

MASTER'S THESIS

Medical concept embedding with ontological representations

Song, Lihong

Date of Award:
2019

[Link to publication](#)

General rights

Copyright and intellectual property rights for the publications made accessible in HKBU Scholars are retained by the authors and/or other copyright owners. In addition to the restrictions prescribed by the Copyright Ordinance of Hong Kong, all users and readers must also observe the following terms of use:

- Users may download and print one copy of any publication from HKBU Scholars for the purpose of private study or research
- Users cannot further distribute the material or use it for any profit-making activity or commercial gain
- To share publications in HKBU Scholars with others, users are welcome to freely distribute the permanent URL assigned to the publication

Abstract

Learning representations of medical concepts from the Electronic Health Records (EHRs) has been shown effective for predictive analytics in healthcare. The learned representations are expected to preserve the semantic meanings of different medical concepts, which can be treated as features and thus benefit a variety of applications. Medical ontologies have also been explored to be integrated with the EHR data to further enhance the accuracy of various prediction tasks in healthcare. Most of the existing works assume that medical concepts under the same ontological category should share similar representations, which however does not always hold. In particular, the categorizations in the categorical medical ontologies were established with various factors being considered. Medical concepts even under the same ontological category may not follow similar occurrence patterns in the EHR data, leading to contradicting objectives for the representation learning. In addition, these existing works merely utilize the categorical ontologies. Actually, it has been noticed that ontologies containing multiple types of relations are also available. However, studies rarely make use of the diverse types of medical ontologies.

In this thesis research, we propose three novel representation learning models for integrating the EHR data and medical ontologies for predictive analytics. To improve the interpretability and alleviate the conflicting objective issue between the EHR data and medical ontologies, we propose techniques to learn medical concepts embeddings with multiple ontological representations. To reduce the reliance on labeled data, we treat the co-occurrence statistics of clinical events as additional training signals, which help us learn good representations even with few labeled data. To

leverage the various domain knowledge, we also consider multiple medical ontologies (CCS, ATC and SNOMED-CT) and propose corresponding attention mechanisms so as to take the best advantage of the medical ontologies with better interpretability. Our proposed models can achieve the final medical concept representations which align better with the EHR data. We conduct extensive experiments, and our empirical results prove the effectiveness of the proposed methods.

Keywords: Bio/Medicine, Healthcare-AI, Electronic Health Record, Representation Learning, Machine Learning Applications

Table of Contents

- Declaration i
- Abstract ii
- Acknowledgements iv
- Table of Contents vi
- List of Tables x
- List of Figures xii
- Chapter 1 Introduction 1**
 - 1.1 Medical Concept Embedding for Predictive Analytics in Healthcare . 1
 - 1.2 Contribution 3
 - 1.2.1 Medical Concept Embedding with Multiple Ontological Rep-
resentations (**Chapter 3** and **Chapter 4**) 4
 - 1.2.2 Medical Concept Embedding with Multi-relational Medical
Ontology Representations (**Chapter 5**) 6
 - 1.3 Thesis Organization 7
- Chapter 2 Related Work 8**
 - 2.1 EHR Data Mining 8
 - 2.2 Representation Learning Methods 11
 - 2.3 Neural Networks 13

2.3.1	Multi-Layer Perceptron	13
2.3.2	Recurrent neural networks	14
2.4	Attention Mechanism	17
2.5	Knowledge Graph Embedding	19
2.6	Summary	21
Chapter 3	Medical Concept Embedding with Multiple Ontological Representations	22
3.1	Introduction	22
3.2	Notations and Preliminaries	25
3.2.1	Basic Notations	25
3.2.2	Learning Representations from the Co-occurrence Statistics	25
3.3	Proposed Model	26
3.3.1	Learning Multiple Ontological Representations	28
3.3.2	Interpretability-Enhanced Predictive Analytics	29
3.4	Experiments	31
3.4.1	Next-admission Diagnosis Prediction	32
3.4.2	Interpretability of the Multiple Ontological Representation	34
3.4.3	Interpretation of the Predictive Attention Patterns (Phenotypes)	40
3.5	Summary	40
Chapter 4	Medical Concept Embedding with Automatically Multiple Ontological Representations	43
4.1	Introduction	43
4.2	Notations and Preliminaries	44
4.2.1	Basic Notations	44
4.3	Proposed Model	45
4.3.1	Automatically Learning Multiple Ontological Representations	46
4.4	Experiments	48

4.4.1	Interpretability of the Multiple Ontological Representations	52
4.4.2	Interpretation of the Predictive Attention Patterns (Phenotypes)	57
4.5	Summary	57
Chapter 5 Medical Concept Embedding with Multi-relational Medical Ontology Representations		59
5.1	Introduction	59
5.2	Notations and Preliminaries	60
5.2.1	Basic Notation	60
5.2.2	Relational Graph Convolutional Networks	62
5.2.3	Knowledge Graph Embedding	62
5.3	Proposed Model	63
5.3.1	Categorical Ontology Attention Layer	63
5.3.2	Multi-relational Attention Layer	64
5.4	Experiments	65
5.4.1	Data and Ontology Preprocessing	66
5.4.2	Performance Evaluation	67
5.4.3	Experimental Results	68
5.4.4	Interpretability of the Multi-relational Medical Ontology Representations	69
5.4.5	Interpretation of the Multi-relational Attention Patterns	72
5.5	Summary	72
Chapter 6 Conclusion and Future Work		75
6.1	Concluding Remarks	75
6.2	Future Research Directions	76
Bibliography		76
List of Publications		90

