Supplementary materials for paper

Drug-Target Interaction Prediction: a Bayesian Ranking Approach

Ladislav Peska[1,2*], Krisztian Buza[2,3]

[1]Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic,

[2]Brain Imaging Centre, Hungarian Academy of Sciences, Budapest, Hungary,

[3]Rheinische Friedrich-Wilhelms-Universität Bonn, Germany.

* Corresponding author.

1 Software, Datasets and Raw Results

Source codes of our approach, as well as datasets used in the evaluation and raw results can be obtained at:
http://www.ksi.mff.cuni.cz/~peska/BRDTI

2 Hyperparameters Tuning and Evaluation

The methods' hyperparameters were tuned via grid-search and 5-fold cross-validation. In BLM-NII, the max function was used to generate final prediction and a weight for combination of structural and collaborative similarity, α, was chosen from {0.0, 0.1, ⋯, 1.0}. In WNN-GIP, the decay hyperparameter T was chosen from {0.1, 0.2, ⋯, 1.0} and a weight for combination of structural and collaborative similarity, α, was chosen from {0.0, 0.1, ⋯, 1.0}. In NetLapRLS, we kept the β and γ parameters the same for both drugs and targets and chosen from $\{10^{-6}, 10^{-5}, \dots, 10^{2}\}$. In matrix factorization approaches, the maximal number of iterations was held constant at 100 and the number of latent factors f was chosen from {50, 100}. Furthermore, in CMF, learning rate λ_l was chosen from $\{2^{-2}, 2^{-1}, 2^{0}\}$ and drug and target regularizations λ_d, λ_t were chosen from $\{2^{-5}, 2^{-4}, 2^{-3}, 2^{-2}\}$. In BRDTI the general regularization $\lambda_g$was chosen from {0.01, 0.05, 0.1, 0.3} and the content alignment regularization λ_c was kept the same for both drugs and targets and chosen from {0.05, 0.1, 0.5, 0.9}. The learning rate was initially set to 0.1 and then updated according to the bold driver heuristics. The volume of nearest neighbors k was held constant at 5. Significance of results was evaluated by one-sided paired t-test. We use the PyDTI implementation of BLMNII, WNN-GIP, NetLapRLS and CMF methods.[1]

3 Results of the Combined Approach

In order to further evaluate contribution of BRDTI method, we implemented a simple combined approach comprising of the all evaluated methods. For each predicted DTI, the combined approach provide mean value of the all single methods prediction. Further combination methods (max, min and probabilistic sum: $S(a,b) = a + b - a * b$) were also tested, however their results were consistently inferior and thus we do not report them. We evaluated combined approach under two conditions, trying to answer following questions:

1) Could combination of methods improve over the results of the single method?

[1] https://github.com/stephenliu0423/PyDTI

2) Considering combined approach, could the incorporation of BRDTI improve its results?

Table S1 contains results w.r.t. both questions. The results indicate that *combined* approach could improve over the results of best single method in the most cases. Furthermore, incorporation of BRDTI method in the *combined* approach improved the results w.r.t. per-drug nDCG significantly over the combined approach without BRDTI. Thus we can conclude that BRDTI cannot be replaced via simple combination of other prediction methods.

Table S1. Comparison of *combined* approach for DTI prediction with and without incorporating BRDTI results for each dataset and CV setting. In each column, the values represent *per-drug nDCG*. The best results and the results without a significant difference to it ($p<0.05$) are in bold.

per-drug nDCG	Best Single	Combined w/o BRDTI	Combined with BRDTI
GPCR	(BRDTI) 0.929	0.928	**0.936**
Ion Channels	(CMF) **0.954**	0.949	**0.956**
Nuclear Receptors	(BRDTI) **0.948**	0.944	**0.952**
Enzymes	(BRDTI) 0.897	0.898	**0.905**