

A web tool for measuring gene semantic similarities by combining Gene Ontology and gene co-function networks

Jiajie Peng^{*}
School of Computer Science
and Technology, Harbin
Institute of Technology, Harbin,
China
MSU-DOE Plant Research
Laboratory, Michigan State
University, East Lansing 48824
jiajiepeng@hit.edu.cn

Hongxiang Li^{*}
School of Computer Science
and Technology, Harbin
Institute of Technology, Harbin,
China
gardeniaxy@hit.edu.cn

Yadong Wang[†]
School of Computer Science
and Technology, Harbin
Institute of Technology, Harbin,
China
ydwang@hit.edu.cn

Jin Chen[†]
MSU-DOE Plant Research
Laboratory, Department of
Computer Science &
Engineering, Michigan State
University, East Lansing 48824
jinchen@msu.edu

ABSTRACT

Gene Ontology (GO) is one of the most popular bioinformatic resource to study functional relationships between genes. However, the existing tools for calculating the semantic similarity are limited since they only rely on GO annotations and structure. We provide a working demonstration of a web tool that calculates functional similarities between GO terms by integrating gene co-function networks and information from GO. We demonstrate the function of the web tool by comparing remote GO terms.

Categories and Subject Descriptors

J.3 [Computer Applications]: LIFE AND MEDICAL SCIENCES

General Terms

Measurement

Keywords

Gene Ontology, Semantic similarity, Software tools

1. INTRODUCTION

The Gene Ontology (GO) is one of the most popular vocabulary system for representing biological knowledge and

describing function information for genes and gene products. As an integrated resource, GO has been widely used to study functional relationships between genes [2]. Despite the fact that various methods have been proposed to measure gene functional similarities [10, 9, 7], our understanding to the semantic relationships between genes or GO terms is still limited. For example, the process ‘response to water deprivation’ (GO:0009414) is intimately related to ‘photosynthesis’ (GO:0015979) and ‘root development’ (GO:0048364) in plants, but none of these terms are closely linked to each other in GO because such tight relationships only exist in plants [8]. Subsequently, the methods purely based on GO annotation and structure may inappropriately measure the functional similarities of the related genes in plants.

In order to accurately measure the functional similarities between genes, we recently developed a new method called NETSIM to measure GO based gene functional similarities by incorporating co-function networks of genes as additional biological knowledge [8]. Specifically, NETSIM combines information from gene co-function networks and the GO structure and annotations in three steps. First, given two GO terms, the functional distance between two corresponding annotated gene sets is calculated using a user provided gene co-function network. The second step identifies the genes annotated to the lowest common parent term of the two given terms by only propagating the genes along the paths from the two given GO terms to the lowest common parent term. Third, similarity between the two given GO terms is calculated based on the functional distance computed in step one and the path-constrained GO annotations in step two. The experimental results on metabolic reaction maps of yeast, Arabidopsis and human indicate that NETSIM can improve the performance of GO term similarities, especially for the genomes with sparse gene annotation data.

^{*}Co-first author

[†]Corresponding author.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage, and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s). Copyright is held by the author/owner(s).

BCB’15, September 9–12, 2015, Atlanta, GA, USA.

ACM 978-1-4503-3853-0/15/09.

<http://dx.doi.org/10.1145/2808719.2816835>.

Various web tools have already been developed to calculate gene functional similarities using GO, such as GossToWeb [1] and G-SESAM [3]. However, none of these tools can incorporate extra information from biological networks to complement the limitation of knowledge encoded in GO. In this demonstration, we present *NETSIMWeb*, an easy-to-use web tool that allows researchers to use NETSIM to measure gene functional similarities. Comparing with the existing web tools, the advantage of our work are:

- *NETSIMWeb* supplies researchers an gene functional similarity measurement that consider both gene co-functional network and information from GO.
- *NETSIMWeb* is an convenient HTML5 based web interface to measure gene functional similarities. Using new features supported by HTML5, such as canvas, It is convenient to integrate graphical content into NETSIMWeb without using third party plugins.

2. DEMONSTRATION

In this demonstration, we provide users the capability to calculate gene functional similarities in three well studied model organisms, i.e. yeast, Arabidopsis and human. The GO annotations of these organisms have been downloaded from the GO website. We provide YeastNET [6], AraNET [4] and HumanNET [5] as co-function networks, respectively. Users can upload a list of genes instead of two to compare them pair-wisely. The submitted job will be managed on the back-end server by a job scheduler. A user will be notified to view the final results when his/her job is done through either the submission webpage or email if user's email is provided.

2.1 Implementation

NETSIMWeb is a Browser/Server (BS) architecture-based web application. In the front end, Asynchronous JavaScript and XML (AJAX) and JavaScript Object Notation (JSON) are used for efficient data transmission between the browser and server. The back end is implemented using Java, Python 2.7 and the web develop framework web.py. MySQL is used for data management. The framework of NETSIMWeb is shown in Figure 1. Based on the GO and gene co-function networks from database, the Java package will calculate the semantic similarities for all the genes provided by user. Then the results will be visualized in user's browser. NETSIMWeb is available at <http://mlg.hit.edu.cn:8090/>.

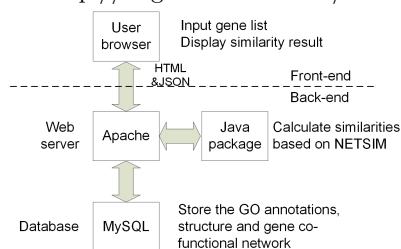


Figure 1: The framework of NETSIMWeb.

2.2 Application: discovering relationships between remote GO terms

To test whether NETSIM could reveal new knowledge comparing with existing measures, we computed the correlation coefficient between the GO term distances in GO structure and their similarity scores computed using NETSIM (we chose Arabidopsis as a running example). We also compared the results with the Wang measure, one of the most

popular semantic similarity measures based on GO [11]. The result shows that the Wang measure has strong correlation with the GO term distances (Figure 2a). However, there is no clear correlation between the NETSIM similarities and GO term distances, especially when the GO term distance is greater than seven (Figure 2b). It thus indicates that NETSIM is capable of discovering close relationships between remote GO terms using extra knowledge from gene co-functional networks.

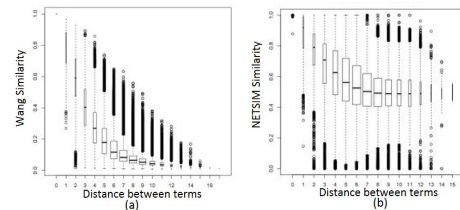


Figure 2: The correlation between the semantic similarities computed using Wang measure (a) and NETSIM (b) and the GO terms distances in GO structure.

3. REFERENCES

- [1] H. Caniza et al. Gossto: a user-friendly stand-alone and web tool for calculating semantic similarities on the gene ontology. *Bioinformatics*, page btu144, 2014.
- [2] G. O. Consortium. Gene ontology consortium: going forward. *Nucleic acids res*, 43(D1):D1049–D1056, 2015.
- [3] Z. Du et al. G-sesame: web tools for go-term-based gene similarity analysis and knowledge discovery. *Nucleic acids res*, pages W345–W349, 2009.
- [4] I. Lee, B. Ambaru, P. Thakkar, E. M. Marcotte, and S. Y. Rhee. Rational association of genes with traits using a genome-scale gene network for arabidopsis thaliana. *Nature biotechnology*, 28(2):149–156, 2010.
- [5] I. Lee, U. M. Blom, P. I. Wang, J. E. Shim, and E. M. Marcotte. Prioritizing candidate disease genes by network-based boosting of genome-wide association data. *Genome research*, 21(7):1109–1121, 2011.
- [6] I. Lee, Z. Li, and E. M. Marcotte. An improved, bias-reduced probabilistic functional gene network of baker's yeast, *saccharomyces cerevisiae*. *PLoS one*, 2(10):e988, 2007.
- [7] J. Peng, H. Li, Q. Jiang, Y. Wang, and J. Chen. An integrative approach for measuring semantic similarities using gene ontology. *BMC syst biol*, 8(Sup 5):S8, 2014.
- [8] J. Peng, S. Uygur, T. Kim, Y. Wang, S. Y. Rhee, and J. Chen. Measuring semantic similarities by combining gene ontology annotations and gene co-function networks. *BMC bioinformatics*, 16(1):44, 2015.
- [9] J. Peng, Y. Wang, and J. Chen. Towards integrative gene functional similarity measurement. *BMC bioinformatics*, 15(Sup 2):S5, 2014.
- [10] C. Pesquita, D. Faria, A. Falcao, P. Lord, and F. Couto. Semantic similarity in biomedical ontologies. *PLoS comput biol*, 5(7):e1000443, 2009.
- [11] Z. Wang, Z. Du, R. Payattakool, Y. Philip, and F. Chen. A new method to measure the semantic similarity of go terms. *Bioinformatics*, 23:1274–1281, 2007.