneer work to apply attention mechanism to healthcare, which considers the historical visit records of patients in a reverse time to learn attentions of different visits.

Distributed Representation for language is proposed to predict the neighbors of a word using a simple neural network such as Skip-gram and Continuous Bag-of-Words (CBOW) [27]. In medical domain, Riccardo et al. propose an unsupervised method to learn the patients representations using a three-layer stack of denoising autoencoders [28]. To improve the interpretation of representations, GRAM employs an attention mechanism based on the hierarchical medical ontology to learn the representation of diseases and drugs [29]. However, GRAM overlooks the severity of diseases when the patients suffer from multiple diseases. Indeed, these works mainly focus on learning representation instead of prescription.

This paper extends prescription methods in a number of important dimensions, including: (1) a deep learning model to learn the patient-level, disease-level, drug-level representations simultaneously from multi-source information of EMRs to achieve patient-specific prescription for comorbidity, and (2) an effective representation of comorbidity learned by hierarchical disease ontologies and a neural attention model.

3 Personalized Prescription for Comorbidity

In this section, we first define the notations of medical ontology and EMRs data, followed by an overview of our approach. Then we introduce the detailed components of learning disease, patient and drug representations, and a fusion method to integrate these representations for personalized prescription.

3.1 Preliminaries

Considering a set of N patients $\mathbb{P} = \{p_1, p_2, ..., p_n, ..., p_N\}$, a patient p_n is specified by his or her patient-specific information P_n (demographic and laboratory information), diagnosis information D_n and medication information Y_n, where $P_n = \{p_n^{age}, p_n^{heartrate}, ...\}, D_n = \{d_1^n, d_2^n, ..., d_i^n, ..., d_I^n\}, Y_n = \{y_1^n, y_2^n, ..., y_k^n, ..., y_K^n\}$. d_i^n denotes the i-th disease in D_n and $y_k^n \in \{0, 1\}$ denotes whether a medication in the k-th medicine class treated for the patient p_n. G is a directed acyclic graph (DAG) of disease (coded in ICD-9) ontology[1]. We only focus on three main levels of ICD-9 ontology (1-digit nodes, 3-digit nodes and leaf-nodes) in this paper to ensure good generalization. Also, the three levels are often used to identify

[1] http://bioportal.bioontology.org/ontologies/ICD9CM.

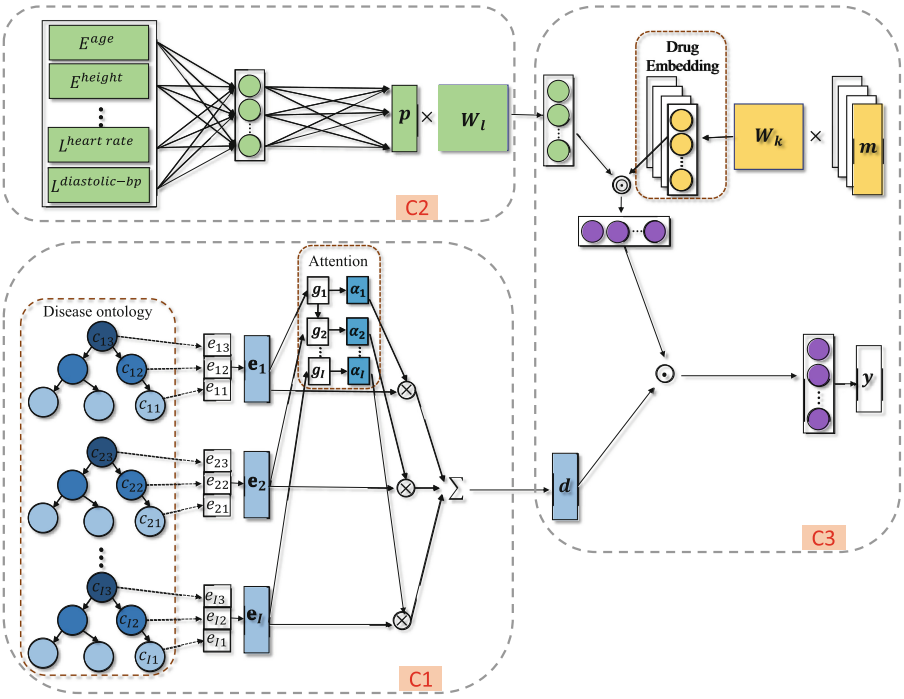

Fig. 1. General framework of PPC.

pharmacological subgroups. There is a hyponymy relation between the high level and low level nodes in G, where the leaf-node c_{i1} (in level-1) represents the i-th disease d_i^n, and the non-leaf nodes c_{i2} (in level-2), c_{i3} (in level-3) show a concept generalized from their child-nodes. Inspired by [29], each node in the three levels is associated with a basic embedding, where $\boldsymbol{e}_{ij}$ represents the basic embedding of node c_{ij} in j-th level.

PROBLEM DEFINITION (Personalized Prescription for Comorbidity.)
For a patient p_n, given his or her patient-specific information P_n and diseases D_n, where $P_n = \{p_n^{age}, p_n^{heartrate}, ...\}, D_n = \{d_1^n, d_2^n, ..., d_I^n\}$, the problem is to predict the personalized treatment Y_n ($Y_n = \{y_1^n, y_2^n, ..., y_K^n\}$) for the patient.

3.2 Algorithm Overview

As shown in Fig. 1, our approach is a deep learning model which includes three main components: **C1:** learning to represent the diagnosis, **C2:** learning to represent the patient, **C3:** fusing representations with a trilinear method.

PPC employs the hierarchical structure of disease ontology in knowledge graph G to learn a interpretable representation of disease d_i^n. It first finds a path from the leaf-node c_{i1} to the highest level node c_{i3} in G. Then, PPC concatenates

the basic embedding vectors $\boldsymbol{e}_{ij} \in \mathbb{R}^m$ $(j = 1, 2, 3)$ of the three nodes as the representation $\hat{\mathbf{e}}_i \in \mathbb{R}^{3m}$ of disease d_i^n. Combing the information of ancestors and children helps to learn a robust and comprehensive representation. Due to the patients in comorbidity have more than two diseases, playing more attention on severity diseases is beneficial to alleviate the symptoms. We use attention mechanism [22] to learn the different severities of diseases and represent the diagnosis of the patient as a single embedding $\mathbf{d}^n$. Simultaneously, we learn the patient representation using a 2-layer MLP with the input P_n, and the medication representation m_k $(k = 1, 2, ..., K)$ is learned by a 1-layer MLP. To learn the interdependencies among diseases, medications and the patient, a trilinear fusion method is adapted to integrate the three representations to predict the personalized treatments for comorbidity.

3.3 C1: Learning to Represent the Diagnosis

Diagnostic information in EMRs consists of the patients' diseases. Medical ontology in this paper is used to facilitate the representation of the diagnosis. We first concatenate the three basic embeddings into a single embedding $\hat{\mathbf{e}}_i \in \mathbb{R}^{3m}$:

$$\hat{\mathbf{e}}_i = [\boldsymbol{e}_{i1}, \boldsymbol{e}_{i2}, \boldsymbol{e}_{i3}], \tag{1}$$

$$\boldsymbol{e}_{i1} = \boldsymbol{W}_{emb1}\boldsymbol{c}_{i1}, \quad \boldsymbol{e}_{i2} = \boldsymbol{W}_{emb2}\boldsymbol{c}_{i2}, \quad \boldsymbol{e}_{i3} = \boldsymbol{W}_{emb3}\boldsymbol{c}_{i3},$$

where $\hat{\mathbf{e}}_i$ is the embedding of disease d_i^n, $\boldsymbol{c}_{ij} \in \mathbb{R}^D$ is the one-hot representation of node c_{ij} $(j = 1, 2, 3)$, $\boldsymbol{W}_{emb1}$, $\boldsymbol{W}_{emb2}$ and $\boldsymbol{W}_{emb3} \in \mathbb{R}^{m \times D}$ are the embedding matrixes corresponding to $\boldsymbol{c}_{i1}$, $\boldsymbol{c}_{i2}$ and $\boldsymbol{c}_{i3}$ respectively.

Then, we use the convex combination of multiple diseases to represent the diagnosis of the patient:

$$\mathbf{d}^n = \sum_{i=1}^{I} \alpha_i \hat{\mathbf{e}}_i, \quad \sum_{i=1}^{I} \alpha_i = 1, \alpha_i \geq 0 \qquad \text{for } d_i^n \in D_n, \tag{2}$$

where I is the number of diseases in D_n. α_i is the attention weight of the disease d_i^n, which also indicates the severity of d_i^n for the patient. The scalar α_i is generated as follows,

$$\alpha_i = \frac{\exp(f(\hat{\mathbf{e}}_i))}{\sum_{j=1}^{I} \exp(f(\hat{\mathbf{e}}_j))}. \tag{3}$$

Using a 1-layer GRU and a 1-layer MLP, we obtain $f(\hat{\mathbf{e}}_i)$ as follows,

$$(\mathbf{g}_1, ..., \mathbf{g}_i, .., \mathbf{g}_I) = \text{GRU}(\hat{\mathbf{e}}_1, ..., \hat{\mathbf{e}}_i, ..., \hat{\mathbf{e}}_I), \tag{4}$$

$$h_i = \mathbf{w}_k^{\mathrm{T}} \mathbf{g}_i + b_k, \tag{5}$$

$$f(\hat{\mathbf{e}}_i) = \tanh(\mathbf{w}_i[h_1,\ h_2, ...,\ h_I]^{\mathrm{T}} + b_i), \qquad for\ \ i = 1, 2, ..., I, \tag{6}$$

where $\mathbf{g}_i \in \mathbb{R}^p$ is the hidden layer of GRU[2]. h_i is the hidden layer of MLP and $\mathbf{w}_k \in \mathbb{R}^p$, b_k, $\mathbf{w}_i \in \mathbb{R}^I$, b_i are parameters to learn. The GRU layer learns the attentions of diseases separately, while the MLP learns the attentions of diseases jointly.

The final representation of diseases $\mathbf{d}_n \in \mathbb{R}^m$ can also be calculated by $\boldsymbol{C}_n \in \mathbb{R}^{3D \times I}$ as shown in Eq. (8), where $\boldsymbol{\alpha} \in \mathbb{R}^I$ is the attention vector. As shown in Eq. (9), $\mathbf{W}_{emb} \in \mathbb{R}^{m \times 3D}$ is the concatenation embedding matrix of disease ontologies in the three levels. Overall, we represent the diagnosis of patients by employing the hierarchical structure of disease ontologies in knowledge graph G and learning different severities of multiple diseases.

$$\mathbf{d}_n = \mathbf{W}_{emb}(\boldsymbol{C}_n\boldsymbol{\alpha}) \tag{7}$$

$$\boldsymbol{C}_n = [\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2, ..., \hat{\mathbf{c}}_I], \qquad \text{where } \hat{\mathbf{c}}_i = [\boldsymbol{c}_{i1}, \boldsymbol{c}_{i2}, \boldsymbol{c}_{i3}], \quad i = 1, 2, ..., I \tag{8}$$

$$\mathbf{W}_{emb} = [\boldsymbol{W}_{emb1}, \boldsymbol{W}_{emb2}, \boldsymbol{W}_{emb3}]. \tag{9}$$

3.4 C2: Learning to Represent the Patient

Demographic and laboratory information belongs to patient-specific indicators. Demographic information consists of age, gender, height, weight, language, ethnicity, etc. Laboratory indicators include blood pressure, temperature, blood oxygen saturation, etc. The patient-specific information is important to the design of therapeutic regimen and dosage.

The demographic information is denoted as E:

$$E = \{E^{age}, E^{height}, ..., E^{weight}\},$$

and the laboratory indicators are denoted as L:

$$L = \{L^{blood-pressure}, L^{temperature}, ..., L^{ph}\}.$$

Each element in E and L indicates a variable of P_n. Let $\widehat{\boldsymbol{p}}_n$ be the intermediate representation of patients where the discrete variables are represented as one-hot codes, and the continuous variables keep invariant.

We use a 2-layer MLP to learn the patient representation:

$$\mathbf{h}_z = f(\mathbf{W}_z\widehat{\boldsymbol{p}}_n + \mathbf{b}_z), \tag{10}$$

$$\mathbf{p}_n = f(\mathbf{W}_u\mathbf{h}_z + \mathbf{b}_u), \tag{11}$$

where $\mathbf{W}_z$ and $\mathbf{b}_z$ are the parameters of first layer, $\mathbf{W}_u$ and $\mathbf{b}_u$ are parameters of second layer, f is the activation function ReLUs, and $\mathbf{p}_n \in \mathbb{R}^n$ is the final representation of the patient.

[2] We have also examined LSTM and other activation functions to learn to represent diagnosis, but they have less efficiency and worse performance.

3.5 C3: Fusing Representations with Trilinear Method

We propose a trilinear fusion method to integrate different sources of information. The input of the trilinear fusion method consists of three types of variables: diagnosis $\boldsymbol{C}_n$, patient-specific information $\mathbf{p}_n$ and candidate medications $\mathbf{m}_k$ $(k = 1, 2, ..., K)$, where $\mathbf{m}_k \in \mathbb{R}^K$ is the one-hot representation of the medicine. $\boldsymbol{C}_n$ is the concatenation of one-hot representations of diseases as shown in Eq. (8). The trilinear fusion method characterizes such a specific treatment event by considering the interdependencies among medications, the patient and diagnosis. Assume $h_{k,n}$ is the index of the probability of the medication m_k recommended for the patient p_n, and the probability is shown in Eq. (13). The trilinear method is described as follows,

$$h_{k,n} = (\mathbf{W}_{emb}(\boldsymbol{C}_n\boldsymbol{\alpha}))^{\mathrm{T}}(\mathbf{W}_m\mathbf{m}_k \odot \mathbf{W}_l\mathbf{p}_n), \tag{12}$$

where $\odot$ denotes the element-wise multiplication and $\boldsymbol{\alpha}$, $\mathbf{W}_{emb}$, $\mathbf{W}_m \in \mathbb{R}^{m\times K}$, $\mathbf{W}_l \in \mathbb{R}^{m\times n}$ are parameters to learn. To predict whether to recommend drug m_k for patient p_n, we use a sigmoid function to predict the probability of recommending m_k as follows:

$$f_{k,n} = \frac{1}{1 + e^{-h_{k,n}}}. \tag{13}$$

3.6 Objective Optimization

To solve this multi-label problem, we optimize the loss function of the K labels simultaneously:

$$Loss = \frac{1}{N}\frac{1}{K}\sum_{n=1}^{N}\sum_{k=1}^{K} l(f_k(\mathbf{C}_n, \mathbf{m}_k, \mathbf{p}_n), y_{k,n}), \tag{14}$$

$$l(f_{k,n}(\mathbf{C}_n, \mathbf{m}_k, \mathbf{p}_n), y_{k,n}) = -(1 - y_{k,n}) * log(1 - f_{k,n}) - y_{k,n} * log(f_{k,n}), \tag{15}$$

where $l(f_{k,n}(\mathbf{C}_n, \mathbf{m}_k, \mathbf{p}_n), y_{k,n})$ is the cross-entropy loss, N is the number of patients in training set. If we believe the solutions with small parameters are more general, we may optionally add a l1-penalty term, which will often make the parameters be nonzero in only a few states to prevent overfitting[3].

4 Experiment

In this section, we conduct experiments to evaluate our proposed method. We first report the dataset and models for comparison, followed by quantitative and qualitative measurements. Quantitative measurements include the common multi-label metrics and mean Jaccard. Qualitative measurements focus on how well the presented method solves the issues mentioned in Sect. 1, such as personalized prescription analysis, the interpretable representation of diseases analysis and the effect of the diseases' severities learned by attention mechanism.

[3] We have examined both l1-norm and l2-norm, and find their performance are similar.

4.1 Dataset Description

The experiments are conducted on a public EMRs dataset MIMIC-3 [30]. MIMIC-3 contains 43K patients in critical care units during 2001 and 2012. There are 6,695 distinct diseases and 4,127 drugs in MIMIC-3. The median number of diseases of each record is 9 (Q1–Q3:6–15). Following the procedure adopted in [13], we extract the top 1,000 most medications and top 2,000 most diseases (ICD-9 codes) in the first 24 h after the admission of patients. Because the patient states always change after 24 h and the first 24 h are the most critical time of the patient. These medications and diseases cover 85.4% of all medication records and 95.3% of all disease records. The medications in patient's diagnosis records are coded in NDC[4]. To obtain the hierarchical information of medications, we map the medication code from NDC into the third level of ATC[5] using the public tool[6]. ATC is another medication code which is hierarchically structured by anatomic and therapeutic classes. Finally, we obtain 180 ATC codes, which is also the number of labels in our multi-label classification task.

For learning the patient representation, we choose 8 demographic features: gender, age, weight, height, religion, language, marital status and ethnicity and 11 clinical variables (followed by the physician's suggestion): diastolic blood pressure, Glascow coma scale, blood glucose, systolic blood pressure fraction of inspired O2, heart rate, pH, respiratory rate, blood oxygen saturation, body temperature, and urine output. These variables are first rescaled to z-scores, then rescaled to [0,1]. We extract the results of clinical variables in the first 24 h after the patients admitted to the intensive care unit. We further fill the missing values by sampling them from the clinically normal interval as defined by clinical physicians. It is reasonable because clinicians often think the variables are norm and do not measure them [18]. For good generalization, we remove the records with more than 10 missing variables. Finally, we obtain 39,260 patients, and randomly divide the dataset for training, validation and testing by the ratio of 80/10/10.

We use the common metrics of multi-label, which contains micro under the ROC curve (micro-AUC), macro under the ROC curve (macro-AUC), label ranking average precision score and label ranking loss to promise fair and honest evaluation [31,32]. Also, we use mean Jaccard to measure the combination of recommended drugs as [14]. Initial PPC and PPC are our proposed methods, while the others are baselines. We describe these methods in detail as follows:

- **Popularity-20 (POP-20):** This is a patten-based method, which considers the top-k most frequent medications prescribed for each disease as predictions. We set K to be 20 for its best performance on validation dataset.
- **Random Forest (RF):** This is a classical machine learning method for multi-label problem. To reduce the massive computation, we use scalar to represent the different diseases, and train the model with 180 independent

[4] http://www.fda.gov/Drugs/DevelopmentApprovalProcess/.
[5] http://www.whocc.no/atc/structure and principles/.
[6] https://www.nlm.nih.gov/research/umls/rxnorm/.

forests, each forest is trained with the total diseases and predict one of the 180 treatments.
- **LM** [13]**:** This is a non-personalized prescription method, which uses a 3-level-MLP to recommend treatments for patients. The goal of LM is to check the errors and omissions in EMRs. The input is historical diseases of the patient in the EMRs records. The output is a single vector which is used to predict the medications treated for the historical diseases of the patient. To test the performance of LM, we use the current diseases of a patient as input and predict the medications for current diseases.
- **LG** [13]**:** This model is with the same setting as LM. But it uses a GRU model instead of MLP.
- **LEAP** [14]**:** LEAP uses a MLP framework to train a multi-label model which uses multiple diseases to predict multiple medications and considers the dependence of medications.
- **Knowledge-based LM (LMK):** We extend LM by incorporating hierarchical structure of disease ontology. The results of LMK can be utilized to test the effectiveness of considering medical ontology.
- **Personalized-infor-based LM (LMKF):** We further extend LMK by concatenating demographic information of patients, clinical measurements and diseases together as input. The results of LMKF can be used to verify the benefit of considering the patient-specific information.
- **initial-PPC (i-PPC):** It is with the same setting as our model, except using GRAM [29] to learn the representation of diseases.
- **PPC:** This is the model proposed in this paper. We aim at comparing it with other methods to demonstrate its advantages in multi-aspects. The basic embeddings $e_{i,j}$ of i-PPC and PPC are both randomly initialized.

The main goals of this section is to answer the following core questions, which guide the design of the experiments.

1. **Prediction Accuracy:** Can patient-specific information support more accurate prescription than other non-personalized prescription? (for issue 1)
2. **Ablation Study:** What is the contribution of each factor (diagnosis, patient, medicine information) to PPC?
3. **Embedding Analysis:** Does medical knowledge help to learn a better representation? (for issue 2)
4. **Attention Analysis:** How well does PPC learn the different severities among diseases?

4.2 Prediction Accuracy

Table 2 shows the performance of aforementioned methods on MIMIC-3. LMK outperforms LM by 2%–5.8%. This result shows combing medical knowledge to learn representations of diseases is significant to improve the accuracy of prescription. LMKF outperforms LMK by 0.1%–2.4%, which verifies the precision treatment is benefit from patient-specific information. Moreover, PPC consistently outperforms other baselines. For non-personalized deep models, such as

LEAP, LG and LM, PPC achieves 0.7%–4.1% improvement, because patient-specific information can help to prescribe more effective medications by identifying different physiologic states and characteristics of patients. It also outperforms LMKF by 0.2%–4.2% because the trilinear fusion method endows PPC with ability of learning rich and integrated representations based on different sources of information. Compared to i-PPC, PPC also achieves better improvements, because learning the different severities can help PPC pay more attention to important diseases.

Table 2. Performance comparisons on test sets for comorbidity prescription (%).

Method	Micro-AUC	Macro-AUC	Label ranking avg. precision	Label ranking loss	Jaccard
POP-20	76.2	55.8	52.7	40.8	37.8
RF	88.3	71.8	60.2	9.8	38.5
LM	89.2	73.2	62.8	9.4	36.6
LG	91.7	77.3	67.0	7.8	39.3
LEAP	92.0	78.9	67.5	7.6	40.8
LMK	92.1	79.0	68.2	7.4	40.1
LMKF	92.2	81.4	68.3	7.1	40.5
i-PPC	92.7	81.0	68.6	7.0	41.3
PPC	**93.1**	**83.0**	**69.9**	**6.90**	**44.7**

As for the other baselines, POP-20 is not effective due to its incapability of learning relation between multiple diseases and medications. RF works poor than deep models, because it fails to learn high-level representations of diseases.

4.3 Ablation Study

We conduct ablation study here to verify the contributions of the three types of information employed in this study. More specifically, we denote PPC-m, PPC-d and PPC-p as the variants of PPC by removing medical information, diagnosis information, and patient-specific information respectively. As the results presented in Table 3, all the information makes a positive contribution to precision treatment, where the contribution of diagnosis information is the most significant.

4.4 Embedding Analysis

To evaluate the effectiveness of disease representations learned by PPC, we use t-SNE [33] to visualize the final embeddings of 2000 diseases in our experiments. As shown in Fig. 2, different colors correspond to different categories of diseases

Table 3. Factor contribution analysis for PPC (%).

Method	Micro-AUC	Macro-AUC	Label ranking avg. precision	Label ranking loss	Jaccard
PPC-m	92.9	82.7	68.5	7.3	44.1
PPC-d	89.5	70.8	61.1	9.8	37.7
PPC-p	92.3	81.3	68.6	7.1	43.2
PPC	**93.1**	**83.0**	**69.9**	**6.90**	**44.7**

in the highest level of G. The names of the categories are represented aside the color-bar. The result shows that the embeddings of diseases in different categories can be roughly separated. In addition, we randomly select two impact point sets in Fig. 2, where the blue digits indicate the leaf-ICD-9 codes. The result shows that the codes are indeed related to their neighbors. However, the most related codes are not with the shortest distance because of the insufficient data. In deed, training the embeddings always need sufficient data, for example, training Skip-gram requires large amount of documents.

4.5 Attention Analysis

The attentions of the diseases can be explained intuitively using a randomly chosen case. Case 1: a patient with 13 diseases and 39 drugs. As mentioned in Sect. 1, learning different "weights" of multiple diseases is still a significant problem to be well addressed. In this section, we validate the availability of diseases' attentions using the domain knowledge and the amount of medications.

Analysis Based on the Domain Knowledge: As verified by a doctor, this is a patient with two main diseases: Parkinson and Chronic airway obstruction. More specifically, the patient is with diseases and symptoms such as: Parkinson, Chronic airway obstruction, depression, constipation, eye infections, esophagitis, indigestion, pneumonia, respiratory failure and congestive heart failures. About 1/3 of Parkinson patients suffer from severe depression and may cause constipation, abnormal gastrointestinal motility, and some eye diseases. Therefore, part of these symptoms and diseases may be caused by Parkinson's disease. In addition, Parkinson's patients are difficult to clean up the sputum, who easily infect pneumonia. Chronic airway obstruction which is unrelated to Parkinson's disease, may cause pulmonary heart disease and lung inflammation. Thus in this case, the patient also suffer from pneumonia, respiratory failure and congestive heart failure. Overall, Parkinson's disease and Chronic airway obstruction are the main diseases in this case and most of the other diseases are complications. As shown in Fig. 3, Parkinson's disease achieves the most attention ($\alpha = 0.12$), while Chronic airway obstruction obtains the third ($\alpha = 0.094$).

Analysis Based on the Amount of Medications: As shown in Fig. 4, to validate our results, we choose level-1 ATC codes to represent the medications. The

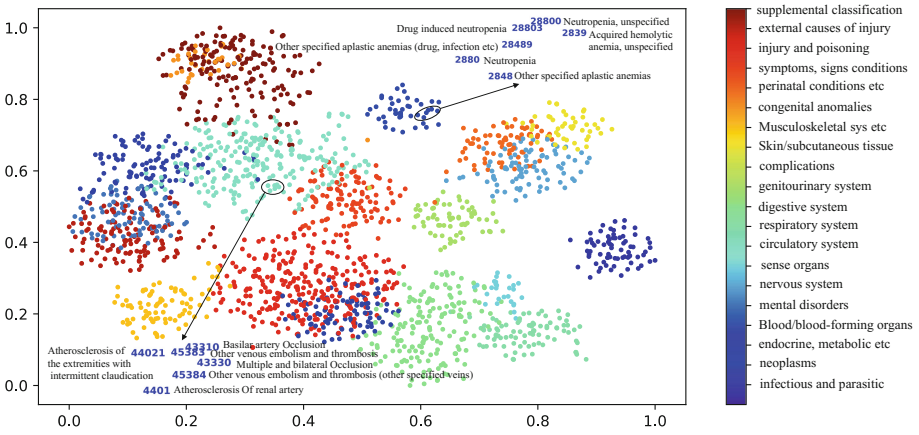

Fig. 2. The visualization (t-SNE, 2-D) of diseases' embeddings learned by PPC. Different colors correspond to different categories of disease in the highest level of G. The name of categories are represented aside the color-bar. The blue digits indicate the leaf-ICD-9 codes (diseases) of two randomly point sets. (Color figure online)

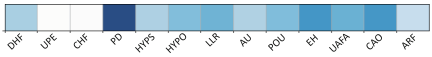

Fig. 3. Attentions learned from a comorbidity patient. Each rectangle represents a disease of this patient. The different color shades shows the volume of the attention of the disease. (DHF: Diastolic heart failure Acute on chronic, UPE: Unspecified pleural effusion, CHF: Congestive heart failure, PD: Parkinson's disease, HYPS: Hyposmolality and/or hyponatremia, HYPO: Hyperpotassemia, LLR: Lymphoid leukemia in remission, AU: Anemia, unspecified, POU: Pneumonia, organism unspecified, EH: Essential hypertension, UAFA: Upper arm and forearm Other cellulitis and abscess, CAO: Chronic airway obstruction, ARF: Acute respiratory failure.)

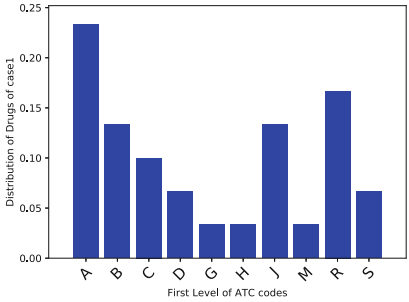

Fig. 4. Distribution of the number of drugs in this case. Abscissa represents the highest level of drug codes (ATC) of this case. The descriptions of partial codes are: A: Alimentary tract and metabolism, R: Respiratory system, J: Antiinfectives for systemic use, B: Blood and blood forming organs. C: Cardiovascular system.

largest amount of the drugs is mainly targeted for disease of Alimentary tract and metabolism (A). As mentioned before, Parkinson is most likely to cause gastrointestinal disease and constipation. So that these drugs are prescribed for symptom of Parkinson. The second largest amount of the drug is for respiratory system (R), such as Chronic airway obstruction and Pneumonia. These results also explain that the main diseases are Parkinson and Chronic airway obstruction, which is in line with our experiment results.

Table 4. Prescriptions for two patients with same diseases.

Diagnosis	Methods	Recommended treatments
Secondary malignant neoplasm of brain and spine, breast malignancy, Other convulsions, Secondary malignant neoplasm of lung, Hypertension Cerebral edema	PPC_{p1}	B05C, B05X, A10A, C08C, A02B, N03A, N02A, N02B, C02D, A12C, A06A, C03A, C03C
	$LEAP_{p1}$	B05C, B05X, A02B, N03A, N02A, N02B, A12C, A06A
	LG_{p1}	B05C, B05X, A10A, A02B, N02A, N02B, A12C, A06A
	PPC_{p2}	B05C, B05X, A10A, C08C, A02B, N03A, A04A, N02A, N02B, C02D, A12C, A06A
	$LEAP_{p2}$	B05C, B05X, A02B, N03A, N02A, N02B, A12C, A06A
	LG_{p2}	B05C, B05X, A10A, A02B, N02A, N02B, A12C, A06A

4.6 Personalized Prescription Analysis

With the subjectively examining performance on 30 randomly selected cases, we find the favorably performs of PPC comparing against other baselines. We choose one of these cases for analysis. In Fig. 5, we show 2 patients with same diseases, where the diseases and mediations recommend by 3 prescription methods are shown in Table 4. For the first patient, PPC_{p_1} recommends a set of medications with 78.6% coverage, where p_i $(i = 1, 2)$ represents the i-th patient. The recommendation coverage of LEAP_{p_1} and LG_{p_1} are both 42.9%. For the second patient, PPC_{p_2} recommends a set of medications with 100% coverage. In contrast, the coverage of LEAP_{p_2} and LG_{p_2} are 88.9% and 77.8%. The case is also the evidence of patient-specific medications. Due to the different physiologic states of patients, the mediations which the patients need are changed. In this case, the first patient is with systolic blood pressure 142 mmHg, while the second patient is 117 mmHg. Considering the patient-specific information, PPC recommends the drugs for p1 with C08C, C02D, C03A, C03C, which targets hypertension. For p2, these drugs were largely reduced. However, as shown

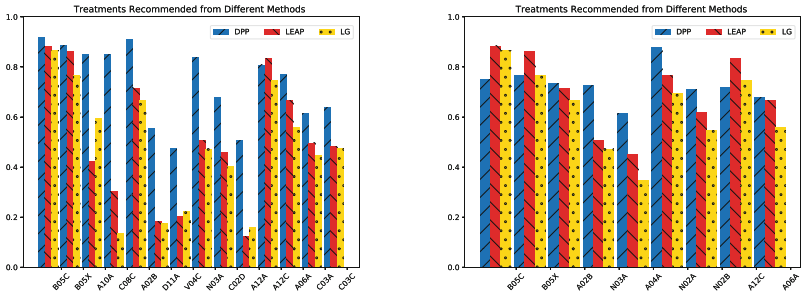

Fig. 5. Medication predictions confidence of two patients with same diseases. ATC codes on abscissa axis represents the prescriptions of doctors, where the hight of the bar indicates the prediction confidence of the three methods. We predict the medicine with the confidence >=0.5.

in Table 4, LEAP and LG that only consider diseases for prescription always recommend the same drugs for the patients with same diseases and ignore the hypertension states of the patients.

5 Conclusion

In order to solve the challenge and issues of personalized prescription for comorbidity, we propose an end-to-end deep learning model PPC. PPC integrates different sources of information to jointly learn representations of patients, diseases and medications and fuses them with a trilinear method to realize personalized prescription. Multiple patient-specific information is exploited to learn patient-level representation, and medical knowledge is combined to learn disease-level representation where an attention mechanism is used to learn different severities of comorbidity. Exploiting multi-source patient-specific information, PPC can recommend customized treatments which may be different for patients even having same diseases but different physiologic states, which achieves better results. Furthermore, PPC learns a good representation of disease and discriminates different severities of multiple diseases of comorbidity patients well. In the future, we will study how to solve the scalability issue for fuller set of medications.

Acknowledgements. This work was partially supported by the National Key Research and Development Program of China under Grant No. 2016YFB1000904, NSFC (61702190), NSFC-Zhejiang (U1609220), Shanghai Sailing Program (17YF1404500) and SHMEC (16CG24).

References

1. PMC2017: The personalized medicine opportunity, challenges, and the future (2017). http://www.personalizedmedicinecoalition.org/Userfiles/PMC-Corporate/file/The-Personalized-Medicine-Report1.pdf
2. Spear, B., Heath-Chiozzi, M., Huff, J.: Clinical application of pharmacogenetics. Trends Mol. Med. **7**(5), 201–204 (2001)
3. Berwick, D.M., Finkelstein, J.A.: Preparing medical students for the continual improvement of health and health care: Abraham flexner and the new public interest. Acad. Med. S56–S65 (2010)
4. Munoz, E., Rosner, F., Friedman, R., Sterman, H., Goldstein, J., Wise, L.: Financial risk, hospital cost, and complications and comorbidities in medical non-complications and comorbidity-stratified diagnosis-related groups. Am. J. Med. **84**(5), 933–939 (1988)
5. Jakovljević, M., Reiner, Ž., Miličić, D., Crnčević, Ž.: Comorbidity, multimorbidity and personalized psychosomatic medicine: epigenetics rolling on the horizon. Psychiatr. Danub. **22**(2), 184–189 (2010)
6. Taylor, A.W., Price, K., Gill, T.K., Adams, R., Pilkington, R., Carrangis, N., Shi, Z., Wilson, D.: Multimorbidity-not just an older person's issue. Results from an Australian biomedical study. BMC Public Health **10**(1), 718 (2010)
7. Bonavita, V., De Simone, R.: Towards a definition of comorbidity in the light of clinical complexity. Neurol. Sci. **29**(1), 99–102 (2008)
8. Valderas, J.M., Starfield, B., Sibbald, B., Salisbury, C., Roland, M.: Defining comorbidity: implications for understanding health and health services. Ann. Family Med. **7**(4), 357–363 (2009)
9. Sun, L., Liu, C., Guo, C., Xiong, H., Xie, Y.: Data-driven automatic treatment regimen development and recommendation. In: KDD, pp. 1865–1874 (2016)
10. Hu, J., Perer, A., Wang, F.: Data driven analytics for personalized healthcare. In: Weaver, C.A., Ball, M.J., Kim, G.R., Kiel, J.M. (eds.) Healthcare Information Management Systems. HI, pp. 529–554. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-20765-0_31
11. Lusted, L.B.: Introduction to medical decision making. Am. J. Phys. Med. Rehabil. **49**(5), 322 (1970)
12. Cheerla, N., Gevaert, O.: Microrna based pan-cancer diagnosis and treatment recommendation. BMC Bioinform. **18**(1), 32 (2017)
13. Bajor, J.M., Lasko, T.A.: Predicting medications from diagnostic codes with recurrent neural networks. In: ICLR (2017)
14. Zhang, Y., Chen, R., Tang, J., Stewart, W.F., Sun, J.: Leap: learning to prescribe effective and safe treatment combinations for multimorbidity. In: KDD, pp. 1315–1324 (2017)
15. Jakovljevi, M., Ostoji, L.: Comorbidity and multimorbidity in medicine today: challenges and opportunities for bringing separated branches of medicine closer to each other. Psychiatr Danub **25**, 18–28 (2013)
16. Hart, A., Wyatt, J.: Connectionist models in medicine: an investigation of their potential. In: Hunter, J., Cookson, J., Wyatt, J. (eds.) AIME 89, pp. 115–124. Springer, Heidelberg (1989). https://doi.org/10.1007/978-3-642-93437-7_15
17. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
18. Lipton, Z.C., Kale, D.C., Elkan, C., Wetzell, R.: Learning to diagnose with LSTM recurrent neural networks. In: ICLR (2016)

19. Chung, J., Gulcehre, C., Cho, K.H., Bengio, Y.: Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555 (2014)
20. Choi, E., Bahadori, M.T., Schuetz, A., Stewart, W.F., Sun, J.: Doctor AI: predicting clinical events via recurrent neural networks. In: Machine Learning for Healthcare Conference, pp. 301–318 (2016)
21. Zhang, P., Wang, F., Hu, J., Sorrentino, R.: Towards personalized medicine: leveraging patient similarity and drug similarity analytics. In: AMIA Joint Summits on Translational Science Proceedings, p. 132 (2014)
22. Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473 (2014)
23. Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhudinov, R., Zemel, R., Bengio, Y.: Show, attend and tell: neural image caption generation with visual attention. In: ICML, pp. 2048–2057 (2015)
24. Yang, L., Ai, Q., Guo, J., Croft, W.B.: aNMM: ranking short answer texts with attention-based neural matching model. In: CIKM, pp. 287–296 (2016)
25. Zhang, W., Wang, W., Wang, J., Zha, H.: User-guided hierarchical attention network for multi-modal social image popularity prediction. In: WWW (2018)
26. Choi, E., Bahadori, M.T., Sun, J., Kulas, J., Schuetz, A., Stewart, W.: Retain: an interpretable predictive model for healthcare using reverse time attention mechanism. In: NIPS, pp. 3504–3512 (2016)
27. Rumelhart, D.E., Hinton, G.E., McClelland, J.L., et al.: A general framework for parallel distributed processing. Parallel Distrib. Process.: Explor. Microstruct. Cogn. **1**, 45–76 (1986)
28. Riccardo, M., Li, L., Kidd, B.A., Dudley, J.T.: Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. Sci. Rep. **6**, 26094 (2016)
29. Choi, E., Bahadori, M.T., Song, L., Stewart, W.F., Sun, J.: GRAM: graph-based attention model for healthcare representation learning. In: KDD, pp. 787–795 (2017)
30. Johnson, A.E.W., Pollard, T.J., Shen, L., Lehman, L.H., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Celi, L.A., Mark, R.G.: MIMIC-III, a freely accessible critical care database. Sci. Data **3**, 160035 EP (2016)
31. Zhang, M.L., Zhou, Z.H.: A review on multi-label learning algorithms. TKDE **26**(8), 1819–1837 (2014)
32. Zhang, W., Wang, L., Yan, J., Wang, X., Zha, H.: Deep extreme multi-label learning. arXiv preprint arXiv:1704.03718 (2017)
33. Maaten, L.V.D., Hinton, G.: Viualizing data using T-SNE. J. Mach. Learn. Res. **9**, 2579–2605 (2008)