Enriching Medical Natural Language Processing via Large-scale Concept Graph from Electronic Medical Records

Xiaokang Jiao , **Xinhui Du** , **Zengtao Jiao*** , **Kangning Cai** and **Jun Yan**
Yidu Cloud Beijing Technology Co., Ltd.
{xiaokang.jiao, xinhui.du, zengtao.jiao, kangning.cai, jun.yan}@yiducloud.cn

Abstract

Artificial intelligence in the medical field is becoming in great demand recent years, but there are still many obstacles to conquer. Applied semantics extraction is one of the key technologies to study. An electronic medical record (EMR) in hospital is the detail record of doctor's diagnosis and treatment, where rich medical knowledge lies. However, considering of the complexity of the clinical process, the diversity of doctors' professional competence level, and even the errors in electronic system, it is difficult to extract knowledge directly from EMR. When building a medical AI application system, it is the common way to customize rules of data extraction based on different target tasks, then to label data by medical experts, and the premise of the application system is finally achieved. These steps are costly and have limited versatility. In this paper, we propose a highly efficient method to perform semantics extraction and analytics with most labor free in medical AI tasks, named as medical knowledge embedding (MKE). Concept graph as the form of medical knowledge is constructed from EMR, and concept embedding representation is generated by graph neural network (GNN) algorithm, which can be used in other medical applied tasks. Due to the limitation on public medical data, a medical named entity recognition (NER) task is carried on to test and verify this concept representation. The experiment shows that this work can outperform classic BILSTM+CRF and the best model in the medical NER track of CCKS 2019.

1 Introduction

In recent years, artificial intelligence (AI) has already been developed in the medical area for decades, otherwise there are lots of difficulties to conquer. The rare access to large scale valuable clinical data is one barrier because of privacy and ethical issues. Besides, the labeling cost of medical data is another obstacle due to high price of medical experts. All these difficulties restrict the development of medical AI to a great extent. Applied semantics extraction and analytics from clinical electronic medical record (EMR) are play a key role to improve current situation.

EMR in hospital is the detail record of doctor's diagnosis and treatment, where rich medical knowledge lies. However, the allowance to access EMR must be granted from ethics committee in hospital to use these data. So, fully utilizing publicly released medical knowledge bases becomes second choice to improve medical AI. Now more and more knowledge bases appear, such as the systematized Nomenclature of Human and Veterinary Medicine Clinical Terms (SNOMED-CT)[1], Unified Medical Language System (UMLS)[2]. These knowledge bases have played important roles in the field of medical data analysis and medical artificial intelligence. Nevertheless, they contain more common sense than professional knowledge in a vertical field of medicine, such as tumor or heart failure, besides the cost of construction is very extremely high.

Even when clinical EMR is accessible, it is difficult to extract knowledge directly, considering of the complexity of the clinical process, the diversity of doctors' professional competence level, and even the errors of electronic system. For building a medical AI application system, it is the usual way to customize rules of data extraction based on different target tasks, then to label data by medical experts, and the premise of the application system is finally achieved. These steps are costly and have limited versatility.

This paper proposes a method to perform semantics extraction and analytics in medical AI tasks with high efficiency, named as medical knowledge embedding (MKE). Concept graph as the form of medical knowledge is constructed from EMR, and medical concept embedding representation is generated by graph neural network (GNN) algorithm, which can be used in other medical applied tasks. A medical named entity recognition task is carried on used the concept representation, and the experiment shows that the semantics embedding of concept graph from EMR can enrich prior information to improve other natural language processing tasks in the medical field.

*zengtao.jiao@yiducloud.cn

[1] $https://www.snomed.org/$
[2] $https://www.nlm.nih.gov/research/umls/index.html$

2 Related Work

EMR has accumulated a large amount of real medical records in hospital, which is bound to attract the interest and attention of researchers in the industry. Dalta et al. developed an automatic clinical diagnosis system based on mimic-iii medical record dataset [Datla *et al.*, 2017]. Goodwin et al. used this dataset to study the knowledge embedding in EMR information extraction [Goodwin and Harabagiu, 2016]. Harabagiu et al. studied EMR information extraction and query expansion on i2b2 dataset [Goodwin and Harabagiu, 2013]. And a clinical support system is developed by Ling [Ling, 2017]. Kang et al. used UMLS to study the concept mapping of EMR terms [Kang *et al.*, 2009]. Khare et al. proposed a method to build a correlation by capturing the assertions related to medical concepts and design a calculation method to predict the indications of new drugs by using the drug labels, which is derived from PubMed and relationship among drug, disease and treatment annotated in LabeledIn [Khare *et al.*, 2014]. Medical knowledge from EMR is a very effective enhancer to medical-applied AI task as above.

Aside from word embedding, which learns word representation via corpus only, graph embedding enhances the representation ability by capturing more relevant information from graph. DeepWalk performs random walks on the graph on the basis of edge weights DeepWalk [Perozzi *et al.*, 2014]. From the sampled sequence, DeepWalk learns the embedding through local predictions on nodes' neighborhood. Furthermore, LINE learns the embedding based on breadth-first search schemes and reserves both the first-order and second-order proximities [Tang *et al.*, 2015]. Node2vec extends DeepWalk by tuning the weights used in random walks to balance homophily and structural equivalence [Grover and Leskovec, 2016]. For all these methods, graph embedding exploits rich information from graph to enhance knowledge representations, which could benefit downstream NLP tasks.

To date, along with other representation techniques, graph embedding methods has provide a variety of real world medical applications. For general representation learning, Wu et al. proposed a graph-based, hierarchical medical entity embedding framework Med2vec to learn representation via EHR data [Wu *et al.*, 2019]. For applications related to link predictions, by performing study on large-scale real-world EMR data, Li et al. introduced a PrTransX algorithm to learn the embedding vectors of a probabilistic knowledge graph (KG) and demonstrate its performance on predictions of disease relations [Li *et al.*, 2020]. Wang et al. proposed a novel framework SMR to provide safe medicine recommendations by constructing a heterogeneous graph (Wang et al. 2017)[12]. Furthermore, Celebi et al. proved that the knowledge embeddings are significant predictors for inferring new drug-drug interactions [Celebi *et al.*, 2019]. For entity recognition, Sun et al. built a graph-based model to detect suspicious claims with inappropriate diagnose medications by integrating the dictionary-based features with the embeddings [Sun *et al.*, 2020]. Among different graph embedding methods, Mao & Fung applied graph convolutional networks (GCN) to measure semantic relatedness between UMLS concepts [Mao and Fung, 2020]. Their graph-based methodology outperforms corpus-based word embeddings.

3 Methods

The schema of our work firstly builds the concept graph from real EMRs, and then generates embedding representation via GCN. The embedding representation as MKE can be used to improve other NLP tasks. The procedure to extract MKE is illustrated below.

3.1 Concept Graph

The concept graph is constructed through dozens of application projects for hospitals scattered all over China which was done by our works in the past, and these projects obtain approval by the ethics board of all hospitals. The graph consists of relevant edges among medical concepts, which include normalized names of symptom, diagnosis, drug, surgery and laboratory test from EMR (Electronic Medical Record) in hospital. The procedure is described as below.

First, the co-occurrence between medical concepts is derived from every EMR. EMR usually consists of semi-structured data such as table, image, free text, etc. The names of diagnosis, drug, surgery, abnormal laboratory test in the graph can be found in specific fields of structured table, such as tables of discharge record, drug order, surgery record, and laboratory test record. The symptom names must yet be extracted from chief complaint or history of present illness through text structure tools developed by previous works, which were developed in terms of the standard information extraction algorithm done with entity and relationship recognition, customization of output medical logic using entity and relationship.

Second, the concept names directly extracted from EMR are expressed with personalized manners, which causes difficulty in identify concept from different EMR in hospital, so the normalization is carried on after the co-occurrence extraction. The diagnosis names are normalized to ICD10[3], which is a diagnostic coding standard called International Classification of Diseases. The surgery names are normalized to code of procedure classification from ICD-9-CM-3[4] for compatibility with historical clinical settings. The drug names are normalized to ATC[5]. The abnormal laboratory test names are normalized to code of LOINC[6]. Nevertheless, lack of public standards, the symptom names are normalized to a schema based on previous clinical knowledge according to symptomatic anatomic site and types of symptom. Considering co-occurrence as an edge, the frequency can be counted after normalization.

Finally, the graph needs be pruned due to the inadequate data quality of EMR, such as mismatch between tables of EMR and error in normalization of concept. Besides, there are odd events in the process of clinical diagnosis and treatment. A statistical method is carried out to quantify the confidence of edge called OR (odds ratio). The OR [Schechtman,

[3] $https://icd.who.int/browse10/2016/en$

[4] $https://www.cdc.gov/nchs/icd/icd9cm.htm$

[5] $https://www.who.int/tools/atc-ddd-toolkit/atc-classification$

[6] $https://loinc.org/$

Category	SYM	DIA	DRU	SUR	LAB
Count	7563	9052	1381	423	1123

Table 1: Concept number of different categories

Edge between concepts		Count
DIA	DIA	105740
DIA	SYM	49417
DIA	DRU	220540
DIA	SUR	16491
DIA	LAB	183178
SYM	SYM	63333
SYM	DRU	20853
DRU	DRU	36811

Table 2: Edge number between concepts

2002] is a measure of association between exposure and outcome. It represents the odds when an outcome occurs given a particular exposure, compared to the odds the outcome occurring in the absence of that exposure. OR>1 indicates positive correlation, otherwise OR<1 indicates negative correlation. The bigger OR is, the more strength of relevance between exposure and outcome is. In this scenario, the nodes on both sides of an edge in concept graph can be considered as exposure and outcome. Besides, the OR is the weight of the edge. Thus the graph is pruned followed by two criteria: OR is bigger than 1.5; frequency of co-occurrence is bigger than 10.

Following above procedure, the concept graph is constructed, which consists of 19542 nodes and 696363 edges. An example of sub-graph is in Figure 1 centering by "冠状病毒感染" (coronavirus infections). The statistics are shown in Table 1 and Table 2. In these tables, SYM is denoted as symptom, DIA as diagnosis, DRU as drug, SUR as surgery, and LAB as laboratory test. Edges besides eight types in Table 2 are ignored as lack of confidence.

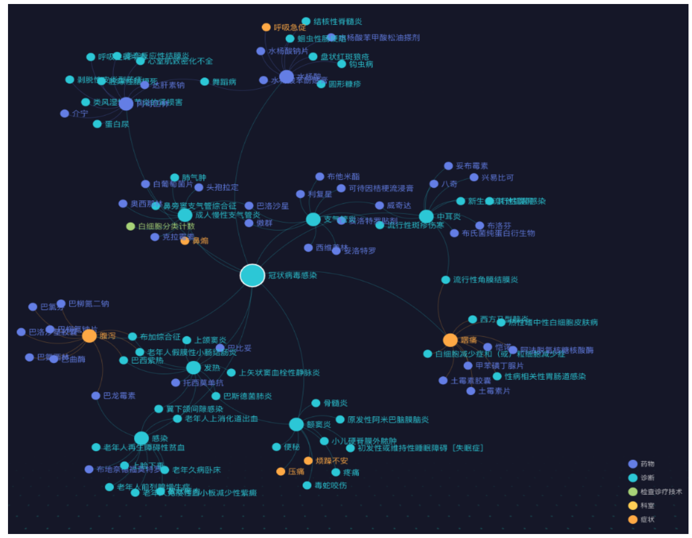

Figure 1: An example subgraph of the concept graph

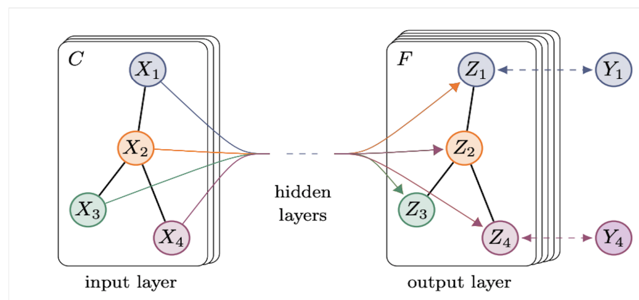

Figure 2: Graph structure inside each layer of GCN

3.2 Concept Embedding

With the concept graph constructed above, Graph Convolutional Networks could be pre-trained to generate entity embeddings [Kipf and Welling, 2017]. Graph convolutional network (GCNs) are variants of graph neural network (GNNs). Gori et al. firstly proposed the architecture of Graph Neural Networks and Scarselli et al. expounded GNNs with more details [Gori *et al.*, 2005; Scarselli *et al.*, 2009]. They use recurrent neural networks to propagate neighbor information until a stable state is reached. Then representation of the target node could be learned through propagation process.

Inspired by CNNs' great success in computer vision, many approaches are dedicated to incorporate convolution methods into GNNs. Kipf & Welling proposed the definition of Graph Convolutional Network, which performs convolution directly on graph structure [Kipf and Welling, 2017]. To explain GCNs, we assume to have a graph of N nodes. Each node has D features. The feature matrix could mark as H with $N \times D$ dimensions. Relationship between each node can be represented as an adjacency matrix A of $N \times N$.

The layer-wise propagation rule will be:

$$H^{l+1} = \sigma(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^l W^l) \quad (1)$$

In the above formula, $\tilde{A} = A + I$ and I is the identity matrix to avoid empty value in diagonal of adjacency matrix. H contains the features for each layer. $\tilde{D}$ is the degree matrix of $\tilde{A}$, which helps to maintain original distribution of H. W^l is a layer-specific trainable weight matrix. σ is the non-linear activation function.

The formula explains the basic transmission mechanism for GCNs. For a specific task, more details could be explained by Figure 2 from Kipf & Welling's paper [Kipf and Welling, 2017], C is the input graph and X_i denotes the feature in C. Finally, X_i will change to be Z_i through propagation. The relationship between each node is shared for each layer in GCNs.

Assuming we have two-layer GCN and we use ReLU and softmax for activation functions. Pre-processing step for adjacency matrix A:

$$\hat{A} = \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} \quad (2)$$

Then the forward propagation will be:

$$Z = f(X, A) = softmax(\hat{A} ReLU(\hat{A} X W^{(0)}) W^{(1)}) \quad (3)$$

For nodes that have labels, denoted as $\mathcal{Y}_L$, we could calculate the cross entropy:

$$\mathcal{L} = - \sum_{l \in \mathcal{Y}_L} \sum_{f=1}^{F} Y_{lf} ln Z_{lf} \tag{4}$$

Then the model could be trained by this node classification task. Link predictions and other tasks will be similar by adjusting loss functions. With the trained GCNs, we could then perform feature extraction to generate node embeddings, which is MKE as wanted.

4 Experiments

To verify the effect of the graph embedding, the public released datasets from real world Chinese EMR is our first choice for considering the privacy of EMR. So a natural language processing task is performed using the open source dataset from the track[7] , on named entity recognition for Chinese EMR in the 12th China Conference on Knowledge Graph and Semantic Computing (CCKS).

4.1 Dataset

The dataset is provided in the medical named entity recognition track of CCKS. A doctor team was organized to produce and annotate the dataset based on clinical experience. Training set consists of 600 documnets and testing set consists of 400 documents, which derived from history of present illness. The annotated data were de-identified and released to evaluation participants with data use agreements.

The entity type of this track focuses on three main categories: symptom, drug, and surgery from history of present illness. Since most of the symptoms appeared in Chinese EMR are structured, this task further subdivided the symptoms into three sub categories: the anatomical site (the subject of the symptoms), the symptom description (the description of the symptoms), and the independent symptoms. Finally, the predefined categories of this task are limited to the following five categories:

- Anatomical Site: The subject of a symptom which refers to a structural unit composed of a variety of tissues that perform certain function, such as “腹部” (abdomen).
- Symptom Description: The description of a symptom which refers to the patient’s own experience and feeling of abnormal physiological function. It needs to be combined with an anatomical part to express a complete symptom, such as “不适” (discomfort), combined with “腹部” (abdomen) to output “腹部不适” (abdominal discomfort).
- Independent Symptom: a complete symptom which can be output independently, such as “眩晕” (dizziness).
- Drug: A chemical substance used to treat, prevent diseases, or promote health.
- Surgery: The treatment of the patient’s body with a medical device such as resection, suturing.

The statistics of the annotated dataset are shown in the following Table 3.

[7] $https://biendata.xyz/competition/CCKS2018\_1/$

Dataset	Training Set	Testing Set
Anatomical Sites	7838	6339
Symptom Descriptions	2066	918
Independent Symptoms	3055	1327
Drugs	1005	813
Surgeries	1116	735
Total	15080	10132

Table 3: The statistics of annotated dataset

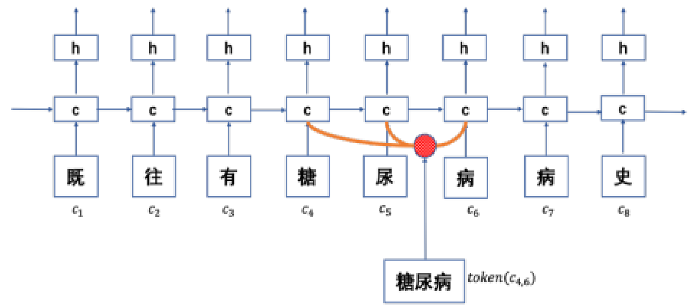

Figure 3: Lattice LSTM structure

4.2 Model Architectures

Named entity recognition (NER) has received constant attention for application and academic research over the recent years as a fundamental task in information extraction. Traditionally, NER has been solved as a sequence labeling problem [Liu *et al.*, 2017; Lample *et al.*, 2016; Long *et al.*, 2018; Liu *et al.*, 2019], where entity boundary and category labels are jointly predicted. The current state-of-the-art for English NER has been achieved by using BILSTM+CRF models [Liu *et al.*, 2017; Lample *et al.*, 2016]. Different from English, word segmentation is a basic task in Chinese, and the potential issue of error propagation can be suffered. So character-based NER is becoming more and more popular. One drawback of character-based NER, however, is that explicit word and word sequence information is not fully exploited, which can be potentially useful.

Inspired from lattice LSTM [Zhang and Yang, 2018], MKE is integrated into a character-based BILSTM+CRF model by representing concept names from the sentence using a lattice structure LSTM, which named as MKE-BILSTM+CRF. As showing in Figure 3, for example, a word-character lattice is constructed by matching a sentence with nodes from the concept graph, and the synonym is extended for the concept node. The red circle as “糖尿病” (diabetes) in the model is the medical concept word in the sentence, which is connected with the corresponding characters in the main model.

Formally, an input sentence S can be denoted as a sequence as $(c_1, c_2, ..., c_n)$, in which c_j is the j-th character in S and can be represented as a D-dimensional vector X_j^c from a pretrain lookup table:

$$X_i^c = e^c(token(C_{j,len(W)}^c)) \tag{5}$$

In our experiment, MKE is integrated into a basic recurrent structure, so the final computation of X_j^c considers the

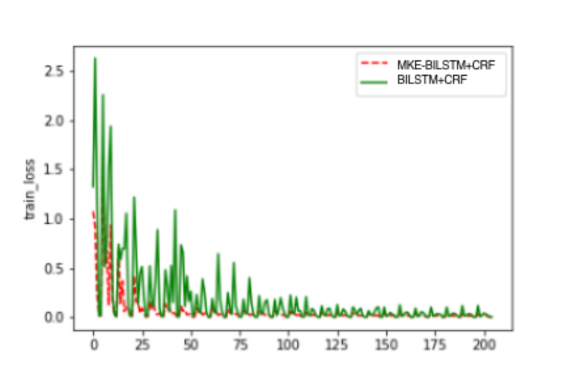

Figure 4: Comparison of named entity recognition loss functions

	P	R	F1
BILSTM+CRF	88.64%	88.63%	87.93%
Best in CCKS 2018	-	-	89.26%
MKE-BILSTM+CRF	91.22%	92.03%	91.68%

Table 4: Evaluation of three models

medical concept is represented as $X_j^{c'}$:

$$X_j^{c'} = [X_j^c, e^t(token(C_{j,m}^c))] \tag{6}$$

where $token(C_{j,len(W)}^c)$ denotes the concept word from character c_j to c_m, and $m - j + 1$ denotes the number of characters in the entity, e^t is the graph embedding lookup table that we have been trained. For every character contained in this concept word, the concept embedding is concatenated to char embedding.

4.3 Experimental Settings

The experiments run on a Tesla P40 (24 GB). Training is performed with batch size 32, dropout probability 0.1, learning rate $\eta = 1\text{e}{-4}$ and training epochs 5. The model is trained both using Adam's Optimizer in a stochastic gradient descent fashion. The medical graph neural network embedding dimension is 200.

4.4 Experimental Results

A comparison of training loss lies in Figure 4, which is loss on different iterations. The baseline is a classic character-based BILSTM+CRF model using a pre-trained character embeddings, which shows that training loss of our method is substantially less compared baseline as the training loss value of MKE-BILSTM+CRF is smoother and stabilizes faster. So the MKE enhances the model to learn quickly, which means less annotated corpus.

According to the track of CCKS, the strict evaluation metrics is employed to measure the performance as below.

The output of NER is denoted as $S = s_1, s_2, ..., s_m$ and the set of the manual annotation (Gold Standard) is denoted as $G = g_1, g_2, ..., g_m$. The element of the collection is an entity mention, which is denoted as a quaternion $< d, b, e, c >$, where d is the document, b and e correspond to the starting and ending position of the entity mention appeared in the document d, respectively. c indicates pre-defined category that the entity mention belongs to. The evaluation is defined $s_i \in S$ is strictly equivalent to $g_i \in G$, if and only if:

$$s_i.d = g_i.d, s_i.b = g_i.b, s_i.e = g_i.e, s_i.c = g_i.c \tag{7}$$

$$P_s = \frac{S \cap_s G}{|S|}, R_s = \frac{S \cap_s G}{|G|}, F1_s = \frac{2 * P_s + *}{P_s + R_s} \tag{8}$$

Evaluations are listed in Table 3 for three models. The first line is a baseline model as BILSTM+CRF, and the F1 reach to 87.93%. MKE-BILSTM+CRF with MKE as priori knowledge gives a 91.68% F1-score based on the same experimental settings and inputs, which is higher compared to the baseline, improved by 4%.

Furthermore, the best result over more than 100 teams on the track lies in the second line of the table, whose F1 is 89.26% [Zhang *et al.*, 2019]. In this model, a large amount of manually defined features are adopted in, including part-of-speech tagging, pinyin characteristics, roots, radicals and dictionary features, etc. These complex feature engineering overcomes all other complicated models, such as hybrid model and representation learning method. On the other contrary, MKE-BILSTM+CRF only includes pre-trained concept embedding based classic BILSTM+CRF to raise F1 to 91.68% with most labor free.

5 Conclusions

In this paper, a method on medical concept embedding is proposed by automatically building a large-scale concept graph from EMRs and embedding medical semantics information through graph neural network, named as medical knowledge embedding (MKE), and the whole procedure is almost manual annotation free. Because of the limitation on public medical AI resources, the paper uses a medical named entity recognition task on text from history of present illness on CCKS is carried on to demonstrate the value of MKE as prior knowledge. MKE can definitely be used in other medical tasks to enrich semantic information.

In the future, to improve MKE, the scale expansion is one aspect for covering more types of hospitals and patients from more regions. On the other way, a special MKE is perhaps a better way to improve tasks for different diseases.

6 Acknowledgments

References

[Celebi *et al.*, 2019] Remzi Celebi, Huseyin Uyar, Erkan Yasar, Ozgur Gumus, Oguz Dikenelli, and Michel Dumontier. Evaluation of knowledge graph embedding approaches for drug-drug interaction prediction in realistic settings. *BMC Bioinformatics*, Dec 2019.

[Datla *et al.*, 2017] Vivek Datla, Sadid A. Hasan, Ashequl Qadir, Kathy Lee, Yuan Ling, Joey Liu, and Oladimeji Farri. Automated clinical diagnosis: The role of content in various sections of a clinical document. In *2017*

IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pages 1004–1011, 2017.

[Gong *et al.*, 2020] Fang Gong, Meng Wang, Haofen Wang, Sen Wang, and Mengyue Liu. Smr: Medical knowledge graph embedding for safe medicine recommendation, 2020.

[Goodwin and Harabagiu, 2013] Travis Goodwin and Sanda M. Harabagiu. Automatic generation of a qualified medical knowledge graph and its usage for retrieving patient cohorts from electronic medical records. In *2013 IEEE Seventh International Conference on Semantic Computing*, pages 363–370, 2013.

[Goodwin and Harabagiu, 2016] Travis Goodwin and Sanda Harabagiu. Embedding open-domain common-sense knowledge from text. *LREC Int Conf Lang Resour Eval*, 2016:4621–4628, May 2016.

[Gori *et al.*, 2005] M. Gori, G. Monfardini, and F. Scarselli. A new model for learning in graph domains. In *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005.*, volume 2, pages 729–734 vol. 2, 2005.

[Grover and Leskovec, 2016] Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks, 2016.

[Kang *et al.*, 2009] Bo-Yeong Kang, Dae-Won Kim, and Hong-Gee Kim. Two-phase chief complaint mapping to the umls metathesaurus in korean electronic medical records. *IEEE Transactions on Information Technology in Biomedicine*, 13(1):78–86, 2009.

[Khare *et al.*, 2014] Ritu Khare, Jiao Li, and Zhiyong Lu. Labeledin: Cataloging labeled indications for human drugs. *J Biomed Inform*, 52:448–456, Dec 2014.

[Kipf and Welling, 2017] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks, 2017.

[Lample *et al.*, 2016] Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer. Neural architectures for named entity recognition, 2016.

[Li *et al.*, 2020] Linfeng Li, Peng Wang, Yao Wang, Shenghui Wang, Jun Yan, Jinpeng Jiang, Buzhou Tang, Chengliang Wang, and Yuting Liu. A method to learn embedding of a probabilistic medical knowledge graph: Algorithm development. *Journal of Medical Internet Research Med Inform*, 8(5):e17645, May 2020.

[Ling, 2017] Yuan Ling. Methods and techniques for clinical text modeling and analytics, 02 2017.

[Liu *et al.*, 2017] Liyuan Liu, Jingbo Shang, Frank F. Xu, Xiang Ren, Huan Gui, Jian Peng, and Jiawei Han. Empower sequence labeling with task-aware neural language model, 2017.

[Liu *et al.*, 2019] Wei Liu, Tongge Xu, Qinghua Xu, Jiayu Song, and Yueran Zu. An encoding strategy based word-character lstm for chinese ner. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 2379–2389. Association for Computational Linguistics, Jun 2019.

[Long *et al.*, 2018] Sun Long, Rao Yuan, Lu Yi, and Li Xue. A method of chinese named entity recognition based on cnn-bilstm-crf model. In Qinglei Zhou, Qiguang Miao, Hongzhi Wang, Wei Xie, Yan Wang, and Zeguang Lu, editors, *Data Science*, pages 161–175. Springer Singapore, 2018.

[Mao and Fung, 2020] Yuqing Mao and Kin Wah Fung. Use of word and graph embedding to measure semantic relatedness between unified medical language system concepts. *Journal of the American Medical Informatics Association*, 27(10):1538–1546, 10 2020.

[Perozzi *et al.*, 2014] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. Deepwalk: Online learning of social representations. *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug 2014.

[Scarselli *et al.*, 2009] Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. The graph neural network model. *IEEE Transactions on Neural Networks*, 20(1):61–80, 2009.

[Schechtman, 2002] Edna Schechtman. Odds ratio, relative risk, absolute risk reduction, and the number needed to treat–which of these should we use? *Value Health*, 5(5):431–436, 2002.

[Sun *et al.*, 2020] Haixia Sun, Jin Xiao, Wei Zhu, Yilong He, Sheng Zhang, Xiaowei Xu, Li Hou, Jiao Li, Yuan Ni, and Guotong Xie. Medical knowledge graph to enhance fraud, waste, and abuse detection on claim data: Model development and performance evaluation. *JMIR Med Inform*, 8(7):e17653, Jul 2020.

[Tang *et al.*, 2015] Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. Line: Large-scale information network embedding. *Proceedings of the 24th International Conference on World Wide Web*, May 2015.

[Wu *et al.*, 2019] Tong Wu, Yunlong Wang, Yue Wang, Emily Zhao, Yilian Yuan, and Zhi Yang. Representation learning of ehr data via graph-based medical entity embedding, 2019.

[Zhang and Yang, 2018] Yue Zhang and Jie Yang. Chinese ner using lattice lstm, 2018.

[Zhang *et al.*, 2019] Jiangtao Zhang, Juanzi Li, Zengtao Jiao, and Jun Yan. Overview of ccks 2018 task 1: Named entity recognition in chinese electronic medical records. In Xiaoyan Zhu, Bing Qin, Xiaodan Zhu, Ming Liu, and Longhua Qian, editors, *Knowledge Graph and Semantic Computing: Knowledge Computing and Language Understanding*, pages 158–164. Springer Singapore, 2019.

[Zhou *et al.*, 2021] Jie Zhou, Ganqu Cui, Shengding Hu, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun. Graph neural networks: A review of methods and applications, 2021.