Biomedical Network Link Prediction using Neural Network Graph Embedding

Sumit Kumar
Birla Institute of Technology
Mesra, India
sumit.atlancey@gmail.com

Raj Ratn Pranesh
Birla Institute of Technology
Mesra, India
raj.ratn18@gmail.com

Ambesh Shekhar
Birla Institute of Technology
Mesra, India
ambesh.sinha@gmail.com

ABSTRACT

In this paper, we aim at Graph embedding learning for automatic grasping of low-dimensional node representation on biomedical networks. The purpose is to use different neural Graph embedding methods for conducting analysis on 3 major biomedical link prediction tasks: drug-disease association (DDA) prediction, drug-drug interaction (DDI) classification, and protein-protein interaction (PPI) classification. We observe that graph embedding method achieve a promising result without the use of any biological features.

ACM Reference Format:
Sumit Kumar, Raj Ratn Pranesh, and Ambesh Shekhar. 2020. Biomedical Network Link Prediction using Neural Network Graph Embedding. In *Proceedings of 8th ACM IKDD CODS and 26th COMAD*. ACM, 1 page. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Graphs are used worldwide to represent biomedical nodes and edges. In the past few years, neural network models mark their success in various fields. Analysis of biomedical graphs would greatly help in predicting potential drug indications as in [3]. Graph embedding [4] automatically learn a low dimensional feature representation for every node, while saving the structural representation of graphs which is used as features for link prediction. For Link prediction task we used 3 different Neural Network-based methods LINE [4], SDNE [5], and GAE [1] each for DDA prediction, DDI prediction, and PPI prediction. We compile 3 benchmark datasets from commonly used biomedical databases for instance NDFRT DDA [1], DrugBank DDI [2], and STRING PPI [3]

2 METHODOLOGY

Link prediction is one of the crucial tasks in biomedical fields. It can be thought of as a given set of biomedical entities and their known interaction, we aim to predict other potential interactions between entities [2]. We perform DDAs prediction using the TMF technique where a DDA matrix is factorized to learn low dimension representation for drugs and diseases in the latent space.

We use OpenNE [4], an open source python package for network embedding to learn node embeddings for LINE [4] and SDNE [5]. We compiled SVD using Numpy [5] to obtain GAE[6] [1] embeddings. For the link prediction task, the respective datasets are split into the training set(80%) and test set(20%). The number of unknown links are greater than known ones. Therefore, we randomly chose broken edges as negative with an equal number of positive samples in both the training and testing stage. For every node pair, we concatenate the embeddings of two nodes as edge feature and made a Logistic Regression Binary Classifier using SKLearn package [7].

3 EXPERIMENTS AND RESULTS

For our experiments, we performed 3 compiled biomedical networks on the dataset. The Performance of the classifier is showen in Table 1. The result on the test set shows that GAE outperformed the other model for link prediction over DrugBank DDI and SPRING PPI dataset. While LINE performed better over the NDFRT DDA dataset. This shows that proposed graph embeddings methods deserve greater attention for future biomedical link prediction and analysis. In future, We want to further expand our research by conducting extensive experiments with multiple datasets and other techniques of graph embedding.

Method	**NDFRT DDA**	**DrugBank DDI**	**STRING PPI**
LINE	**96%**	90.3%	85.6%
SDNE	94%	90.5%	87.6%
GAE	80.6%	**91.7%**	**89.9%**

Table 1: Model performance score for various dataset

REFERENCES

[1] Thomas N Kipf and Max Welling. 2016. Variational graph auto-encoders. *arXiv preprint arXiv:1611.07308* (2016).
[2] Linyuan Lü and Tao Zhou. 2011. Link prediction in complex networks: A survey. *Physica A: statistical mechanics and its applications* 390, 6 (2011), 1150–1170.
[3] Maya Rotmensch, Yoni Halpern, Abdulhakim Tlimat, Steven Horng, and David Sontag. 2017. Learning a health knowledge graph from electronic medical records. *Scientific reports* 7, 1 (2017), 1–11.
[4] Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. 2015. Line: Large-scale information network embedding. In *Proceedings of the 24th international conference on world wide web*. 1067–1077.
[5] Daixin Wang, Peng Cui, and Wenwu Zhu. 2016. Structural deep network embedding. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*. 1225–1234.

[1] https://www.nlm.nih.gov/research/umls/sourcereleasedocs/current/NDFRT/
[2] https://www.drugbank.ca/
[3] https://string-db.org/
[4] https://github.com/thunlp/OpenNE
[5] http://www.numpy.org/
[6] https://github.com/tkipf/gae
[7] https://scikit-learn.org/stable/

8th ACM IKDD CODS and 26th COMAD, 2021, Online

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn