Check for
updates

OntoMedRec: Logically-pretrained model-agnostic ontology encoders for medication recommendation

Weicong Tan[1] · Weiqing Wang[1] · Xin Zhou[1] · Wray Buntine[2] · Gordon Bingham[3] ·
Hongzhi Yin[4]

Abstract

Recommending medications with electronic health records (EHRs) is a challenging task for
data-driven clinical decision support systems. Most existing models learnt representations
for medical concepts based on EHRs and make recommendations with the learnt represen-
tations. However, most medications appear in EHR datasets for limited times (the frequency
distribution of medications follows power law distribution), resulting in insufficient learning
of their representations of the medications. Medical ontologies are the hierarchical classifi-
cation systems for medical terms where similar terms will be in the same class on a certain
level. In this paper, we propose **OntoMedRec**, the *logically-pretrained* and *model-agnostic*
medical **Onto**logy Encoders for **Med**ication **Rec**ommendation that addresses data sparsity
problem with medical ontologies.
We conduct comprehensive experiments on real-world EHR datasets to evaluate the effec-
tiveness of OntoMedRec by integrating it into various existing downstream medication
recommendation models. The result shows the integration of OntoMedRec improves the
performance of various models in both the entire EHR datasets and the admissions with
few-shot medications. We provide the GitHub repository for the source code. (https://github.
com/WaicongTam/OntoMedRec)

Keywords Medication recommendation · Logic tensor networks · Medical ontology

1 Introduction

The mass application of electronic health records (EHRs) has made data-driven clinical
decision-support systems possible [1]. Deep learning models designed to assist clinical prac-

---

✉ Weiqing Wang
  teresa.wang@monash.edu

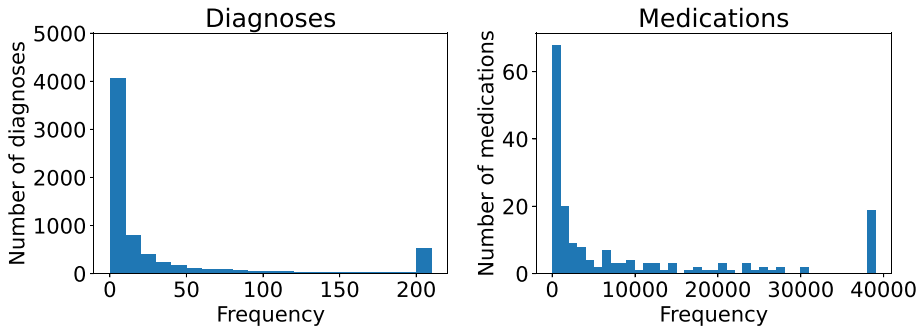Extended author information available on the last page of the article

Figure 1 Frequency distribution of diagnoses and medications in MIMIC-III dataset. The last bin is the cropped diagnoses/medications with a frequency higher than 200/40000

titioners in a range of tasks have emerged, with notable categories encompassing patient risk prediction, re-admission forecasting, the generation of EHR representations, and medication recommendations for prescribers. To assist medical practitioners in prescribing medications, recommending sets of medications for them accurately and efficiently has become a challenging yet crucial task. Therefore, numerous data-driven medication recommendation models have been developed, exemplified by notable solutions such as 4SDrug [2], EDGE [3], and SafeDrug [4]. These models aim to predict the most suitable medication regimen based on a patient's diagnoses, medical procedures, and/or prior prescription history, as demonstrated by systems like COGNet [5] and SARMR [6]. Existing medication recommendation models fall into two categories: instance-based models and longitudinal models. Instance-based models (e.g., LEAP [7] and 4SDrug [2]) recommend sets of drugs with patients' diagnoses in the current admission, whereas longitudinal models (e.g., MICRON [8], SafeDrug [4] and COGNet [5]) utilise patients' previous admissions.

For both instance-based models and longitudinal medication recommendation models, we identify one challenge that has not been sufficiently addressed: **data sparsity issue (challenge 1)**. Similar to the user-interaction sparsity challenge in other recommender system models [9, 10], medication recommendation models suffer from data sparsity issues deriving from the frequency distribution of medical concepts. As demonstrated in Figure 1, the majority of diagnoses and medications only appear at limited times in the entire MIMIC-III dataset and their occurrence follows the power law distribution. This inevitably leads to insufficient learning of the indication relationships between diagnoses and medications (i.e., for what medical conditions a medication was designed) in instance-based models and their respective embeddings in longitudinal models. As proven many other recommendation tasks (e.g., [11] and [12]), utilising external knowledge bases can alleviate the cold-start effect. One category of the notable knowledge base for medication recommendation models is medical ontologies. Therefore, to alleviate the data sparsity issue (**challenge 1**), similar to [13, 14], we leverage external structured knowledge (i.e., medical ontologies) [13, 14] as it provides prior knowledge for the medical terms in EHRs. In EHRs, diagnoses, procedures and medications are encoded in standardised hierarchical classification systems called as medical ontologies. Each medical term is a node of the ontology and the relation between them is "is-a" (e.g., benproperine is a cough suppressant).

Figure 2 shows part of ATC ontology which is an ontology of medications. In this ontology, similar medications fall into the same parent node, yet there are definitive differences that distinguish them (i.e., the difference between siblings). For example, as demonstrated in Fig-
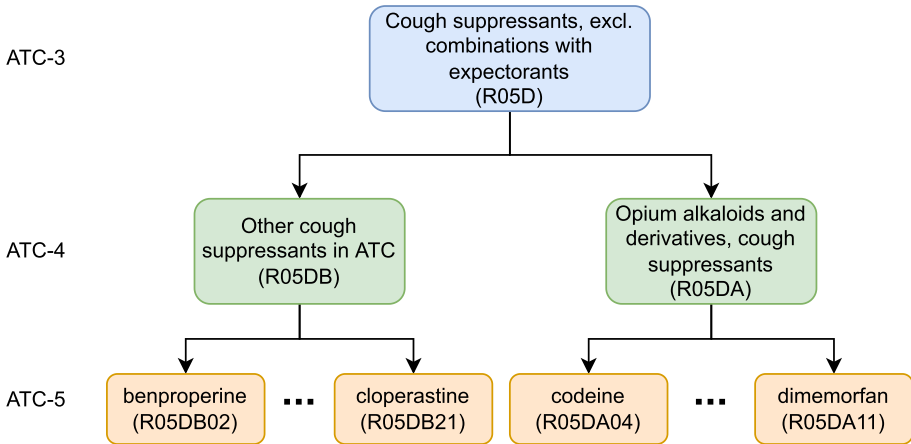
Figure 2 An excerpt of the ATC ontology. Some nodes are omitted

ure 2, medications in "Other cough suppressant in ATC" (R05DB) and "Opium alkaloids and derivatives, cough suppressants" (R05DA) fall into the same category "Cough suppressants, excl. combinations with expectorants"(R05). However, they are intrinsically different since codeine cough suppressants (i.e., R05DA) and non-codeine cough suppressants (i.e., R05DB) have different clinical characteristics (e.g., physical dependency and drug-drug interaction). Benproperine and cloperastine have the same therapeutical classification (i.e., they are both non-codeine cough suppressants), yet they are two different chemicals. As we can see from this example and some existing studies in recommender models (e.g., [15] and [16]), effectively modelling the parental, ancestral and sibling relationships (similarities and differences) is beneficial to the medication recommendation task.

Even though there are some works exploiting the modelling of medical ontologies in the medication recommendation task, **these existing works cannot effectively model ontology relationships to benefit the medication recommendation task (challenge 2)**. Notable models integrating ontology information in medication recommendation include G-BERT [13] and KnowAugNet [14]. G-BERT uses a Graph Attention Network (GAT) [17] encoder trained end-to-end along with the medication recommendation module. KnowAugNet pretrains ontology encoders with an unsupervised contrastive learning method. However, both models encode ontology with GAT and treat ontology as an undirected graph, whereas ontology is by definition a direct acyclic graph (DAG). Moreover, they cannot model some important relationships such as the difference between siblings as shown in Figure 2. There are also models designed for other tasks that utilised medical ontologies (e.g., GRAM [18] and KAME [19]). However, the modelling of the ontology in these methods is deeply coupled with their downstream tasks which are not medication recommendation.

To effectively model the ontology relationships to improve the medication recommendation task (**challenge 2**), we propose a model OntoMedRec based on logic tensor networks (LTN) [20] in this paper. As we know, LTN aims at combining symbolic rules and neural computation together. The advantage of using LTN in our task is that it allows us to easily integrate the modelling of various identified ontology relationships as symbolic rules (e.g., the parental and sibling relations in Figure 2) into the training process (i.e., neural computation). Recent advances in logic tensor networks (LTN) [20] have shown its effectiveness in graph learning tasks such as ontology deduction and reasoning [21]. However, **we find**

that directly applying existing LTN technique to our task is challenging for two reasons (challenge 3). The first reason is that the existing LTN works are designed for different task. To adapt to medication recommendation task, we need to design new sets of predicates, axioms, constants and variables. The second reason is that directly applying existing LTN methods is memory consuming. Existing LTN studies that designed for ontology data (e.g., [20] and [21]) are based on smaller ontologies (i.e., ≤ 100 nodes) with small representation dimensions. To model an ontology of $|\mathcal{N}|$ nodes with n variables and d as model dimension, the space complexity is $O(n|\mathcal{N}|d)$, which requires large amount of memory when the ontology is large (e.g., 17,737 nodes in our task). This affects the efficiency of the training process since the memories in GPUs are usually more scarce than RAMs. The high space complexity calls for an efficient sampling method for the effective training of larger node representations on larger ontologies. We devise a sampling method based on the structure of medical ontologies and our modelling method. It decreases the space complexity to $O(nbd)$ where $b << |\mathcal{N}|$ is the batch size. The contribution of this paper can be summarised as follows:

- **Logically-pretrained ontology encoder:** We carefully design an LTN-based encoder by devising novel predicates, axioms, constants, and variables for the self-supervised logical training on medical ontologies. *The design is based on the insights of what structural information is beneficial for the medication recommendation task.* The devised axioms are naturally interpretable for humans. Moreover, for the efficient training of the model, we also designed an *axiom-oriented sampling method* to enable the learning of larger node representations on large ontologies. Furthermore, to infuse the indication relationships between medications and medical diagnoses, we utilised the MEDI dataset [22] to *logically align the representation space* of diagnoses and medications.
- **Model-agnostic ontology representation learning model for medication recommendation:** Once the encoder is well trained, its output can be loaded into various existing medication recommendation models to improve their performance as the initialisation of the embeddings of medical codes (i.e., diagnoses, procedures and/or medication embeddings). Thus, similar to other pretrained models, our encoder is *"once trained and ready to use for any medication recommendation models"*.
- **Comprehensive experiments:** Comprehensive experiments have been done (with code published) to validate the effectiveness of our model in improving different existing medication recommendation methods including both *instance-based methods and longitudinal methods* for both *normal scenarios and few-shot scenarios*. The results show that: 1) our model is able to improve the performance of both instance-based and longitudinal downstream recommendation methods but the improvements on longitudinal methods are more obvious compared to instance-based methods; 2) our model is able to improve the performance of existing recommendation methods in both normal scenarios and few-shot scenarios but the improvement is more obvious for few-shot scenarios.

2 Related work

2.1 Instance-based medication recommendation

Instance-based models recommend a set of medications based on the current admission. LEAP [7] was an early model that predicted the prescribed medications as sequences, and it made inferences with beam search. SMR [23] recommends drugs based on knowledge graph

embeddings of diagnoses and medications. More recently, 4SDrug [2] was proposed. It is a set-based model trained by comparing the difference between medication sets with similar corresponding diagnoses sets.

2.2 Longitudinal medication recommendation

Longitudinal models make use of patients' previous diagnoses and procedures records. RETAIN [24] was a representation learning model that encodes a patient's EHR into a representation, and it can be used for the medication recommendation task with extra output layers. DCw-MANN [25] used all past medications to predict current medications using a LSTMs-based encoder-decoder model. GAMENet [26] used memory bank matrices to associate past diagnoses and procedures with medications. SafeDrug [4] uses a global molecule encoder and a local molecule substructure encoder to encode medications. COGNet [5] uses Transformer-based [27] to encode the patient's diagnoses, procedures and medication history. MICRON [8] is a model designed for predicting the change in prescriptions, it models the change of prescribed medications with residual vectors. In addition to a patient's EHR, MERITS [28] used the neural ordinary differential equation to model the irregular time series of the patient's vital signs. The model proposed by Yao et al. [29] used RNN to model the path from the root node to medical concepts on medical ontologies. Other than medication recommendation, some longitudinal models use longitudinal EHR data to perform other tasks such as diagnoses prediction (e.g., KAME [19]) and representation learning (e.g., GRAM [18]). These two models also had medical ontology encoding modules, but they were trained end-to-end with downstream tasks.

2.3 Existing solutions to data sparsity issue

Some existing models have attempted to address the data sparsity issue. G-BERT [13] used GAT encoders to encode diagnoses and medications. However, the pretraining data used in G-BERT is the patient records with one admission. These admission data still follow the distribution we described in Figure 1. kampnet [14] used unsupervised contrastive learning to pre-train encoders for medical ontologies and medication-diagnoses co-existence graph. EDGE [3] considers drugs that never appear in a certain time range in the EHR dataset as novel drugs, and uses meta-learning to alleviate the cold-start effect of those drugs. However, an interpretable and EHR-independent pretrained encoder for medical ontologies has not been proposed. Moreover, the data sparsity issue in medication recommendation has not been sufficiently addressed.

3 Preliminaries

3.1 Electronic medical record

An electronic health record (EHR) dataset can be considered as a collection of $|\mathcal{U}|$ patients' medical records $\mathcal{U} = \{\mathcal{U}^{(n)}\}_{n=1}^{|\mathcal{U}|}$ where a patient's medical record $\mathcal{U}^{(n)}$ is constituted by their admissions $[\mathcal{V}_t^{(n)}]_{t=1}^{T^{(n)}}$ to the hospital. For the sake of brevity, we will omit the (n) superscript in future formulae where there is no confusion. In each admission, a set of medical diagnosis codes $(\mathcal{D}_t)$, a set of medical procedure codes $(\mathcal{P}_t)$ and a set of prescribed medication codes

$(\mathcal{M}_t)$ will be recorded as $\mathcal{V}_t = \{\mathcal{D}_t, \mathcal{P}_t, \mathcal{M}_t\}$. Note that, in some medication recommendation models (e.g., COGNet [5] and LEAP [7]), the set of medical diagnosis codes $(\mathcal{D}_t)$ and the set of medical diagnosis codes $(\mathcal{D}_t)$ are considered sequences.

It is worth noticing that the diagnosis does not only record the chief complaints (i.e., the prominent symptoms that cause this specific admission to the hospital [30]) of the patient's admission. It also records other medical conditions of the patient. Assume that there is a diabetic patient with existing liver conditions who was admitted to the hospital due to a broken arm. Not only the bone fracture will be recorded, but the cause of the fracture (e.g., falling), the diabetes and liver conditions will be recorded as well. All the diagnosis information is codified as codes on a medical ontology that can be modelled by OntoMedRec.

3.2 Medical concept ontologies

A medical ontology $\mathcal{T}_* = \{\mathcal{N}_*, \mathcal{E}_*, \mathbf{E}_*\}$ is a hierarchical taxonomy of medical concepts in a certain domain. It is a directed acyclic graph (DAG) where $\mathcal{N}_*$ is the set of nodes, $\mathcal{E}_*$ is the set of edges and $\mathbf{E}_*$ is the matrix of node features. An edge $e_j = \langle n_a, n_b \rangle \in \mathcal{E}_*$ represents that n_b is a more specific concept deriving from n_a (i.e., n_a is a parent of n_b). Take the excerpt in Figure 2 again as an example. There is a directed edge from "Cough suppressants, excl. combinations with expectorants (R05D)" to "Other cough suppressant in ATC" (R05DB)" since R05DB is a more specific term to classify a medication.

For OntoMedRec, we will use three non-overlapping taxonomies respectively for diagnoses, medical procedures and medications, namely $\mathcal{T}_d$, $\mathcal{T}_p$ and $\mathcal{T}_m$. Note that, $\mathcal{T}_d$, $\mathcal{T}_p$ and $\mathcal{T}_m$ are publicly available and shared by all EHR datasets by their linkage to $\mathcal{D}_t$, $\mathcal{P}_t$ and $\mathcal{M}_t$ respectively. More specifically, each medical code is a node on the corresponding medical ontology. They can be either a leaf node or a parent node. Since the medical ontology is independent of EHR datasets and all medical concepts in EHR datasets belong to the medical ontology, the pretrained representations of OntoMedRec can be integrated into any downstream recommendation models trained and tested with EHR datasets.

3.3 Medication recommendation

Following the task definition in in Sec.1 and Sec.2, the medication recommendation task can be formulated as follows:

- Longitudinal models predict $\mathcal{M}_T$ given a patient's admission history $[\mathcal{D}_t, \mathcal{P}_t]_{t=1}^{T}$. Some of them add past medication records $[\mathcal{M}_t]_{t=1}^{T-1}$ (e.g., MICRON [8] and COGNet [5]).
- Instance-based models predict $\mathcal{M}_T$ given a patient's diagnosis information in the current visit $[\mathcal{D}_T, \mathcal{P}_T]$.

3.4 Logic tensor networks

Logic Tensor Networks (LTNs) are the neural networks for data modelling with quantifiable and human-interpretable rules. They are based on real logic [21] defined on a first-order language $\mathcal{L}$. $\mathcal{L}$ is composed of [20]:

- **A set of constants.** In our case, it is the node feature matrices $\mathbf{E}_* \in \mathbb{R}^{|\mathcal{N}_*| \times d}$ where $|\mathcal{N}_*|$ is the number of nodes and d is the dimension of the node representations.
- **A set of variables.** They are the symbols created over the subset of the constants to describe the logical relationships in the graph.

- **A set of predicates.** They are a set of functions $\{f_1(\cdot), f_2(\cdot), \cdots, f_n(\cdot)\}$ that take variables as inputs and calculate the satisfiability scores of a logical relationship.
- **A set of connectives.** They are logical operators and aggregation operators (e.g., "and $(\wedge)$" and "not $(\neg)$").

There, a knowledge base can be defined as a triple $< \mathcal{K}, \mathcal{G}(\cdot|\theta), \Theta >$, where

- $\mathcal{K}$ is a set of closed formulae (i.e., axioms) defined by the variables, predicates and connectives in $\mathcal{L}$ and the set of domain symbols. They are highly interpretable propositional logic expressions.
- $\mathcal{G}(\cdot|\theta)$ is the parameter groundings of the symbols and logical operators,
- Θ is the set of parameters in the groundings. This includes the trainable parameters of predicates and constants.

The training of an LTN model aims to find the set of optimal parameters Θ^* that maximise the aggregated satisfiability of $< \mathcal{K}, \mathcal{G}(\cdot|\theta) >$

$$\Theta^* = \underset{\theta \in \Theta}{\mathrm{argmax}} \ \underset{\phi \in \mathcal{K}}{\mathrm{SatAgg}}(\mathcal{G}_\theta(\phi)) \tag{1}$$

$$\underset{\phi \in \mathcal{K}}{\mathrm{SatAgg}}(\mathcal{G}_\theta(\phi)) = 1 - (\frac{1}{|\mathcal{K}|} \sum_{\phi \in \mathcal{K}} (\mathcal{G}_\theta(\phi)))^{\frac{1}{p}} \tag{2}$$

where $\mathrm{SatAgg}(\cdot)$ is the function that aggregates the satisfiabilities of each axiom and p is a hyperparameter.

Therefore, the training goal can be formulated as the minimisation of the loss $\mathcal{L}$:

$$\mathcal{L} = 1 - \underset{\phi \in \mathcal{K}}{\mathrm{SatAgg}}(\mathcal{G}_\theta(\phi)) \tag{3}$$

A more detailed and illustrative description of how these components are used to describe the logical relationship in medical ontologies is at Section 4.1

3.5 The indication relationship between medications and diagnoses

If a medication m was designed for treating a medical diagnosis d, an indication relationship $< m, d >$ can be defined. A medication can be designed to treat a set of medical conditions. If a medication is able to treat a parent node on the diagnosis ontology, it can be considered that it can cure all its children nodes. It is worth noticing that the indication relations graph does not enumerate all the existing indication relations.

4 The OntoMedRec model

4.1 Pre-training ontology encoders

By definition, the chosen medical ontologies have the following characteristics:

- **Explicit directed edges.** An edge in a medical ontology refers to a parent-child relationship between the two nodes. This relationship is not interchangeable or reflexive.
- **Implicit deductive relationships.** Besides explicit edges in $\mathcal{E}_*$, there are deductive relationships in medical ontologies.

- – Two nodes with the same parent node are sibling nodes. They have definitive differences on their level. ancestor node).
- – Ancestor nodes are multi-hop parent nodes. Ancestral relationships are not commutative or reflexive. We define one-hop ancestors as "parents" but not "ancestors".

- Each node (except the root node) has only one parent.

To accurately model the structural characteristics of medical ontologies, we pre-train three medical ontology node encoders, respectively for diagnoses, procedures and medications using logic tensor networks. Following the axioms used for the ontology deduction task in [21], we devise a set of additional axioms regarding the explicit and deductive relationships among nodes. Additionally, we devise axioms to define the sibling relationships in the ontology.

Since medical ontologies are much larger compared to the ontology in [21], it is impractical to define variables over all the nodes in these three ontologies. For instance, to describe the axiom "the parent node of the parent node of a node is an ancestor node", three variables are required ("$\forall x, y, z : P(x, y) \land P(y, z) \rightarrow A(x, z)$" where x, y, z are variables and $P(\cdot, \cdot)$ and $A(\cdot, \cdot)$ are the predicates that calculate the satisfiability of the parent and ancestor relation). Each time a new variable is created over the entire ontology, a new copy of the embedding matrix of all the nodes ($\mathbf{E}_* \in \mathbb{R}^{|\mathcal{N}| \times d}$) is required, which is a task of the space complexity of $O(n|\mathcal{N}|d)$, where n is the number of variables. To achieve efficient and effective training of the encoders, we design an axiom-oriented sampling method. We, firstly, randomly sample a batch of nodes from the ontology. Then, we sample all their respective ancestors, parents and siblings. This set of nodes constitutes a training node batch. All the directed edges between two nodes in the set constitute the positive edge samples, whereas all the node pairs without directive edges between them constitute the negative edge samples. With the adoption of the sampling method, the space complexity of the creation of variables are reduced to $O(nbd)$, where b is the batch size.

4.1.1 The knowledge formulation of ontology data

Therefore, the knowledge of an ontology can be formulated as follows, using the notations in [20].

- **Domain** Medical terms in the ontology
- **Variables** x, y and z, ranging over a batch of sampled nodes $N_b \subset \mathcal{T}_*$
- **Predicates** $P_*(x, y)$ as the parent scorer, $S_*(x, y)$ as the sibling scorer and $A_*(x, y)$ as the ancestor scorer
- **Axioms**
  - – Parental relationships are not reflexive and commutative: $\forall x \in N_b : \neg P_*(x, x)$, $\forall x, y \in N_b : P_*(x, y) \rightarrow \neg P_*(y, x)$
  - – Ancestral relationships are not reflexive and commutative: $\forall x \in N_b : \neg A_*(x, x)$, $\forall x, y \in N_b : A_*(x, y) \rightarrow \neg A_*(y, x)$
  - – The definition of sibling relationships (nodes with the same parent node): $\forall x, y, z \in N_b : P_*(x, y) \land P_*(x, z) \rightarrow S_*(y, z)$
  - – Sibling relationships are not reflexive but commutative: $\forall x \in N_b : \neg S_*(x, x)$, $\forall x, y \in N_b : S_*(x, y) \rightarrow S_*(y, x)$
  - – The parent node of a parent node is an ancestor node: $\forall x, y, z \in N_b : P_*(x, y) \land P_*(y, z) \rightarrow A_*(x, z)$

- The parent node of an ancestor node is an ancestor node: $\forall x, y, z \in N_b : P_*(x, y) \wedge A_*(y, z) \rightarrow A_*(x, z)$
- Positive and negative edges in the batch: $\forall(x, y) \in P_b : P_*(x, y), \forall(x, y) \notin P_b : \neg P_*(x, y)$

- **Grounding**

  - Let $\mathbf{v}_n$ be the representation of node n, $\mathcal{G}(\mathbf{v}_n) = \mathbb{R}^d$
  - $\mathcal{G}(x|\theta) = \mathcal{G}(y|\theta) = \mathcal{G}(z|\theta) = [\mathbf{v}_n | n \in N_n]$
  - $\mathcal{G}(P_*|\theta), \mathcal{G}(S_*|\theta)$ and $\mathcal{G}(A_*|\theta)$ are $\sigma(\text{MLP}(x, y))$ with one output neuron and sigmoid function ($\sigma(\cdot)$) as the activation of the final layer

Ontology encoders are trained to maximise the aggregated satisfiability of all these axioms describing the structural characteristics of the ontology. For each ontology, we use a different set of predicates with the same structure. The three sets of predicates are optimised separately.

4.1.2 The alignment of diagnosis and medication representations

Intuitively, aligning the representations of medications and diagnoses after they have been respectively trained shortens the distance of these representations. The representations of diagnoses and medications are infused with the indication relationship. Therefore, using the pretrained representations from OntoMedRec as a starting point can improve the performance of the model, particularly in admissions with few-shot medications. Similarly, the knowledge of the MEDI dataset can be formulated as follows:

- **Domains:** Medical terms in the medication and diagnoses ontology
- **Variables**

  - Medication m ranging over all the medications in the batch of sampled indication pairs
  - Diagnoses s_x and s_y ranging over all the medications in the batch of sampled indication pairs

- **Predicates** $I(m, d)$ for the indication relationship
- **Axioms:** Let $\mathcal{I}$ be all the indication pairs in a sampled batch in MEDI dataset: $\forall(m, s_x) \in \mathcal{I}_b : I(m, s_x)$
- **Grounding**

  - Let $\mathbf{m}$ and $\mathbf{d}$ and be the representation of medication m and diagnosis d, $\mathcal{G}(\mathbf{m}) = \mathcal{G}(\mathbf{d}) = \mathbb{R}^d$
  - $\mathcal{G}(m|\theta) = [\mathbf{m}_m | m \in \mathcal{I}]$, $\mathcal{G}(d|\theta) = [\mathbf{d}_d | d \in \mathcal{I}]$
  - $\mathcal{G}(I|\theta)$ is $\sigma(\text{MLP}(m, d))$ with one output neuron and sigmoid function ($\sigma(\cdot)$) as the activation of the final layer

In each pretraining epoch, we train the three encoders sequentially then align medication and diagnosis embeddings with the indication dataset. We save the procedure embeddings with the highest satisfiability on the procedure ontology, and the medication and diagnoses embeddings with the highest satisfiability on the indication dataset.

4.2 Fine-tuning with downstream models

Following the pre-training phase, the embeddings of medical terms are integrated with downstream medication recommendation models for further fine-tuning. We choose both

Table 1 The statistical characteristics of the medical ontologies

	Diagnosis	Procedure	Medication
# nodes	17737	4670	6441
# edges	17736	4669	6440
Max depth	7	4	5

instance-based models (Leap [7] and 4SDrug [2]) and longitudinal models (RETAIN [24], SafeDrug [4] and MICRON [8]) to fine-tune and evaluate OntoMedRec.

The representations of diagnoses and procedures (and medications, where possible) are loaded as a starting point for the respective embedding table and are further end-to-end fine-tuned with the medication recommendation task.

5 Experiments

5.1 Experimental setup

5.1.1 Dataset

We use the ATC ontology for medications and the ICD9-CM ontology for diagnoses and procedures from BioPortal [31] to pretrain OntoMedRec. ICD9-CM is split into two sub-ontologies, respectively for diagnoses and procedures. The characteristics of these ontologies are listed in Table 1.

We use the benchmark dataset MIMIC-III [32] to fine-tune and evaluate the performance of downstream models integrated with the representations of OntoMedRec and other baselines. The statistical characteristics of the datasets are described in Table 2. To explore the performance of downstream models with or without OntoMedRec in sparse cases, we reserve the patients with only one admission, low-frequency diagnoses and low-frequency medications that were discarded in previous studies (e.g., in [4]). The ratio of training, testing and validation set is 4 : 1 : 1. We split out a set of admissions with few-shot medications. We use TWOSIDES dataset [33] as the ground truth of drug-drug interactions (DDIs). In contrast to previous studies, we reserve the drug pairs with lower numbers of DDIs.

5.1.2 The generation of few-shot medications test cases

We sort all medications in the EHR dataset by their frequencies. The medications with the lowest 30% frequency (i.e., tail percentage) are designated as few-shot medications.

Table 2 The statistical characteristics of the MIMIC-III dataset

Item	MIMIC-III
# patients	35441
# admissions	44129
# medications	120
# procedures	1975
# diagnoses	6658

Prescriptions in the test set with more than 1 few-shot medication are added to the few-shot test set.

5.1.3 Baselines

There are two major categories of existing medical ontology modelling methods: EHR-independent models (KAMPNet [14]) and EHR-dependent models (G-BERT [13]). Both KAMPNet and G-BERT use GAT [17] to model medical ontology. Thus, we also choose GAT as one of the baselines to validate the effectiveness of our model. KAME [19] and GRAM [18] are not comparable to our models because their ontology training is deeply coupled with their downstream task which are not medication recommendation. GCN is commonly used for the modelling of EHR and DDI graphs [26]. Therefore, we choose randomly initiated naive embedding table, GAT [17] and GCN as baselines. GAT and GCN are fine-tuned along with the downstream models. We use link prediction as the pretraining task for these two models. The two baselines are trained 20 epochs, the best checkpoints with the lowest loss are selected.

5.1.4 Evaluation metrics

Following the evaluation protocol of many medication recommendation models, we use the following metrics:

- **Jaccard coefficient.** It is the most common benchmark score for medication recommendation models. The Jaccard coefficient of all the the n patient's T admissions is calculated as follows:

$$\text{Jaccard}_t^{(n)} = \frac{|\{i : \mathbf{m}_{t,i}^{(n)} = 1\} \cap \{i : \hat{\mathbf{m}}_{t,i}^{(n)} = 1\}|}{|\{i : \mathbf{m}_{t,i}^{(n)} = 1\} \cup \{i : \hat{\mathbf{m}}_{t,i}^{(n)} = 1\}|} \tag{4}$$

$$\text{Jaccard}^{(n)} = \frac{1}{T^{(n)}} \sum_{t=1}^{T^{(n)}} \text{Jaccard}_t^{(n)} \tag{5}$$

  where $\{i : \mathbf{m}_{t,i}^{(n)} = 1\}$ is the set of indices where the element at i on multi-hot encoding vector is 1. The higher the Jaccard coefficient is, the more accurate the recommendation is (i.e., the recommended set is more similar to the label).

- **Drug-Drug Interaction (DDI) score.** It is the percentage of medication pairs with known DDIs in the recommended set of medications. The lower it is, the fewer DDIs there are in the generated medication recommendation, and the safer the recommended medication combination can be considered. The DDI score of the n patient's admissions is calculated as follows:

$$\text{DDI}^{(n)} = \frac{\sum_t^{T^{(n)}} \sum_{j,k \in \hat{\mathbf{m}}_{t,i}^{(n)} = 1} \mathbf{1}\{\mathbf{D}_{j,k} = 1\}}{\sum_{j,k \in \hat{\mathbf{m}}_{t,i}^{(n)} = 1} 1} \tag{6}$$

  where $\mathbf{D} \in \mathbb{R}^{|\mathcal{M}| \times |\mathcal{M}|}$ is the DDI matrix retrieve from the TWOSIDES dataset [33] and $\mathbf{1}\{\cdot\}$ is an indicator function that returns 1 when the input is true and 0 otherwise. $\mathbf{D}_{j,k} = 1$ indicates that medication j and medication k have at least one adverse effect when prescribed together.

Table 3 Performance of selected models on the MIMIC-III dataset

		Jaccard	F1	DDI	No. of drugs
LEAP	Naive	0.4689 ± 0.0019	0.6287 ± 0.0019	0.0603 ± 0.0004	17.8810 ± 0.0405
	+GAT	0.4178 ± 0.0009	0.5815 ± 0.0009	0.0631 ± 0.0000	19.9971 ± 0.0014
	+GCN	0.3853 ± 0.0012	0.5500 ± 0.0013	0.0769 ± 0.0000	12.9998 ± 0.0002
	+OMR	$\mathbf{0.4732 \pm 0.0017}$	$\mathbf{0.6322 \pm 0.0016}$	$\mathbf{0.0596 \pm 0.0004}$	17.1709 ± 0.0465
SafeDrug	Naive	0.5431 ± 0.0016	0.6920 ± 0.0014	0.0600 ± 0.0002	21.9985 ± 0.0668
	+GAT	0.5232 ± 0.0022	0.6740 ± 0.0020	$\mathbf{0.0569 \pm 0.0002}$	21.6908 ± 0.0838
	+GCN	0.5202 ± 0.0013	0.6707 ± 0.0010	0.0580 ± 0.0002	22.1046 ± 0.1074
	+OMR	$\mathbf{0.5481 \pm 0.0024}$	$\mathbf{0.6965 \pm 0.0021}$	0.0589 ± 0.0002	22.1837 ± 0.0790
MICRON	Naive	0.5147 ± 0.0020	0.6645 ± 0.0020	$\mathbf{0.0504 \pm 0.0005}$	15.7760 ± 0.1218
	+GAT	0.4991 ± 0.0026	0.6508 ± 0.0023	0.0510 ± 0.0003	15.4560 ± 0.1239
	+GCN	0.5003 ± 0.0017	0.6524 ± 0.0016	0.0511 ± 0.0004	15.4916 ± 0.0648
	+OMR	$\mathbf{0.5203 \pm 0.0019}$	$\mathbf{0.6696 \pm 0.0019}$	0.0517 ± 0.0003	16.1107 ± 0.1218
4SDrug	Naive	0.4667 ± 0.0017	0.6261 ± 0.0016	0.0478 ± 0.0004	13.7054 ± 0.0463
	+GAT	0.4666 ± 0.0016	0.6260 ± 0.0017	0.0477 ± 0.0005	13.7442 ± 0.0624
	+GCN	$\mathbf{0.4670 \pm 0.0017}$	$\mathbf{0.6263 \pm 0.0016}$	0.0478 ± 0.0004	13.7361 ± 0.0559
	+OMR	0.4662 ± 0.0015	0.6257 ± 0.0014	$\mathbf{0.0474 \pm 0.0005}$	13.7321 ± 0.0724
RETAIN	Naive	0.5433 ± 0.0023	0.6913 ± 0.0019	0.0646 ± 0.0006	17.0477 ± 0.1014
	+GAT	0.4264 ± 0.0023	0.5871 ± 0.0023	$\mathbf{0.0554 \pm 0.0006}$	14.9248 ± 0.0877
	+GCN	0.4335 ± 0.0020	0.5909 ± 0.0018	0.0574 ± 0.0006	16.4566 ± 0.0917
	+OMR	$\mathbf{0.5536 \pm 0.0019}$	$\mathbf{0.7001 \pm 0.0015}$	0.0642 ± 0.0005	17.7567 ± 0.0836

- **F1 score.** It is commonly used as a metric for classification tasks. It is the harmonic mean of the precision and recall score and is calculated as follows:

$$\text{Precision}_t^{(n)} = \frac{|\{i : \mathbf{m}_{t,i}^{(n)} = 1\} \cap \{i : \hat{\mathbf{m}}_{t,i}^{(n)} = 1\}|}{|\{i : \hat{\mathbf{m}}_{t,i}^{(n)} = 1\}|} \tag{7}$$

$$\text{Recall}_t^{(n)} = \frac{|\{i : \mathbf{m}_{t,i}^{(n)} = 1\} \cap \{i : \hat{\mathbf{m}}_{t,i}^{(n)} = 1\}|}{|\{i : \mathbf{m}_{t,i}^{(n)} = 1\}|} \tag{8}$$

$$\text{F1}_t^{(n)} = 2 \cdot \frac{\text{Precision}_t^{(n)} \cdot \text{Recall}_t^{(n)}}{\text{Precision}_t^{(n)} + \text{Recall}_t^{(n)}} \tag{9}$$

5.2 Results discussion

Table 3 lists the results of the performance of downstream models in the entire testing set, and Table 4 lists the results with the few-shot medications testing set. The drop in performance between Tables 3 and 4 in all downstream models proves our assumption that the data sparsity issue harms the performance of downstream models. Although OntoMedRec cannot achieve the lowest DDI in some models in both test settings, they are lower than the ground truth DDI score (0.078 among the entire test set and 0.069 in the few-shot test set).

Table 4 Performance of selected models on the test set with few-shot medications

		Jaccard	F1	DDI	No. of drugs
Leap	Naive	0.4328 ± 0.0060	0.5981 ± 0.0060	0.0522 ± 0.0018	18.3266 ± 0.1629
	+GAT	0.3978 ± 0.0064	0.5636 ± 0.0067	0.0632 ± 0.0000	20.0000 ± 0.0000
	+GCN	0.3451 ± 0.0046	0.5083 ± 0.0051	0.0769 ± 0.0000	13.0000 ± 0.0000
	+OMR	$\mathbf{0.4341 \pm 0.0071}$	$\mathbf{0.5986 \pm 0.0071}$	$\mathbf{0.0571 \pm 0.0023}$	17.4800 ± 0.2528
SafeDrug	Naive	0.5141 ± 0.0083	0.6705 ± 0.0074	0.0577 ± 0.0007	23.8330 ± 0.5112
	+GAT	0.5044 ± 0.0054	0.6618 ± 0.0049	0.0564 ± 0.0010	23.8990 ± 0.2941
	+GCN	0.4940 ± 0.0082	0.6517 ± 0.0079	$\mathbf{0.0562 \pm 0.0009}$	24.6039 ± 0.4459
	+OMR	$\mathbf{0.5206 \pm 0.0071}$	$\mathbf{0.6769 \pm 0.0062}$	0.0587 ± 0.0011	24.7879 ± 0.4228
MICRON	Naive	0.4849 ± 0.0099	0.6411 ± 0.0101	0.0615 ± 0.0018	20.5828 ± 0.6436
	+GAT	0.4626 ± 0.0061	0.6215 ± 0.0067	$\mathbf{0.0583 \pm 0.0023}$	19.9857 ± 0.9167
	+GCN	0.4660 ± 0.0062	0.6240 ± 0.0052	0.0628 ± 0.0010	20.0187 ± 0.6505
	+OMR	$\mathbf{0.4876 \pm 0.0094}$	$\mathbf{0.6428 \pm 0.0087}$	0.0622 ± 0.0012	20.6576 ± 0.5616
4SDrug	Naive	0.4310 ± 0.0071	0.5953 ± 0.0064	0.0385 ± 0.0020	14.4564 ± 0.2999
	+GAT	0.4298 ± 0.0064	0.5947 ± 0.0062	0.0385 ± 0.0016	14.3754 ± 0.3247
	+GCN	0.4304 ± 0.0068	0.5949 ± 0.0065	0.0380 ± 0.0023	14.2446 ± 0.3319
	+OMR	$\mathbf{0.4316 \pm 0.0079}$	$\mathbf{0.5961 \pm 0.0075}$	$\mathbf{0.0379 \pm 0.0018}$	14.4943 ± 0.3259
Retain	Naive	0.5057 ± 0.0077	0.6615 ± 0.0069	0.0608 ± 0.0018	19.7118 ± 0.6843
	+GAT	0.3650 ± 0.0076	0.5283 ± 0.0081	$\mathbf{0.0585 \pm 0.0014}$	16.1552 ± 0.2670
	+GCN	0.3637 ± 0.0069	0.5263 ± 0.0076	0.0700 ± 0.0011	18.2791 ± 0.3432
	+OMR	$\mathbf{0.5229 \pm 0.0089}$	$\mathbf{0.6780 \pm 0.0078}$	0.0609 ± 0.0015	20.6692 ± 0.6741

5.2.1 Results on the entire MIMIC-III dataset

As we can observe from Table 3, integrating the representations of diagnoses, procedures and medications (where possible) from OntoMedRec can improve the performance of most selected medication recommendation models in the entire dataset compared to all baselines. This demonstrates the representation of OntoMedRec is model-agnostic for downstream medication recommendation models.

5.2.2 Results on the few-shot cases

The representations of OntoMedRec improve all compared downstream models in the few-shot medication test set. We can observe from the result that the representations improve the performance for all compared longitudinal models. For compared instance-based models, although the representations of OntoMedRec do not improve their performance by a large margin, we notice that 1) they have lower performance scores compared to selected longitudinal models, and 2) the performance after the integration of OntoMedRec is not lower than the best performance by a large margin. We speculate the reason is that instance-based models adopted fewer pretrained embeddings comparing to longitudinal models.
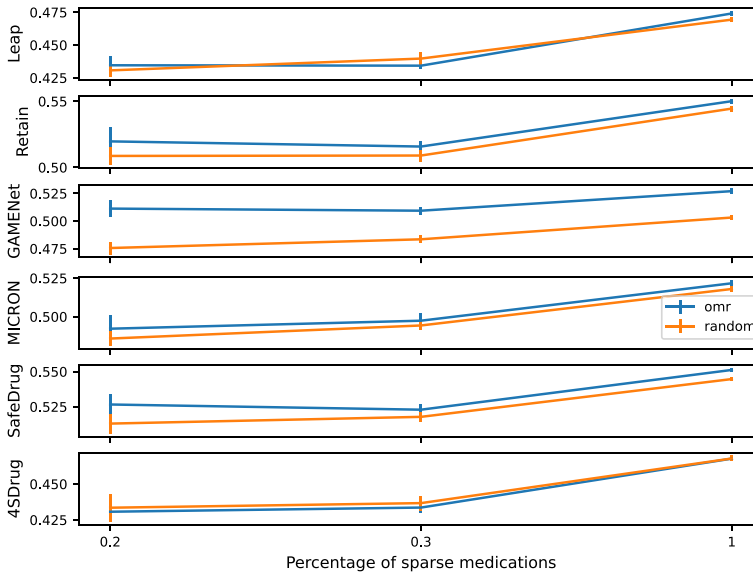
Figure 3 Performance of randomly initialised embedding table and OntoMedRec embedding table in different downstream models and tail percentages

5.2.3 Further investigation of sparse scenarios

To further investigate how few-shot medications affect the performance of medication recommendation models with OntoMedRec, we further compare OntoMedRec representations and randomly initialised embedding table with different sparse settings starting with the lowest frequency being 20% (as visualised in Figure 3). The representation of OntoMedRec can improve the performance of longitudinal downstream models in all three test sets. Overall, the performance gap between models with OntoMedRec pretraining and models without pretraining is larger when the data is sparser (20% is the sparsest scenario) which shows that medication recommendation models can benefit more from OntoMedRec in sparser scenarios. For LEAP, OntoMedRec can improve its performance for the entire testing set and the test set with the 20%-least-frequent medications. For 4SDrug, the performance margin is small.

6 Conclusion

In this paper, we proposed OntoMedRec, the self-supervised, logically-pretrained model-agnostic ontology encoders for medication recommendation. We devise axioms that collectively define the structure of medical ontologies, and use logical tensor networks (LTNs) to maximise the satisfiability of the representations. Furthermore, we align the representations of diagnoses and medications with medication indication information. The ontology-enhanced representation can be integrated into various downstream medication recommendation models to alleviate the negative effect brought by the data sparsity issue. We conducted experiments to evaluate the efficacy of OntoMedRec. Results show that the representation of OntoMedRec can improve the performance of most selected models in the

entire testing dataset, and that it can improve the performance of all longitudinal models in the few-shot medications test set.

Data Availability No datasets were generated or analysed during the current study.

Declarations

References

1. Yu, K.-H., Beam, A.L., Kohane, I.S.: Artificial intelligence in healthcare. Nat. Biomed. Eng. **2**(10), 719–731 (2018)
2. Tan, Y., Kong, C., Yu, L., Li, P., Chen, C., Zheng, X., Hertzberg, V.S., Yang, C.: 4sdrug: Symptom-based set-to-set small and safe drug recommendation. KDD '22, pp. 3970–3980. Association for Computing Machinery, New York, NY, USA (2022). https://doi.org/10.1145/3534678.3539089
3. Wu, Z., Yao, H., Su, Z., Liebovitz, D.M., Glass, L.M., Zou, J., Finn, C., Sun, J.: Knowledge-driven new drug recommendation. Preprint arXiv:2210.05572 (2022)
4. Yang, C., Xiao, C., Ma, F., Glass, L., Sun, J.: Safedrug: Dual molecular graph encoders for recommending effective and safe drug combinations. In: Zhou, Z.-H. (ed.) Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21, pp. 3735–3741 (2021). https://doi.org/10.24963/ijcai.2021/514 . Main Track. https://doi.org/10.24963/ijcai.2021/514. Main Track
5. Wu, R., Qiu, Z., Jiang, J., Qi, G., Wu, X.: Conditional generation net for medication recommendation. In: Proceedings of the ACM Web Conference 2022. WWW '22, pp. 935–945. Association for Computing Machinery, New York, NY, USA (2022). https://doi.org/10.1145/3485447.3511936
6. Wang, Y., Chen, W., PI, D., Yue, L., Wang, S., Xu, M.: Self-supervised adversarial distribution regularization for medication recommendation. In: Zhou, Z.-H. (ed.) Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21, pp. 3134–3140 (2021). https://doi.org/10.24963/ijcai.2021/431. Main Track
7. Zhang, Y., Chen, R., Tang, J., Stewart, W.F., Sun, J.: Leap: Learning to prescribe effective and safe treatment combinations for multimorbidity. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '17, pp. 1315–1324. Association for Computing Machinery, New York, NY, USA (2017). https://doi.org/10.1145/3097983.3098109

8. Yang, C., Xiao, C., Glass, L., Sun, J.: Change matters: Medication change prediction with recurrent residual networks. In: Zhou, Z.-H. (ed.) Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21, pp. 3728–3734 (2021). Main Track

9. Yin, H., Wang, Q., Zheng, K., Li, Z., Zhou, X.: Overcoming data sparsity in group recommendation. IEEE Trans. Knowl. Data Eng. **34**(7), 3447–3460 (2022). https://doi.org/10.1109/TKDE.2020.3023787

10. Lin, R., Tang, F., He, C., Wu, Z., Yuan, C., Tang, Y.: Dirs-kg: a kg-enhanced interactive recommender system based on deep reinforcement learning. World Wide Web **26**(5), 2471–2493 (2023). https://doi.org/10.1007/s11280-022-01135-x

11. Yang, N., Ma, Y., Chen, L., Yu, P.S.: A meta-feature based unified framework for both cold-start and warm-start explainable recommendations. World Wide Web **23**(1), 241–265 (2020). https://doi.org/10.1007/s11280-019-00683-z

12. Zhong, T., Zhang, S., Zhou, F., Zhang, K., Trajcevski, G., Wu, J.: Hybrid graph convolutional networks with multi-head attention for location recommendation. World Wide Web **23**(6), 3125–3151 (2020). https://doi.org/10.1007/s11280-020-00824-9

13. Shang, J., Ma, T., Xiao, C., Sun, J.: Pre-training of graph augmented transformers for medication recommendation. In: Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19, pp. 5953–5959 (2019)

14. An, Y., Tang, H., Jin, B., Xu, Y., Wei, X.: Kampnet: multi-source medical knowledge augmented medication prediction network with multi-level graph contrastive learning. BMC Med. Inform. Decis. Mak. **23**(1), 1–19 (2023)

15. Zheng, Q., Liu, G., Liu, A., Li, Z., Zheng, K., Zhao, L., Zhou, X.: Implicit relation-aware social recommendation with variational auto-encoder. World Wide Web **24**(5), 1395–1410 (2021)

16. Wang, H., Liu, G., Liu, A., Li, Z., Zheng, K.: Dmran: A hierarchical fine-grained attention-based network for recommendation. In: Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19, pp. 3698–3704 (2019). https://doi.org/10.24963/ijcai.2019/513

17. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., Bengio, Y.: Graph attention networks. In: International Conference on Learning Representations (2018). https://openreview.net/forum?id=rJXMpikCZ

18. Choi, E., Bahadori, M.T., Song, L., Stewart, W.F., Sun, J.: Gram: Graph-based attention model for healthcare representation learning. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '17, pp. 787–795. Association for Computing Machinery, New York, NY, USA (2017)

19. Ma, F., You, Q., Xiao, H., Chitta, R., Zhou, J., Gao, J.: Kame: Knowledge-based attention model for diagnosis prediction in healthcare. CIKM '18, pp. 743–752. Association for Computing Machinery, New York, NY, USA (2018)

20. Badreddine, S., Garcez, A.D., Serafini, L., Spranger, M.: Logic tensor networks. Artif. Intell. **303**, 103649 (2022)

21. Bianchi, F., Hitzler, P.: On the capabilities of logic tensor networks for deductive reasoning. In: AAAI Spring Symposium: Combining Machine Learning with Knowledge Engineering (2019)

22. Wei, W.-Q., Mosley, J.D., Bastarache, L., Denny, J.C.: Validation and enhancement of a computable medication indication resource (medi) using a large practice-based dataset. In: AMIA Annual Symposium Proceedings, vol. 2013, p. 1448 (2013). American Medical Informatics Association

23. Gong, F., Wang, M., Wang, H., Wang, S., Liu, M.: Smr: medical knowledge graph embedding for safe medicine recommendation. Big Data Res. **23**, 100174 (2021)

24. 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. Adv. Neural Inf. Process. Syst. **29** (2016)

25. Le, H., Tran, T., Venkatesh, S.: Dual control memory augmented neural networks for treatment recommendations. In: Advances in Knowledge Discovery and Data Mining, pp. 273–284. Springer, Cham (2018)

26. Shang, J., Xiao, C., Ma, T., Li, H., Sun, J.: Gamenet: Graph augmented memory networks for recommending medication combination. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 1126–1133 (2019)

27. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Adv. Neural Inf. Process. Syst. **30** (2017)

28. Zhang, S., Li, J., Zhou, H., Zhu, Q., Zhang, S., Wang, D.: Merits: Medication recommendation for chronic disease with irregular time-series. In: 2021 IEEE International Conference on Data Mining (ICDM), pp. 1481–1486 (2021). https://doi.org/10.1109/ICDM51629.2021.00192

29. Yao, Z., Liu, B., Wang, F., Sow, D., Li, Y.: Ontology-aware prescription recommendation in treatment pathways using multi-evidence healthcare data. ACM Trans. Inf. Syst. (2023)

30. Wagner, M.M., Hogan, W.R., Chapman, W.W., Gesteland, P.H.: Chief complaints and icd codes. Handbook of biosurveillance, **333** (2006)
31. Whetzel, P.L., Noy, N.F., Shah, N.H., Alexander, P.R., Nyulas, C., Tudorache, T., Musen, M.A.: Bioportal: enhanced functionality via new web services from the national center for biomedical ontology to access and use ontologies in software applications. Nucleic Acids Res.**39**(suppl_2), 541–545 (2011)
32. Johnson, A.E., Pollard, T.J., Shen, L., Lehman, L.-W.H., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Anthony Celi, L., Mark, R.G.: Mimic-iii, a freely accessible critical care database. Sci. Data **3**(1), 1–9 (2016)
33. Tatonetti, N.P., Ye, P.P., Daneshjou, R., Altman, R.B.: Data-driven prediction of drug effects and interactions. Sci. Transl. Med. **4**(125), 125–3112531 (2012)

Authors and Affiliations

Weicong Tan[1] · Weiqing Wang[1] · Xin Zhou[1] · Wray Buntine[2] · Gordon Bingham[3] · Hongzhi Yin[4]

Weicong Tan
weicong.tan@monash.edu

Xin Zhou
xin.zhou@monash.edu

Wray Buntine
wray.b@vinuni.edu.vn

Gordon Bingham
g.bingham@alfred.org.au

Hongzhi Yin
db.hongzhi@gmail.com

[1]   Faculty of Information Technology, Monash University, Victoria 3800, Australia

[2]   College of Engineering and Computer Science, VinUniversity, Hanoi, Vietnam

[3]   Nursing Services, Alfred Health, Melbourne, Victoria 3004, Australia

[4]   Faculty of Engineering, The University of Queensland, Brisbane, Queensland 4072, Australia