Learning Electronic Health Records through Hyperbolic Embedding of Medical Ontologies

Qiuhao Lu
University of Oregon
Eugene, OR, USA
luqh@cs.uoregon.edu

Nisansa de Silva
University of Oregon
Eugene, OR, USA
nisansa@cs.uoregon.edu

Sabin Kafle
University of Oregon
Eugene, OR, USA
skafle@cs.uoregon.edu

Jiazhen Cao
University of Oregon
Eugene, OR, USA
jcao@uoregon.edu

Dejing Dou
University of Oregon
Eugene, OR, USA
dou@cs.uoregon.edu

Thien Huu Nguyen
University of Oregon
Eugene, OR, USA
thien@cs.uoregon.edu

Prithviraj Sen
IBM Research
San Jose, CA, USA
senp@us.ibm.com

Brent Hailpern
IBM Research
San Jose, CA, USA
bth@us.ibm.com

Berthold Reinwald
IBM Research
San Jose, CA, USA
reinwald@us.ibm.com

Yunyao Li
IBM Research
San Jose, CA, USA
yunyaoli@us.ibm.com

ABSTRACT

Unplanned intensive care units (ICU) readmissions and in-hospital mortality of patients are two important metrics for evaluating the quality of hospital care. Identifying patients with higher risk of readmission to ICU or of mortality can not only protect those patients from potential dangers, but also reduce the high costs of healthcare. In this work, we propose a new method to incorporate information from the Electronic Health Records (EHRs) of patients and utilize hyperbolic embeddings of a medical ontology (i.e., ICD-9) in the prediction model. The results prove the effectiveness of our method and show that hyperbolic embeddings of ontological concepts give promising performance.

CCS CONCEPTS

• **Computing methodologies** → **Knowledge representation and reasoning**; Ontology engineering; • **Applied computing** → **Health informatics**.

KEYWORDS

medical ontology, graph embedding, readmission prediction, mortality prediction

ACM-BCB '19, September 7–10, 2019, Niagara Falls, NY, USA

ACM ISBN 978-1-4503-6666-3/19/09…$15.00
https://doi.org/10.1145/3307339.3342148

ACM Reference Format:
Qiuhao Lu, Nisansa de Silva, Sabin Kafle, Jiazhen Cao, Dejing Dou, Thien Huu Nguyen, Prithviraj Sen, Brent Hailpern, Berthold Reinwald, and Yunyao Li. 2019. Learning Electronic Health Records through Hyperbolic Embedding of Medical Ontologies. In *10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics (ACM-BCB '19), September 7–10, 2019, Niagara Falls, NY, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3307339.3342148

1 INTRODUCTION

Patients who are readmitted to intensive care units (ICU) after transfer or discharge are at high risk of mortality, and the readmissions are usually costly for both the patients and hospitals. Therefore, efficiently and accurately identifying patients who are prematurely discharged or transferred from ICU can not only reduce the risk of mortality, but also help decrease the high but avoidable costs of healthcare. According to a recent study [1], unplanned hospital readmission was estimated to have cost nearly $26 billion annually in the U.S. In addition to hospital readmission, ICU readmission is also a major problem. Around 10% of ICU patients are readmitted during the same hospitalization [20], due to premature discharge or premature transfer from ICU. This highlights the importance of predicting the ICU readmission risk for healthcare systems.

In the past few years, there have been several published studies [11, 14, 23, 33] on this unplanned readmission prediction task. Most of the studies are conducted by physicians and medical researchers, and they generally focus on selecting statistically significant features from ICU patients' Electronic Health Records (EHRs) and combining them with traditional machine learning methods, such as logistic regression [33]. These studies prove effectiveness by achieving good prediction accuracy, but they still can be improved

by incorporating more sophisticated features, such as the latent embeddings of ontological medical concepts in the patients' EHRs.

Similar to the unplanned readmission prediction task, for in-hospital mortality prediction, there are studies [6, 7] that outperform the traditional scoring systems [10, 31] with machine learning methods. However, they have common limitations: They are not using any external knowledge to improve their models. Therefore, their approaches can be improved by incorporating external knowledge such as medical ontologies.

Medical ontologies are primarily characterized by hierarchical relationships and textual descriptions, along with non-hierarchical features. While Euclidean space is the default geometry for word-based embedding methods [16], embeddings learned in hyperbolic spaces [17] are capable of representing the hierarchies more efficiently. Although there has been some progress in learning word embeddings in hyperbolic spaces [5], effective leveraging of hierarchies with other data sources remains an open problem.

Hyperbolic space-based representation learning provides an effective way to learn latent embeddings for medical ontologies, which are inherently hierarchical in nature. This significantly helps solve the problem for learning medical ontology embeddings by providing more efficiency and lower dimensions. However, this also poses a significant challenge to medical applications. It is a well-discovered fact that in Euclidean spaces, the learned embeddings are the function of context, which is defined during training. When syntactic structures are taken as the context, words are considered semantically similar when they are surrounded by that same context. Comparable analogies can also be drawn for medical concepts. For example, medical concepts are used by medical providers for billing purposes; and thus, similar concepts for such tasks are concepts which co-occur in a diagnosis as well as in a similar hierarchical structure.

In this study, we propose a new method to leverage latent information in the textual data from ICU patients' EHRs, by training and combining the hyperbolic embeddings of the medical concepts in them. We implement our method based on the state-of-the-art method [14] on ICU readmission prediction and the widely accepted benchmark [6] on in-hospital mortality prediction, and we show improvement in both tasks. We also evaluate the hyperbolic embeddings of medical concepts by comparing them with other popular graph embedding methods, both intrinsically and extrinsically. All the experiments are conducted on the MIMIC-III dataset [8].

Our contributions are summarized as follows:

(1) Our method of adding embeddings of ICD-9 codes from discharge summaries proves effective and it helps improve the performance of the state-of-the-art method on ICU readmission prediction with different graph embeddings. It also outperforms the benchmark results in the task of mortality prediction.
(2) We prove that the hyperbolic embeddings of medical concepts give promising performance in different evaluations, outperforming Euclidean-based graph embeddings in intrinsic evaluation and give comparable performance in extrinsic evaluation.

The rest of this paper is organized as follows. In Section 2, we review previous studies related to our work. We describe our method in Section 3. In Section 4, we demonstrate the experimental results with analysis. The conclusion and future work are discussed in Section 5.

2 RELATED WORK

2.1 Hyperbolic Representation Learning

Representation learning is one of the fundamental characteristics of deep learning advances, with representation learning of words as vectors enabling significant advantages over traditional feature engineering methods. While it is common to use the output of the layer before the last layer of a Convolutional Neural Network (CNN) as an image representation for downstream tasks, it is no surprise to see similar approaches for representation learning being applied to other data sources, such as knowledge bases (KBs) [3]. Recently, a similar idea, in which the process includes representation learning of the medical concepts represented in a large scale KB (e.g., UMLS, SNOMED) [4], has been applied to the medical domain. This has enabled application of deep learning advances into the medical domain, and at the same time, it has increased the range of applications of medical KBs.

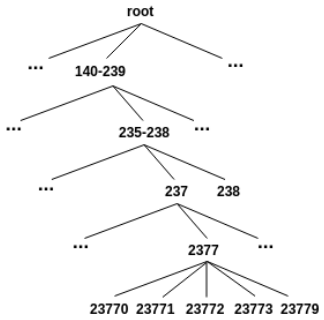
It has been shown that linear representations need a much higher number of dimensions in order to represent the hierarchies, which are the most common aspect in the KB [17]. Consequently, representation learning in hyperbolic spaces, as opposed to in Euclidean spaces, has been proposed, with hyperbolic space embedding found to perform better in representation of hierarchies, especially with lower dimension of features [24]. There have been several methods proposed for learning representations in hyperbolic spaces in which it has also been found that, for data which is not inherently hierarchical (e.g., words), such methods do not yield significant performance gain [5, 13].

Medical ontologies are usually hierarchically organized. This kind of tree-like structure can be well-represented in hyperbolic space. To better illustrate, we visualize the Poincaré embedding by training a 2-D embedding of a subtree of the ICD-9 ontology [26]. As shown in Figure 1, the embedding looks like a continuous version of "tree," in which the low-level nodes (the leaves) are on the edge, and the high-level nodes (root) are on the center area. This is consistent with the features of hyperbolic space.

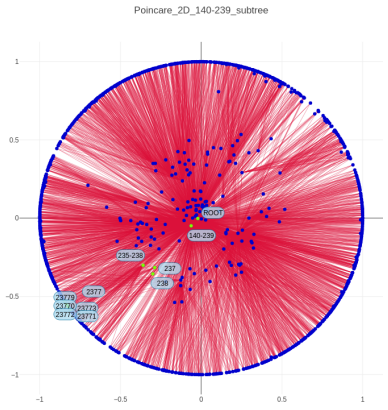
2.2 Unplanned ICU Readmission Prediction

Unplanned ICU readmission prediction, along with unplanned hospital readmission prediction, is an important task in the healthcare field. Apart from research conducted by physicians, which is usually based on specific feature engineering and traditional machine learning methods [33], there is also some solid work that is from the angle of natural language processing, which focuses on predicting readmissions based on medical textual notes from EHRs [23]. Some researchers exploit representation learning techniques to solve this problem, where they learn either embeddings of patients or embeddings of medical concepts from patients' data [11, 14].

To the best of our knowledge, [14] proposes the best Area Under the Receiver Operating Characteristics curve (AUROC) score of 0.791 for the ICU readmission prediction task on the MIMIC-III dataset [8]. They take three types of information as input, i.e., chart events information, basic demographic information, and diagnosis



(a) Hierarchical Structure of the Subtree



(b) Visualization of the 2-D Hyperbolic Embeddings of the Subtree

Figure 1: Hierarchy and Corresponding 2-D Hyperbolic Embeddings of "140-239" Subtree of ICD-9

information (in the form of ICD-9 codes). All 3 of the types of features are concatenated and put into the prediction model, which is a sequential combination of two LSTM layers and one multi-filter CNN layer. They also utilize the embeddings of diagnosis (in the form of ICD-9 codes) to improve the prediction, which are learned from a private medical claims dataset which consists of the health claims data of roughly four million people from 2005 to 2013 [4].

Though their work proves effective, they completely overlook the important information that is encoded in the medical text notes in patients' EHRs. Researchers have proved that the medical text notes in patients' EHRs contain enough information that can be used to support the readmission prediction task [23]. In this study, we incorporate the important textual information by extracting ICD-9 codes from the medical notes and encode them with different graph embedding methods.

2.3 In-Hospital Mortality Prediction

In-hospital mortality prediction is another important task in the medical domain which aims at predicting the mortality of patients when they are in hospital. Early studies develop systems that calculate the predictions based on expert knowledge or data driven approaches, such as the APACHE [31], the APACHE II [30], the SAPS [9], and the SAPS II [10].

Recently, researchers have used machine learning techniques to deal with this problem. Johnson et al. [7] use three traditional methods to solve this task, i.e., Logistic Regression, SVM, and Random Forest. The benchmark [6] we use also compares different approaches with traditional scoring systems, which include Logistic Regression and LSTM-based models. Although these works advance the current state-of-the-art performance in mortality prediction, few of them utilize external knowledge, such as medical ontologies. Hence, in this study, we make use of the ICD-9 codes, and we represent them with graph embeddings, to see whether it can improve the performance of in-hospital mortality prediction.

2.4 ICD-9 and other Medical Ontologies

There are multiple medical ontologies:

- **UMLS**: The Unified Medical Language System is a compendium of many controlled vocabularies in the biomedical sciences [2].
- **SNOMED CT**: SNOMED Clinical Terms is a systematically organized computer processable collection of medical terms providing codes, terms, synonyms and definitions used in clinical documentation and reporting [27].
- **ICD-9**: ICD-9 is the 9-th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD), a medical classification list by the World Health Organization (WHO) [26]. Besides ICD-9, more recent versions (i.e., ICD-10 and ICD-11) are widely used as well.
- **MeSH**: Medical Subject Headings (MeSH) is a comprehensive controlled vocabulary for the purpose of indexing journal articles and books in the life sciences; it serves as a thesaurus that facilitates searching [15].

In this study, we conduct experiments on the MIMIC-III dataset, which takes ICD-9 as their coding ontology. ICD-9 Clinical Modification (ICD-9-CM) is a modification of ICD-9. This national variant of ICD-9 is provided by the Centers for Medicare and Medicaid Services (CMS) and the National Center for Health Statistics (NCHS), and the use of ICD codes are now mandated for all inpatient medical reporting requirements.

3 METHOD

3.1 Hyperbolic Medical Concept Embeddings

We refer the readers to [5, 13, 17] for more detailed treatment of hyperbolic spaces, and their characterization and differences with respect to the Euclidean space geometry. Any metric space is characterized by the distance between two points, with the distance being defined in hyperbolic space, specifically for Poincaré ball model for two points $u, v \in \mathbb{B}^d$ is

$$d_H(u,v) = \operatorname{arcosh}\left(1 + 2\frac{||u-v||^2}{(1-||u||^2)(1-||v||)^2)}\right) \quad (1)$$

For a unit Poincaré ball space, $||u|| < 1$. As is evident from Equation 1, the distances between two points near the boundary of Poincaré ball tend to ∞. Also, for a hierarchical structure (e.g., a tree) that is embedded into the space, the root node will be placed in the

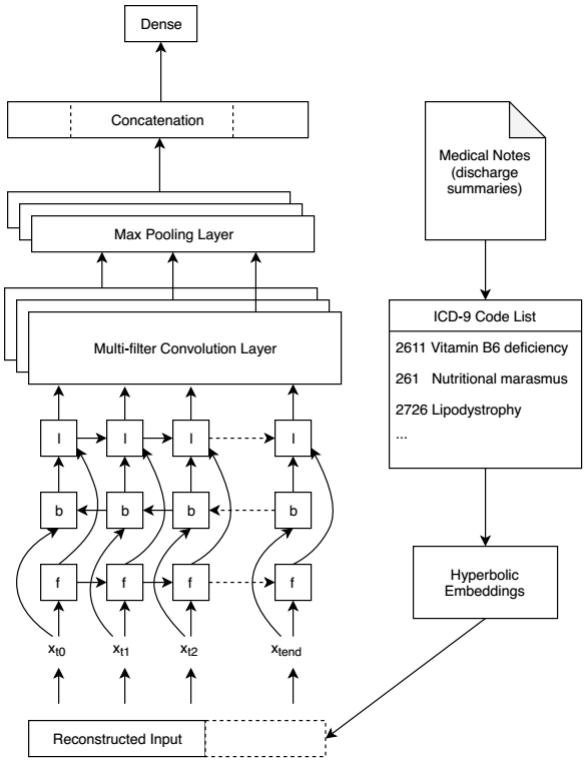


Figure 2: Framework of Readmission Prediction

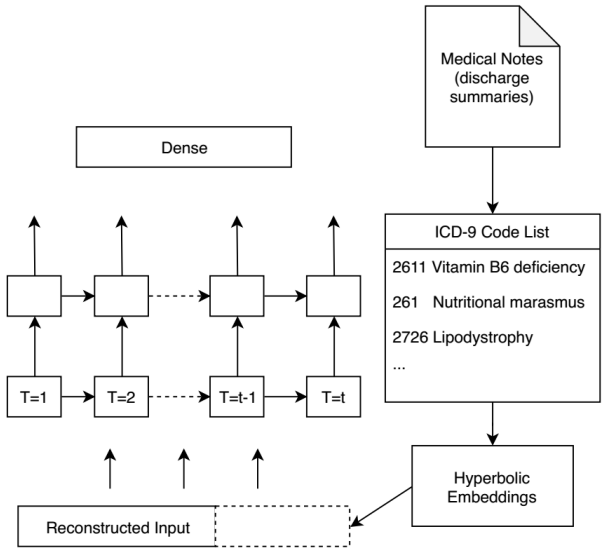


Figure 3: Framework of Mortality Prediction

center area of the space while the leaf nodes will be placed near the boundary area.

In order to learn embeddings from hierarchical medical ontologies, we follow the work of [17] to use Riemannian-SGD to optimize the loss function:

$$L = \sum_{(u,v)\in S} \log \frac{\exp^{-d_H(u,v)}}{\sum_{v'\in N(u)} \exp^{-d_H(u,v')}} \tag{2}$$

where $(u, v) \in S$ is a hierarchical (i.e., subclassof) relationship in a Knowledge Base (KB) S and $N(u) = \{v|(u, v) \notin S\} \cup \{u\}$ is a set of negative examples for u. Equation 2 can be observed as a soft ranking-loss, where related objects should be closer than objects for which we do not observe a relationship.

There are different kinds of medical ontologies that hierarchically organize medical concepts, including diseases, articles, medicines, etc. Since we conduct experiments on the MIMIC-III dataset, which encodes disease information of patients based on the 9-th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-9), we explore embeddings of the ICD-9 medical concepts for further evaluation.

3.2 Incorporating Textual Information from EHRs for Readmission Prediction

In this study, we propose to incorporate the medical text notes, i.e., the discharge summaries, to improve the prediction. We extract ICD-9 codes from the discharge summaries of ICU patients' EHRs, using an automatic medical code assignment tool for ICD-9 [19]. The extracted ICD-9 codes are then embedded with the method described in Section 3.1, the embeddings of which are used to reconstruct the input for the prediction model. In the experiment, we also use different graph embedding methods for fair comparison. Inspired by the use of deep learning models in Lin et al's approach [14], the framework of our method is shown in Figure 2.

Note that in Figure 2, the reconstructed input contains "medical notes ICD-9," which consists of the embeddings of the list of medical codes extracted from the textual notes (i.e., discharge summaries) in ICU patients' EHRs. Just like the "diagnosis ICD-9" of the original input [14], they are also in the form of embeddings [4]. But unlike the "diagnosis ICD-9" which is generated manually by professional coders, the "medical notes ICD-9" tends to be redundant, yet more informative for predicting ICU readmission. Thus, though it is possible that there exists some overlapping between the two lists, our hypothesis is that adding a new list of related ICD-9 codes will help improve the model. The experimental results do show that the reconstructed input demonstrates an advantage over the original.

3.3 Incorporating Embeddings for Mortality Prediction

Harutyunyan et al.'s work [6] is a widely accepted benchmark in in-hospital mortality prediction. We use their benchmark model for our experiment, which is an LSTM network that takes a 48-hour sequence of numerical features (e.g., the Glascow Coma Scale, Heart Rate, etc.) as input. To test our method of incorporating embeddings, we follow the same framework of the readmission prediction experiment. We concatenate the embeddings of the diagnoses of patients (in the form of ICD-9 codes), along with ICD-9 codes extracted from discharge summaries, to the original input and see if any performance gain can be achieved. The framework is shown in Figure 3.

4 EVALUATION

In this section, we evaluate our proposed method both intrinsically and extrinsically. For intrinsic evaluation, we test different embeddings of the ICD-9 ontology by comparing the similarities between medical concepts in the embedding spaces, to prove that the hyperbolic embedding method is a good fit for hierarchical representations, i.e., the ICD-9 ontology. For extrinsic evaluation, we test our method based on the state-of-the-art ICU readmission prediction model [14] and the in-hospital mortality prediction benchmark [6] on the MIMIC-III dataset, with different graph embeddings, to see (1) whether our method improves the performance of ICU readmission prediction and in-hospital mortality prediction; and (2) whether the hyperbolic embeddings of medical concepts from ICD-9 show any advantage over other prevalent embedding algorithms.

4.1 Intrinsic Evaluation

In this subsection, we intrinsically evaluate the different embeddings over the ICD-9 ontology. Basically, we want to compare and demonstrate how the similarities of medical concepts from ICD-9 are retained in the embedding spaces.

4.1.1 *Setup.*

Since we do not have a publicly available gold standard test where the similarities of medical concepts are assigned by professionals, nor the expertise to assign them by ourselves, we take an alternative by randomly selecting a certain number of pairs of medical concepts from ICD-9 (20,000 in our test), and computing the similarities between them based on several prevalent ontology-based similarity measurements, i.e., the Wu & Palmer similarity [32], the Leacock & Chodorow similarity [12], the Resnik similarity [22], and the RADA similarity [21]. Thus, we have 4 sequences of ontology-based term pair similarities over the same set of selected medical concepts.

We then compute the distance-based term pair similarities in the embedding spaces, and we evaluate the embeddings by comparing the Pearson Correlation Coefficients between the sequences of distance-based term pair similarities and the sequences of ontology-based term pair similarities. Note that for the hyperbolic embeddings (Poincaré), we compute the Poincaré distance to denote the *dissimilarity* based on Equation 1, and we use the negative value of it to denote the *similarity*. For Euclidean embeddings, we use the Euclidean distance as the *dissimilarity*, and convert it to *similarity* based on $s = \frac{1}{1+d}$.

Intuitively, higher correlation coefficients imply that the similarities between concepts are better retained in the corresponding embedding space.

The 4 ontology-based similarity measurements are defined as follows:

$$\text{Sim}_{\text{WUP}}(C_1, C_2) = \frac{2 * N_3}{N_1 + N_2 + 2 * N_3} \tag{3}$$

where N_1 and N_2 are the distance from the least common subsumer (LCS) to C_1 and C_2 respectively. N_3 is the depth of the least common subsumer. The least common subsumer of two concept nodes C_1 and C_2 is the lowest node that can be a parent for C_1 and C_2.

$$\text{Sim}_{\text{LCH}}(C_1, C_2) = -\log\left(\frac{\text{ShortestPath}(C_1, C_2)}{2 * depthmax}\right) \tag{4}$$

Table 1: Pearson Correlation Coefficients for Different Embeddings of ICD-9

Method	Dim	Measurement			
		WUP	LCH	RESNIK	RADA
Poincaré	10	**0.5720**	**0.6797**	**0.5784**	**0.7278**
	100	**0.5866**	**0.6902**	**0.5977**	**0.7351**
	300	**0.6042**	**0.7046**	**0.6007**	**0.7491**
ComplEx	10	0.4279	0.3169	0.4320	0.3036
	100	0.2265	0.2018	0.2094	0.1774
	300	0.1307	0.1432	0.1134	0.1141
DistMult	10	0.4297	0.3621	0.4174	0.3410
	100	0.1941	0.1827	0.1964	0.1521
	300	0.1204	0.1223	0.1161	0.0922
transE	10	0.0483	0.0269	0.0709	0.0130
	100	0.4159	0.3682	0.3494	0.3658
	300	0.4355	0.3958	0.3862	0.3912
Rescal	10	0.4108	0.2884	0.4364	0.2952
	100	0.2522	0.2756	0.1986	0.2523
	300	0.1243	0.1355	0.1166	0.1039

where *depthmax* is the maximum depth of any node in the tree and $\text{ShortestPath}(C_1, C_2)$ is the length of the shortest path between C_1 and C_2.

$$\text{Sim}_{\text{RESNIK}}(C_1, C_2) = \text{IC}(\text{LCS}(C_1, C_2)) \tag{5}$$

where LCS refers to the least common subsumer and IC refers to information content. Note that since we cannot compute the term frequency of the medical concepts, we use another ontology-based information content as an alternative [25]:

$$\text{IC}(c) = 1 - \frac{\log(\text{hypo}(c) + 1)}{\log(maxnodes)} \tag{6}$$

where hypo(c) refers to the number of hyponyms of concept c and *maxnodes* refers to the maximum number of concepts in the taxonomy.

$$\text{Sim}_{\text{RADA}}(C_1, C_2) = 2 * depthmax - \text{ShortestPath}(C_1, C_2) \tag{7}$$

where *depthmax* and $\text{ShortestPath}(C1, C2)$ are the same as Equation 4.

4.1.2 *Experiment.*

We randomly pick 20,000 concept pairs from the ICD-9 ontology and compute the mentioned 4 kinds of ontology-based similarities between them. Then we compute the distance-based similarities over these pairs for the several compared embeddings. Finally, we calculate the Pearson Correlation Coefficients between the above two kinds of sequences, as shown in Table 1.

Table 1 shows that the Poincaré embeddings significantly outperform the TransE [3], DistMult [34], ComplEx [28, 29] and Rescal [18] embeddings, in that the Poincaré embeddings show much higher correlation coefficients with the ontology-based similarity sequences. Generally it shows that the similarities between concepts are better retained in the hyperbolic embedding space than in the other embedding spaces.

Table 2: Performance on ICU Readmission Prediction Without Discharge Summaries

Embedding	Acc	Pre-0	Pre-1	Re-0	Re-1	**A.R**	A.P
Poincaré	0.7223	0.9035	0.3740	0.7361	0.6655	0.7786	0.4827
ComplEx	0.6621	0.9141	0.3306	0.6423	0.7454	0.7591	0.4236
Distmult	0.6426	0.9126	0.3172	0.6169	0.7508	0.7534	0.4243
TransE	0.7254	0.9062	0.3789	0.7366	0.6782	**0.7876**	0.4875
Rescal	0.6544	0.9160	0.3264	0.6303	0.7562	0.7661	0.4456

*Acc: Accuracy, Pre: Precision, Re: Recall, A.R: AUC under ROC, A.P: AUC under PRC

Table 3: Performance on ICU Readmission Prediction With Discharge Summaries

Embedding	Acc	Pre-0	Pre-1	Re-0	Re-1	**A.R**	A.P
Poincaré	0.7481	0.8993	0.4005	0.7766	0.6310	0.7851	0.4819
ComplEx	0.6705	0.9101	0.3342	0.6565	0.7263	0.7602	0.4341
Distmult	0.6678	0.9067	0.3303	0.6565	0.7151	0.7606	0.4327
TransE	0.7536	0.9039	0.4100	0.7779	0.6511	**0.7882**	0.4957
Rescal	0.6399	0.9209	0.3197	0.6067	0.7801	0.7684	0.4454

*Acc: Accuracy, Pre: Precision, Re: Recall, A.R: AUC under ROC, A.P: AUC under PRC

Table 1 also demonstrates that the Poincaré embeddings are capable of representing information with very few dimensions. As shown in this table, the Poincaré embeddings with low dimensions give good performance, similar to the one with higher dimensions. It thus proves that using hyperbolic-based embedding approaches is a good way to capture semantics in hierarchical data, such as ICD-9.

To sum up, in this subsection we intrinsically evaluate the hyperbolic embeddings over the ICD-9 medical ontology by comparing such with other graph embedding methods. The experimental results demonstrate that the method works well and outperforms other embedding approaches.

4.2 Extrinsic Evaluation 1: 30-day Unplanned ICU Readmission Prediction

In this subsection, we evaluate our proposed method described in Section 3.2, to see whether any performance improvement on 30-day unplanned ICU readmission prediction can be gained. Moreover, since we apply the hyperbolic embeddings of the ICD-9 medical ontology in the proposed method, this subsection can also be regarded as an extrinsic evaluation test for the medical embeddings.

4.2.1 Setup.

This portion of our experiments is conducted based on the MIMIC-III Critical Care (Medical Information Mart for Intensive Care III) Database, which is a large, freely-available database composed of deidentified health-related EHR data associated with over 40,000 patients who stayed in the critical care units (ICU) of the Beth Israel Deaconess Medical Center between 2001 and 2012.

The database contains a large variety of EHR data of ICU patients, including basic demographic information, bedside vital sign measurements, laboratory test results, medications, procedures, medical text notes (e.g., discharge summaries), and so on.

In this experiment, we follow the data preprocessing procedure of [6, 14] and generate a dataset of 48,411 ICU stay records. Each ICU stay record corresponds to one ICU patient, and each patient may have multiple ICU stay records. We then split the entire dataset into the training set (80%), the validation set (10%) and the testing set (10%) for further evaluation.

For fair comparison, we use the same setup and benchmark with Lin et al. [14] and consider 4 types of positive ICU stay records, including the patients (and the corresponding ICU stay record) who were transferred to low-level wards from ICU and readmitted to ICU later; the patients who were transferred out of ICU and died later; the patients who were discharged and readmitted to ICU later; and the patients who were discharged and died later. Note that the "later" here means "within 30 days."

4.2.2 Experiment.

We experiment with the hyperbolic embeddings (Poincaré) of the ICD-9 ontology and several state-of-the-art graph embedding methods, i.e., ComplEx, Distmult, TransE and Rescal. The results of using our method, with different embeddings, on the ICU readmission prediction task are shown in Table 2 and Table 3.

Note that in the readmission prediction task, most researchers are using the Area Under the Receiver Operating Characteristics curve (AUROC) as the main metric to evaluate their approaches. Generally, a higher AUROC score means a better model, for this task. Along with AUROC (A.R), some additional metrics are proposed, to better illustrate the comparison. However, these additional metrics can be unstable, and they are better used for additional evaluation.

In Table 2, we present the performance of different embeddings without ICD-9 codes extracted from discharge summaries. Note that in Table 2 we only use the human-annotated ICD-9 codes for each patient, without using any extractions from the discharge summaries. In Table 3, we present the corresponding results with ICD-9 codes extracted from discharge summaries as described in

Table 4: Performance on Readmission Prediction with Different Dimensions of Poincaré Embeddings

Embedding	dim	Acc	Pre-0	Pre-1	Re-0	Re-1	**A.R**	A.P
Poincaré	100	0.7086	0.8769	0.3410	0.7440	0.5590	**0.7193**	0.4098
Poincaré	10	0.6721	0.8886	0.3214	0.6796	0.6403	0.7165	0.3994
TransE	100	0.7115	0.8787	0.3455	0.7461	0.5655	0.7101	0.4067
TransE	10	0.7218	0.8702	0.3475	0.7710	0.5146	0.7099	0.3930

Table 5: Performance on Mortality Prediction Without Discharge Summaries

Embedding	Acc	Pre-0	Pre-1	Re-0	Re-1	**A.R**	A.P
Poincaré	0.8814	0.8977	0.6350	0.9737	0.2912	0.8722	0.5543
ComplEx	0.8947	0.9165	0.6701	0.9662	0.4380	0.8915	0.6104
Distmult	0.8988	0.9152	0.7115	0.9730	0.4243	**0.8956**	0.6247
TransE	0.8888	0.9019	0.6930	0.9777	0.3211	0.8854	0.5717
Rescal	0.8913	0.9019	0.7239	0.9809	0.3188	0.8954	0.6025

*Acc: Accuracy, Pre: Precision, Re: Recall, A.R: AUC under ROC, A.P: AUC under PRC

Section 3.2. It shows that adding extra ICD-9 codes from discharge summaries does improve the overall performance on this readmission prediction task. It also shows that the Poincaré embeddings outperform all other graph embedding methods except TransE. Note that we also test this method with the Claims embeddings [4] that are used by Lin et al. [14], the results of which (0.7943) also demonstrates an advantage over their best reported A.R score (0.791). We do not think it is fair to compare Lin et al.'s results [14] with the reported graph embedding methods in Table 2 and 3, because they only use ICD-9 codes to generate embeddings.

As is described in Section 2.1, hyperbolic embeddings have the ability to represent hierarchical data with lower dimensions. So, in Table 4, we test our method using the Poincaré and TransE embeddings with lower dimensions. The results are consistent, showing that lower dimensions of Poincaré embeddings give better performance than that of TransE, especially when in 300 dimensions TransE actually does better than Poincaré.

To sum up, in this subsection we evaluate our method in the ICU readmission prediction task, and we also extrinsically evaluate the hyperbolic embeddings of the ICD-9 ontology. The results prove the effectiveness of our method by showing a better AUROC over the model without discharge summaries. The results also demonstrate the good qualities of the hyperbolic embeddings, in that they give comparable performance with the state-of-the-art graph embeddings methods.

4.3 Extrinsic Evaluation 2: In-Hospital Mortality Prediction

In this subsection, we further evaluate our method and the hyperbolic embeddings of the ICD-9 medical ontology by incorporating the embeddings into existing methods of in-hospital mortality prediction and comparing their performance.

4.3.1 Setup.

This part of our experiments is also conducted on the MIMIC-III dataset [8]. We follow the data preprocessing pipeline with the benchmark [6]. The data contains 42, 276 ICU stays of 33, 798 unique, de-identified patients, who are at least 18 years old. For fair comparison, we adopt the same split of 15% for validation and 85% for training.

4.3.2 Experiment.

To be consistent with the experiment on 30-day unplanned ICU readmission prediction, we experiment with the same group of graph embeddings (i.e., Poincaré, ComplEx, Distmult, TransE and Rescal). The results of our method with different embeddings on in-hospital mortality prediction are shown in Table 5 and Table 6.

In the task of in-hospital mortality prediction, Area Under the Receiver Operating Characteristics curve (AUROC) is still the widely accepted metric for evaluation. Higher AUROC indicates better performance of a model for a certain embedding method. In Table 5 and 6, the same set of metrics are represented for additional comparison.

The work of Harutyunyan et al. [6] is a widely accepted benchmark for mortality prediction, which gives an A.R score of 0.8607. As in the readmission experiment, we present the results with and without ICD-9 codes extracted from discharge summaries. Note that in Table 5 we only use the human-annotated ICD-9 codes for each patient, without using any extractions from the discharge summaries. It shows that adding ICD-9 codes can generally improve the performance of mortality prediction with different embeddings. Every embedding method in the experiment leads to an improvement on the AUROC (A.R) score over the baseline (0.8607). In Table 7, we test the Poincaré embeddings with different dimensions, and the results are very stable, which is consistent with the earlier assumption.

In summary, we extrinsically evaluate our method and the hyperbolic embeddings of the ICD-9 medical ontology on the task of in-hospital mortality prediction. The results prove the effectiveness of our method by representing higher AUROC than the benchmark, though in this task the hyperbolic embeddings do not outperform all other embeddings. However, adding ICD-9 codes extracted from

Table 6: Performance on Mortality Prediction With Discharge Summaries

Embedding	Acc	Pre-0	Pre-1	Re-0	Re-1	**A.R**	A.P
Poincaré	0.8882	0.9128	0.6338	0.9626	0.4128	0.8756	0.5760
ComplEx	0.8972	0.9012	0.8220	0.9895	0.3073	0.8958	0.6312
Distmult	0.8941	0.9267	0.6330	0.9529	0.5183	0.8959	0.6218
TransE	0.8882	0.8995	0.7043	0.9802	0.3004	0.8852	0.5771
Rescal	0.8929	0.9141	0.6654	0.9669	0.4197	**0.8979**	0.6042

*Acc: Accuracy, Pre: Precision, Re: Recall, A.R: AUC under ROC, A.P: AUC under PRC

Table 7: Performance on Mortality Prediction with Different Dimensions of Poincaré Embeddings

Embedding	dim	Acc	Pre-0	Pre-1	Re-0	Re-1	**A.R**	A.P
Poincaré	300	0.8882	0.9128	0.6338	0.9626	0.4128	0.8756	0.5760
Poincaré	100	0.8938	0.9081	0.7061	0.9759	0.3692	**0.8789**	0.5841
Poincaré	10	0.8904	0.9100	0.6627	0.9691	0.3876	0.8755	0.5900

discharge summaries does improve the overall performance with almost every embedding method on the task of mortality prediction, which is consistent with the readmission prediction experiment.

5 CONCLUSION AND FUTURE WORK

In this study, we present a new method to incorporate the medical notes in patients' EHR data to improve the state-of-the-art model on ICU readmission prediction, and the widely accepted benchmark on in-hospital mortality prediction. We specially leverage the hyperbolic embeddings of the ICD-9 ontology in our proposed method. To the best of our knowledge, we are the first to do so and achieve promising results.

We are exploring the following directions for further research:

(1) Though the hyperbolic embeddings (Poincaré) of ICD-9 perform well, they are learned from very limited information (i.e., the hierarchical structure). We will try to train a better and joint embedding with the hierarchy of medical ontologies, the textual descriptions of concepts, and even patients' EHR data in the hyperbolic space.
(2) In the mortality prediction task, the hyperbolic embeddings are not as good as all the other embeddings. We will try to incorporate task-specific supervision to derive the embeddings of medical concepts.
(3) Readmission and mortality predictions are two important tasks for the healthcare system. We will explore more sophisticated and effective models to solve these tasks. Based on our research, we have found that useful and valuable information exists in the medical notes (e.g., discharge summaries) of patients' EHRs, which supports prediction; so we will try new ways to leverage the textual information.

ACKNOWLEDGMENTS

This research is supported by NSF grant CNS-1747798 to the IUCRC Center for Big Learning.

REFERENCES

[1] Christopher Baechle, Ankur Agarwal, Ravi Behara, and Xingquan Zhu. 2017. Latent topic ensemble learning for hospital readmission cost reduction. In *2017 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 4594–4601.
[2] Olivier Bodenreider. 2004. The unified medical language system (UMLS): integrating biomedical terminology. *Nucleic acids research* 32, suppl_1 (2004), D267–D270.
[3] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *Advances in neural information processing systems*. 2787–2795.
[4] Youngduck Choi, Chill Yi-I Chiu, and David Sontag. 2016. Learning low-dimensional representations of medical concepts. *AMIA Summits on Translational Science Proceedings* 2016 (2016), 41.
[5] Bhuwan Dhingra, Christopher J Shallue, Mohammad Norouzi, Andrew M Dai, and George E Dahl. 2018. Embedding Text in Hyperbolic Spaces. *NAACL HLT 2018* (2018), 59.
[6] Hrayr Harutyunyan, Hrant Khachatrian, David C. Kale, and Aram Galstyan. 2017. Multitask Learning and Benchmarking with Clinical Time Series Data. *CoRR* abs/1703.07771 (2017).
[7] Alistair EW Johnson, Andrew A Kramer, and Gari D Clifford. 2014. Data preprocessing and mortality prediction: the Physionet/CinC 2012 challenge revisited. In *Computing in Cardiology Conference (CinC)*. IEEE, 157–160.
[8] Alistair EW Johnson, Tom J Pollard, Lu Shen, H Lehman Li-wei, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. 2016. MIMIC-III, a freely accessible critical care database. *Scientific data* 3 (2016), 160035.
[9] Le Gall JR, Loirat P, Alperovitch A, Glaser P, Granthil C, Mathieu D, Mercier P, Thomas R, and Villers D. 1984. A simplified acute physiology score for ICU patients. *Crit Care Med* 12, 11 (1984), 975–977.
[10] Le Gall JR, Lemeshow S, and Saulnier F. 1993. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *JAMA* 270, 24 (1993), 2957–2963.
[11] Denis Krompaß, Cristóbal Esteban, Volker Tresp, Martin Sedlmayr, and Thomas Ganslandt. 2015. Exploiting latent embeddings of nominal clinical data for predicting hospital readmission. *KI-Künstliche Intelligenz* 29, 2 (2015), 153–159.
[12] Claudia Leacock and Martin Chodorow. 1998. Combining local context and WordNet similarity for word sense identification. *WordNet: An electronic lexical database* 49, 2 (1998), 265–283.
[13] Matthias Leimeister and Benjamin J Wilson. 2018. Skip-gram word embeddings in hyperbolic space. *arXiv preprint arXiv:1809.01498* (2018).
[14] Yu-Wei Lin, Yuqian Zhou, Faraz Faghri, Michael J Shaw, and Roy H Campbell. 2018. Analysis and Prediction of Unplanned Intensive Care Unit Readmission using Recurrent Neural Networks with Long Short-Term Memory. *bioRxiv* (2018), 385518.
[15] Carolyn E Lipscomb. 2000. Medical subject headings (MeSH). *Bulletin of the Medical Library Association* 88, 3 (2000), 265.
[16] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*. 3111–3119.

[17] Maximillian Nickel and Douwe Kiela. 2017. Poincaré embeddings for learning hierarchical representations. In *Advances in neural information processing systems*. 6338–6347.

[18] Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. 2011. A Three-Way Model for Collective Learning on Multi-Relational Data.. In *ICML*, Vol. 11. 809–816.

[19] Adler Perotte, Rimma Pivovarov, Karthik Natarajan, Nicole Weiskopf, Frank Wood, and Noémie Elhadad. 2013. Diagnosis code assignment: models and evaluation metrics. *Journal of the American Medical Informatics Association* 21, 2 (2013), 231–237.

[20] Carolina R Ponzoni, Thiago D Corrêa, Roberto R Filho, Ary Serpa Neto, Murillo SC Assunção, Andreia Pardini, and Guilherme PP Schettino. 2017. Readmission to the intensive care unit: incidence, risk factors, resource use, and outcomes. A retrospective cohort study. *Annals of the American Thoracic Society* 14, 8 (2017), 1312–1319.

[21] Roy Rada, Hafedh Mili, Ellen Bicknell, and Maria Blettner. 1989. Development and application of a metric on semantic nets. *IEEE transactions on systems, man, and cybernetics* 19, 1 (1989), 17–30.

[22] Philip Resnik. 1995. Using information content to evaluate semantic similarity in a taxonomy. In *Proceedings of the 14th international joint conference on Artificial intelligence-Volume 1*. Morgan Kaufmann Publishers Inc., 448–453.

[23] A Rumshisky, M Ghassemi, T Naumann, P Szolovits, VM Castro, TH McCoy, and RH Perlis. 2016. Predicting early psychiatric readmission with natural language processing of narrative discharge summaries. *Translational psychiatry* 6, 10 (2016), e921.

[24] Frederic Sala, Chris De Sa, Albert Gu, and Christopher Ré. 2018. Representation tradeoffs for hyperbolic embeddings. In *International Conference on Machine Learning*. 4457–4466.

[25] Nuno Seco, Tony Veale, and Jer Hayes. 2004. An intrinsic information content metric for semantic similarity in WordNet. In *ECAI*, Vol. 16. 1089.

[26] Vergil N Slee. 1978. The International classification of diseases: ninth revision (ICD-9). *Annals of internal medicine* 88, 3 (1978), 424–426.

[27] Michael Q Stearns, Colin Price, Kent A Spackman, and Amy Y Wang. 2001. SNOMED clinical terms: overview of the development process and project status.. In *Proceedings of the AMIA Symposium*. American Medical Informatics Association, 662.

[28] Théo Trouillon, Christopher R Dance, Éric Gaussier, Johannes Welbl, Sebastian Riedel, and Guillaume Bouchard. 2017. Knowledge graph completion via complex tensor factorization. *The Journal of Machine Learning Research* 18, 1 (2017), 4735–4772.

[29] Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. 2016. Complex embeddings for simple link prediction. In *International Conference on Machine Learning*. 2071–2080.

[30] Knaus WA, Draper EA, and Wagner DP. 1985. APACHE II: a severity of disease classification system. *Crit Care Med* 13, 10 (1985), 818–829.

[31] Knaus WA, Zimmerman JE, Wagner DP, Draper EA, and Lawrence DE. 1981. APACHE-acute physiology and chronic health evaluation: a physiologically based classification system. *Crit Care Med* 9, 8 (1981), 591âĂŞ597.

[32] Zhibiao Wu and Martha Palmer. 1994. Verbs semantics and lexical selection. In *Proceedings of the 32nd annual meeting on Association for Computational Linguistics*. Association for Computational Linguistics, 133–138.

[33] Y Xue, D Klabjan, and Luo Yuan. 2018. Predicting ICU readmission using grouped physiological and medication trends. *Artificial intelligence in medicine* (2018), 4.

[34] Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2014. Embedding entities and relations for learning and inference in knowledge bases. *arXiv preprint arXiv:1412.6575* (2014).