

MEETING ABSTRACT

Open Access

Integrative biclustering of heterogeneous datasets using a Bayesian nonparametric model with application to chemogenomics

Dazhuo Li, Eric C Rouchka*

From 10th Annual UT-ORNL-KBRIN Bioinformatics Summit 2011
Memphis, TN, USA. 1-3 April 2011

Motivation

The identification of protein function and the prediction of ligand-target interaction is an active research field that is facilitated by means of categorizing ligands and proteins into biologically sensible groups. Because of the pharmacological fact that related drugs can bind to receptors without obvious sequence or structural similarity, it is appropriate to categorize proteins based not only on their sequence or structures but also on the chemical structure and the phenotypic side-effect of their ligands. In chemogenomic studies where the complete set of ligands for a protein is not known *a priori*, integrating the *de novo* detection of interacting ligand and protein groups into the categorization process can guide the process towards more biologically sensible solutions.

Results

We present the Weighted Infinite Relational Model (WIRM) that jointly detects biologically sensible ligand groups and protein groups by integrating the clustering of various data types including chemical compound descriptors, protein sequences, ligand-target bindings and pharmaceutical effects. WIRM takes advantage of the Bayesian nonparametric paradigm for integrating multiple data types, for allowing for missing values (e.g. unknown ligand-target interaction) in the data, for automatically inferring the number of clusters without explicit model comparison, and for predicting the ligand-target interactions. Because some of these data types, to varying degrees, may suggest relationships having no

implication for ligand-target interactions or for biological sensible ligand and protein groups, WIRM allows different types of data to have different weights based on prior knowledge of their quality or relevance.

Conclusion

We apply WIRM to the ion channel proteins and G-protein-coupled receptors. We validate its performance using functional annotation and ligand-target interaction. We also test the relationship among multiple data types by varying the weights which indicate the impact of each data type on the model. The categories and interactions inferred by WIRM both confirm known biology and suggest novel predictions.

Acknowledgements

This work was supported in part by the National Institutes of Health (P2ORR16481, P2ORR16481S1, P30ES014443) and the Department of Energy (DE-EM0000197). Its contents are solely the responsibility of the authors and do not represent the official views of NCTR, NIEHS, NIH, or DOE.

Published: 5 August 2011

doi:10.1186/1471-2105-12-S7-A6

Cite this article as: Li and Rouchka: Integrative biclustering of heterogeneous datasets using a Bayesian nonparametric model with application to chemogenomics. *BMC Bioinformatics* 2011 **12**(Suppl 7):A6.

* Correspondence: eric.rouchka@louisville.edu
Department of Computer Engineering and Computer Science, University of Louisville, Louisville, KY, 40292, USA