RDF Graph Embeddings for Prediction of Drug-Drug Interactions

RDF Graph Embeddings for Prediction of Drug-Drug Interactions

Remzi Çelebi*, Erkan Yaşar*, Özgür Gümüş*, Oğuz Dikenelli*
*Department of Computer Engineering, Ege University, Izmir, Turkey

Drug-drug interactions (DDIs) are a very important topic in drug discovery and public health. Drug-drug interaction (DDI) may occur when multiple drugs are co-prescribed, and these interactions have the potential to lead to patient death or drug withdrawal, a potential which makes it the object of great interest from both academia and industry. Prediction of potential drug drug interaction helps reduce unanticipated drug interactions, drug development costs and optimizes the treatments in the drug design. The underlying assumption of the similarity based DDI prediction approach is that similar drugs may interact with same drugs. There are already several similarity-based approaches, the most common of these are structural, therapeutic, phenotypic and genomic similarities [1]. The study by Zhang, Wen, et al. [2] collected a wide variety of drug data and thus predicted drug-drug interactions by integrating chemical, biological, phenotypic and network data. Gottlieb et al. [3] built a novel framework, INDI, which considered seven kinds of drug-drug similarities. The *Villar et al.* study [4] designed a novel molecular fingerprint similarity based on DDI profiles and developed a useful *in silico* model to predict new drug interactions. Cheng et al. [5] presented a HNAI framework to predict drug interactions utilizing the drug phenotypic, therapeutic, structural, and genomic similarities.

Linked Open Data (LOD) is a technique for publishing, describing, and linking data. Linked open data is a potential source of background knowledge for modeling predictive machine learning and building content-based recommender systems. LOD identifies resources with Uniform Resource Identifiers (URI) and uses the RDF (Resource Description Framework) standard, a powerful data model, to describe and exchange resources on the Web. An important LOD-based project is Bio2RDF which integrates numerous Life Sciences databases available on different websites. Bio2RDF provides a service for data integration service from various sources as a resource for scientific researchers. Bio2RDF scripts convert heterogeneously formatted data such as SQL and XML into RDF format. Bio2RDF creates a large RDF graph that interlinks data from major biological databases, including Drugbank, KEGG and SIDER.

In this work, we have applied the RDF2Vec method to extract feature vector representation of linked open drug data from a subset of Bio2RDF datasets to predict potential drug-drug interactions. RDF2Vec is a recently published

methodology that adapts the language modeling approach of Word2Vec to RDF Graph Embeddings. Word2Vec trains a neural network model to learn vector representation of words, called word embeddings. It maps each word to a vector of latent numerical values in which semantically and syntactically closer words will appear closer in the vector space. The hypothesis which underlies this approach is that closer words in word sequence are statistically more dependent. RDF2Vec applies a similar approach to RDF Graph considering the entities and relations between entities by converting the graph into a set of sequences of entities (walks or paths) and trains the same neural network model to learn vector representation of entities in the RDF graph.

We generated walks to be used as input for vector representation of drugs on RDF graph data. A similarity or relatedness score between any two drugs could be calculated by taking the cosine of those drug vector representations. We extended our previous work [1] by integrating the RDF Graph Embedding based drug similarities to train a logistic regression classifier for DDI prediction. Our preliminary results suggest that drug vector representation based similarities could enhance existing pharmacological similarity-based DDI prediction methods. The AUC value has been increased from 0.67 to 0.76 based on fivefold cross-validation with these new similarities.

References

1- Çelebi, R., Mostafapour, V., Yasar, E., Gümüs, Ö., & Dikenelli, O. "Prediction of Drug-Drug Interactions Using Pharmacological Similarities of Drugs." *Database and Expert Systems Applications (DEXA), 2015 26th International Workshop on*. IEEE, 2015.

2- Zhang, W., Chen, Y., Liu, F., Luo, F., Tian, G., & Li, X. (2017). Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data. *BMC bioinformatics*, *18*(1), 18.

3- Gottlieb A, Stein GY, Oron Y, Ruppin E, Sharan R. INDI: a computational framework for inferring drug interactions and their associated recommendations. Mol Syst Biol. 2012;8:592.

4- Vilar, S., Uriarte, E., Santana, L., Tatonetti, N. P., & Friedman, C. (2013). Detection of drug-drug interactions by modeling interaction profile fingerprints. *PloS one*, *8*(3), e58321.

5- Cheng F, Zhao Z. Machine learning-based prediction of drug-drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties. J Am Med Inform Assoc. 2014;21(e2):e278–86.