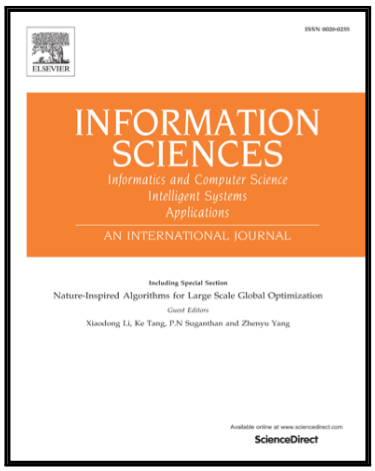

Identification of Drug-Target Interactions via Multiple Information Integration

Yijie Ding, Jijun Tang, Fei Guo

PII: S0020-0255(17)30777-6
DOI: 10.1016/j.ins.2017.08.045
Reference: INS 13049

To appear in: *Information Sciences*

Received date: 5 June 2017
Revised date: 9 August 2017
Accepted date: 11 August 2017

Please cite this article as: Yijie Ding, Jijun Tang, Fei Guo, Identification of Drug-Target Interactions via Multiple Information Integration, *Information Sciences* (2017), doi: 10.1016/j.ins.2017.08.045

Identification of Drug-Target Interactions via Multiple Information Integration

Yijie Ding[a,b], Jijun Tang[a,b,c], Fei Guo[a,b,*]

[a] *School of Computer Science and Technology, Tianjin University, Tianjin 300350, China*
[b] *Tianjin University Institute of Computational Biology, Tianjin 300350, China*
[c] *Department of Computer Science and Engineering, University of South Carolina, Columbia, SC 29208, USA*

Abstract

Identifying Drug-Target Interactions (DTIs) is an important process in drug discovery. Traditional experimental methods are expensive and time-consuming for detecting DTIs. Therefore, computational approaches provide many effective strategies to deal with this issue. In recent years, most of computational methods only use the information of drug-drug similarity or target-target similarity, which cannot perfectly capture all characteristics to identify DTIs. In this paper, we propose a novel computational model of DTIs prediction, based on machine learning methods. To improve the performance of prediction, we further use molecular substructure fingerprints, Multivariate Mutual Information (MMI) of proteins and network topology to represent drugs, targets and relationship between them. Moreover, we employ Support Vector Machine (SVM) and Feature Selection (FS) to construct model for predicting DTIs. Experiments of evaluation show that proposed approach achieves better results than other outstanding methods for feature-based DTIs prediction. The proposed approach achieves AUPRs of 0.899, 0.929, 0.821 and 0.655 on Enzyme, Ion Channel (IC), GPCR and Nuclear Receptor datasets, respectively. Compared with existing best methods, AUPRs are increased by 0.016 on Ion Channel datasets. In addition, our method obtains the second best performance on

☆

*Corresponding author
Email address: fguo@tju.edu.cn (Fei Guo)