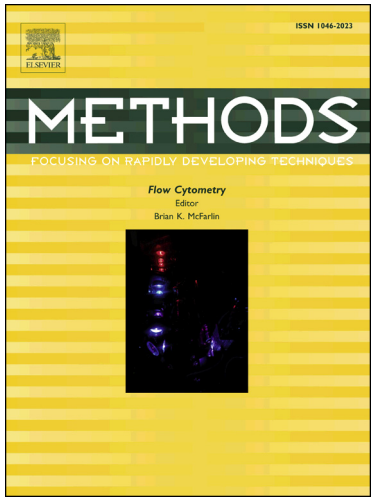

Drug-Target Interaction Prediction using Ensemble Learning and Dimensionality Reduction

Ali Ezzat, Min Wu, Xiao-Li Li, Chee-Keong Kwoh

PII: S1046-2023(17)30025-7
DOI: http://dx.doi.org/10.1016/j.ymeth.2017.05.016
Reference: YMETH 4216

To appear in: *Methods*

Please cite this article as: A. Ezzat, M. Wu, X-L. Li, C-K. Kwoh, Drug-Target Interaction Prediction using Ensemble Learning and Dimensionality Reduction, *Methods* (2017), doi: http://dx.doi.org/10.1016/j.ymeth.2017.05.016

Drug-Target Interaction Prediction using Ensemble Learning and Dimensionality Reduction

Ali Ezzat[a], Min Wu[b], Xiao-Li Li[b,*], Chee-Keong Kwoh[a]

[a]*School of Computer Science and Engineering, Nanyang Technological University, Singapore*
[b]*Institute for Infocomm Research (I^2R), A*Star, Singapore*

Abstract

Experimental prediction of drug-target interactions is expensive, time-consuming and tedious. Fortunately, computational methods help narrow down the search space for interaction candidates to be further examined via wet-lab techniques. Nowadays, the number of attributes/features for drugs and targets, as well as the amount of their interactions, are increasing, making these computational methods inefficient or occasionally prohibitive. This motivates us to derive a reduced feature set for prediction. In addition, since ensemble learning techniques are widely used to improve the classification performance, it is also worthwhile to design an ensemble learning framework to enhance the performance for drug-target interaction prediction.

In this paper, we propose a framework for drug-target interaction prediction leveraging both feature dimensionality reduction and ensemble learning. First, we conducted feature subspacing to inject diversity into the classifier ensemble. Second, we applied three different dimensionality reduction methods to the subspaced features. Third, we trained homogeneous base learners with the reduced features and then aggregated their scores to derive the final predictions. For base learners, we selected two classifiers, namely *Decision Tree* and *Kernel Ridge Regression*, resulting in two variants of ensemble models, *EnsemDT* and *EnsemKRR*, respectively.

*Corresponding author
Email address: xlli@i2r.a-star.edu.sg (Xiao-Li Li)