Prediction of Drug-target interactions from heterogeneous information network based on LINE embedding model

Bo-Ya Ji
Xinjiang Technical Institutes of Physics and Chemistry, Chinese Academy of Sciences, Urumqi 830011, China. University of Chinese Academy of Sciences, Beijing 100049, China

Zhu-Hong You
Xinjiang Technical Institutes of Physics and Chemistry, Chinese Academy of Sciences, Urumqi 830011, China. University of Chinese Academy of Sciences, Beijing 100049, China

Han-Jing Jiang
Xinjiang Technical Institutes of Physics and Chemistry, Chinese Academy of Sciences, Urumqi 830011, China. University of Chinese Academy of Sciences, Beijing 100049, China

Zhen-Hao Guo
Xinjiang Technical Institutes of Physics and Chemistry, Chinese Academy of Sciences, Urumqi 830011, China. University of Chinese Academy of Sciences, Beijing 100049, China

Kai Zheng
School of Computer Science and Engineering, Cen-tral South University, Changsha, 410083, China

Abstract

Background: The prediction of potential drug-protein target interactions (DTIs) not only provides a better comprehension of biological processes but also is critical for identifying new drugs. However, due to the disadvantages of costly and high time-consuming traditional experiments, only a small section of interactions between drugs and targets in the database is verified experimentally. Therefore, it is meaningful and important to develop new computational methods with good performance for predicting DTIs. At present, many existing computational methods only utilize a single type of molecule without paying attention to the interactions and influences between other types of molecules.

Methods: In this work, we developed a novel network embedding-based heterogeneous information integration model to predict potential DTIs. Firstly, a heterogeneous information network is built by combining the known associations among protein, drug, lncRNA, disease, and miRNA. Secondly, the Large-scale Information Network Embedding (LINE) model is used to learn behavior information of nodes in the network. Hence, the known drug-protein interaction pairs can be represented as a combination of attribute information (e.g. protein sequences information and drug molecular fingerprints) and behavior information of themselves. Thirdly, the Random Forest classifier is used for training and predicting.

Results: In the results, under the 5-fold cross validation, our method obtained 85.83% prediction accuracy with 80.47% sensitivity at the AUC of 92.33%. Moreover, in the case studies of three common drugs, the top 10 candidate targets have 8 (Caffeine), 7 (Clozapine) and 6 (Pioglitazone) are respectively verified to be associated with corresponding drugs.

Conclusions: In short, these results indicate that our method can be a powerful tool for predicting drug-protein interactions and finding unknown targets for certain drugs or unknown drugs for certain targets.

Full Text

However, the manuscript can be downloaded and accessed as a PDF.

Figures

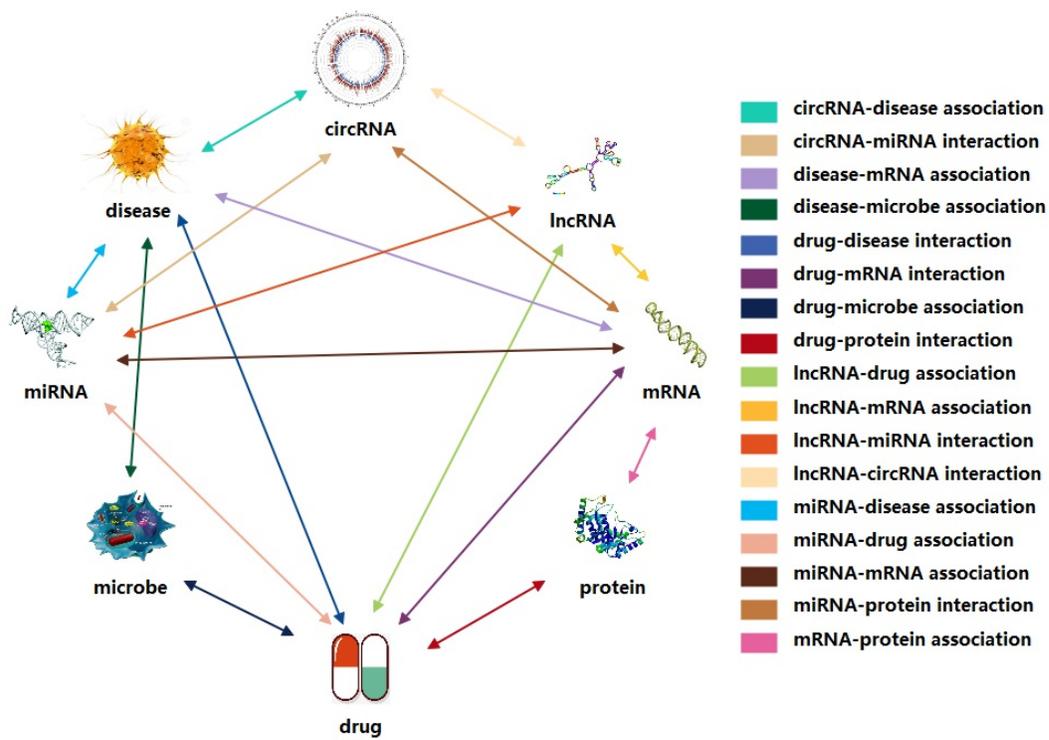



Figure 1

The heterogeneous information network.

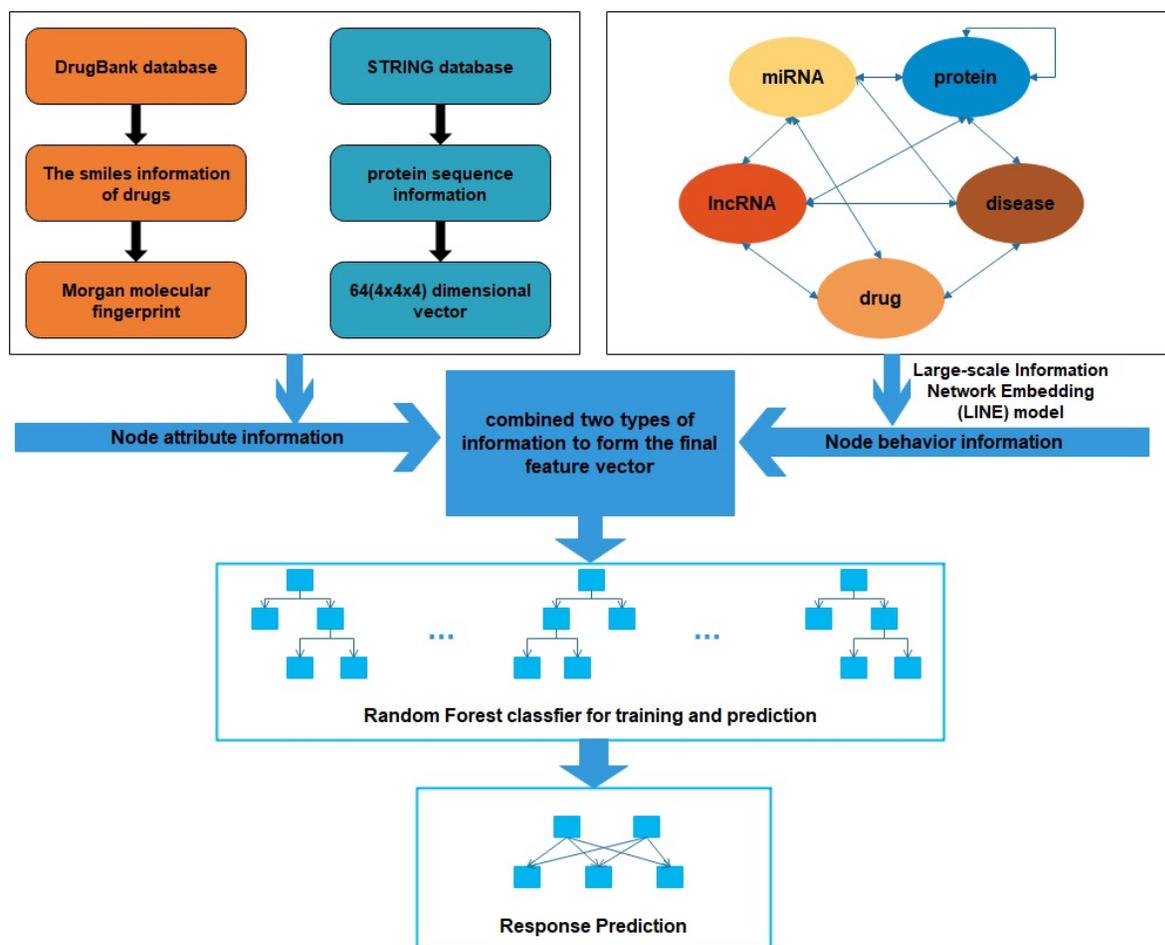



Figure 2

Flowchart of our model to predict potential drug-protein interactions

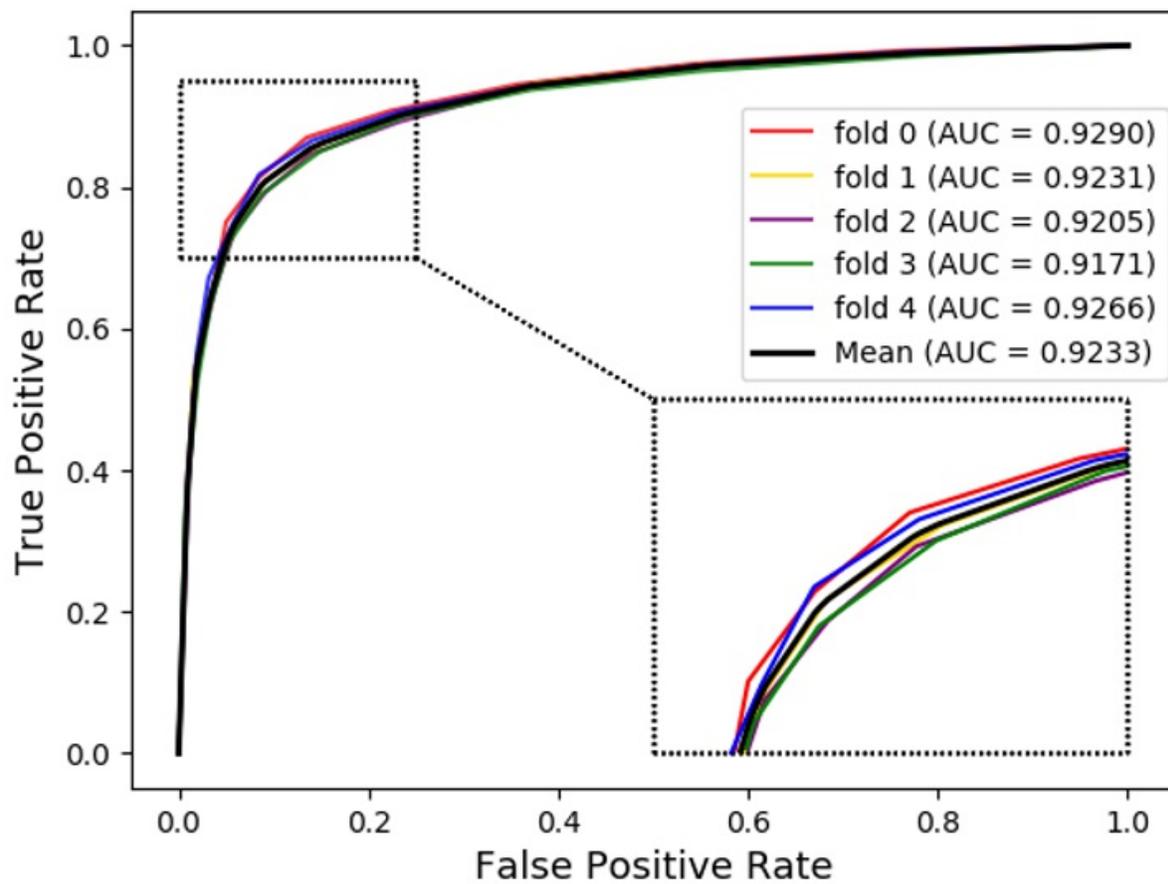



Figure 3

The ROC curves of our method in drug-target interaction prediction under 5-fold cross validation

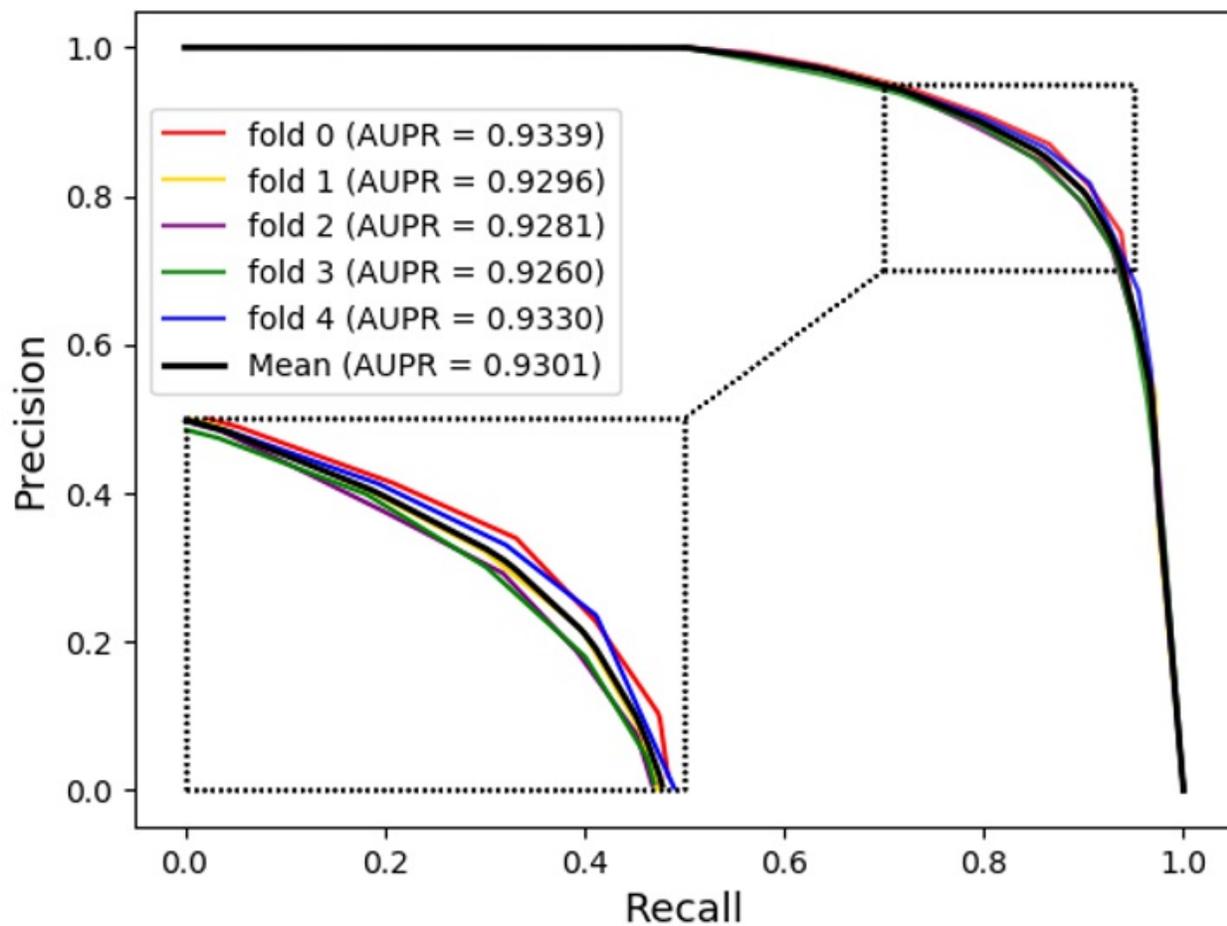



Figure 4

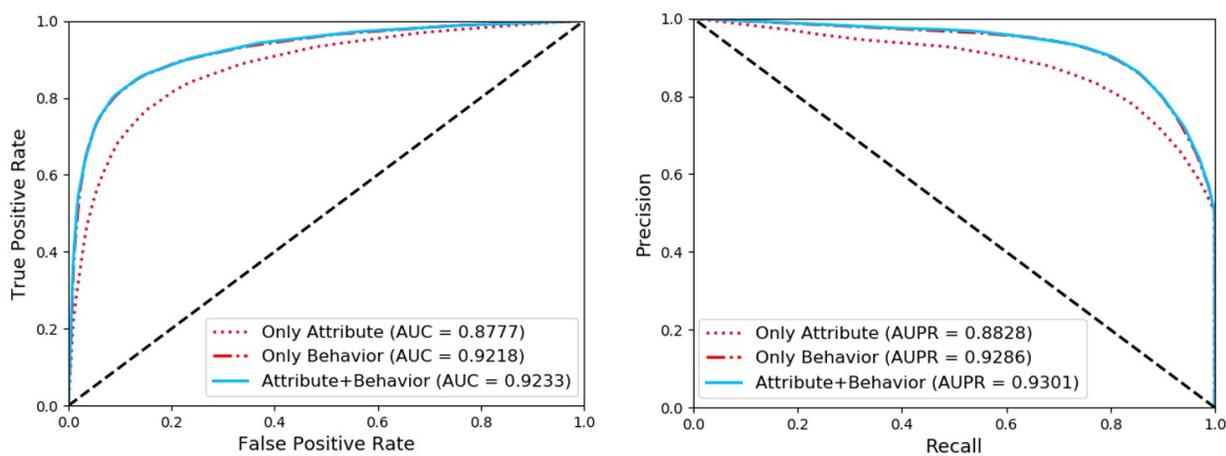
The PR curves of our method in drug-target interaction prediction under 5-fold cross

validation

Figure 5

Comparison of our method with different combinations under 5-fold cross-validation

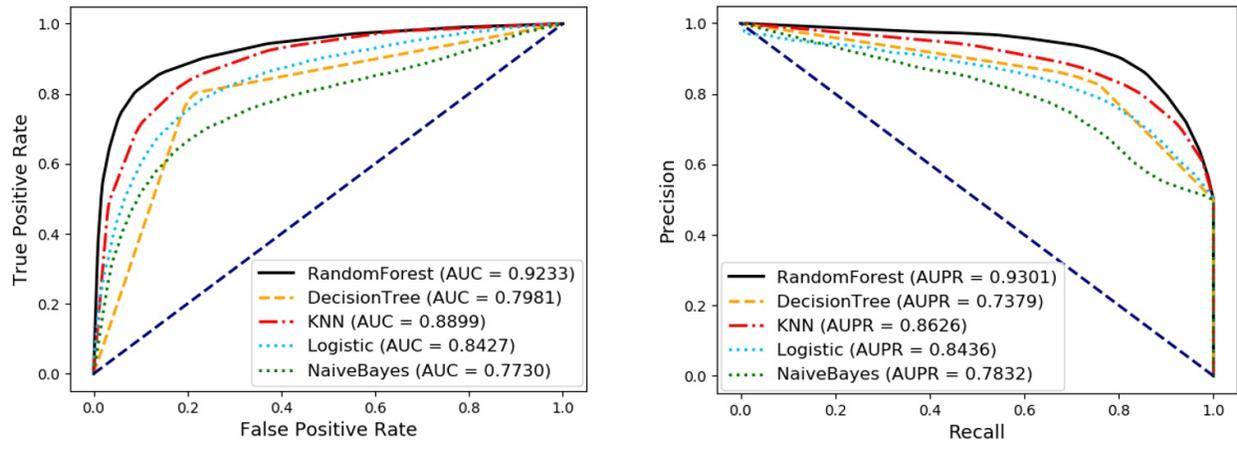



Figure 6

Comparison of our method with Random Forest, KNN, Naive Bayes, Logistic and Decision

Tree classifiers under five-fold cross-validation