Bioinformatics
doi.10.1093/bioinformatics/xxxxxx
Advance Access Publication Date: Day Month Year
Manuscript Category

OXFORD

Subject Section

Combining Bottom-up and top-down approaches for drug-target-interaction prediction

Tilman Hinnerichs [1,*] and Robert Hoehndorf [2]

[1]Department, Institution, City, Post Code, Country and
[2]Department, Institution, City, Post Code, Country.

*To whom correspondence should be addressed.

Associate Editor: XXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Motivation: Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text.
Results: Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text
Availability: Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text Text
Contact: tilman.hinnerichs@kaust.edu.sa
Supplementary information:10264703 Supplementary data are available at *Bioinformatics* online.

1 Introduction

Wang u.a. [2018]
In history, traditional remedies, that were known for their medicinal properties lead to drugs by extraction of the functional ingredients. Alternatively, characteristics and features of potential drugs were detected by accident. More recently, biological drug targets were found *in silico* through discovery of suitable computational predictors.
The challenge of accurately predicting drug-target-interactions (DTI) has shown its importance in the fields of drug repurposing and repositioning, and in the exploration of novel drugs and their interaction partners. Knowledge about those links between compounds and their target proteins help in an array of medical and pharmaceutical studies. Additionally, those associations can be utilized to identify disease specific targets, leading to desirable therapeutic effects.
With the rapidly growing field of machine learning approaches and their application to bioscientifical problems in the realm of bioinformatics, different kinds of data, such as long DNA sequences could be utilized for feature generation, while rapid advances were made. Almost all state of the art models for drug-target-interaction prediction were based on the usage of neural networks with increasing size.
Only recently, the technique of graph learning was introduced by (T. Kipf et. al) through graph convolution algorithms, and improved and altered under usage of attention mechanisms (GAT), random walks (SAGE, APPNP, PPNP) and mixtures of both (copy pytorch geometric citations). While based on diverse systems, they can be relevant for testing distinct hypothesis for given graphs. While convolutional filters are suitable for finding patterns among the the given graph, attention mechanisms are more relevant for discovery of important regions within. Lately, graph learning approaches found application for computing compound representations for DTI prediction.
Approaches on this rather sophisticated problem can divided into top-down or network approaches (**CITATION**), and bottom-up or molecular approaches (**CITATION**). Top-down approaches take advantage of other data such as diseases (CITATION), side effects, knowledge graphs or ontologies, in order to learn representations for both compound and protein. On the other hand, bottom-up aspirations attempt to learn from chemical properties. For drugs, molecular structure (CITATION GraphDTA), molecular fingerprints, similarity to other drugs (See Bioinf Survey), and other molecular features. On the protein side, secondary structure prediction (CITATION), contact prediction (CITATION), or simply convolution over the amino acid sequences served to obtain a feature for the given proteins. However, both kinds contain and share some problems, that are not solvable within themselves. Thus, bottom-up approaches share the lack of ability to generalize, which we will show in later sections, and usually focus on engineering sophisticated features for the drugs, while neglecting to formulate meaningful features on the protein side. Top-down approaches lack the ability to spot small differences to cope with small differences within the drug structure and rely heavily on given data for the considered

picture(0,0)(-35,0)(1,0)30 (0,35)(0,-1)30 picture

picture(0,0)(35,0)(-1,0)30 (0,35)(0,-1)30 picture

drug-target pair. The latter is not suitable for predictions on novel or unseen compounds, as e.g., data on side effects or its impact on diseases is seldom given for novel drugs.

In order to design such a feature for proteins and drugs, respectively, we make use of the interaction networks for both proteins and compounds. Drug-drug interaction networks were introduced by (Boyce et al. CITATION) have have been used for . . . (CITATION). Drug-drug interaction networks may give a hint on common targeted pathways or . . . (See Boyce paper what they could be used for). As an additional compound feature we will use semantic side effect similarity, which we will discuss later on.

Protein-protein interaction networks have shown great results in . . . (CITATION, copy some stuff and ideas from them).

2 Approach

3 Methods

4 Discussion

5 Conclusion

Acknowledgements

Funding

This work has been supported by the... Text Text Text Text.

References

Wang, Chen / Kurgan, Lukasz(2018): *Review and comparative assessment of similarity-based methods for prediction of drug–protein interactions in the druggable human proteome*, 6: 2066-2087.

picture(0,0)(-35,0)(1,0)30 (0,-35)(0,1)30 picture

picture(0,0)(35,0)(-1,0)30 (0,-35)(0,1)30 picture