A new method to measure the semantic similarity of GO terms

James Z. Wang 1,*, Zhidian Du 1, Rapeeporn Payattakool 1, Philip S. Yu 2 and Chin-Fu Chen 3

1School of Computing, Clemson University, Clemson, SC 29634, USA, 2IBM T. J. Watson Research Center, 19 Skyline Drive, Hawthorne, NY 10532, USA and 3Department of Genetics and Biochemistry, Clemson University, Clemson, SC 29634, USA

*To whom correspondence should be addressed.

Abstract

Motivation: Although controlled biochemical or biological vocabularies, such as Gene Ontology (GO) (http://www.geneontology.org), address the need for consistent descriptions of genes in different data sources, there is still no effective method to determine the functional similarities of genes based on gene annotation information from heterogeneous data sources.

Results: To address this critical need, we proposed a novel method to encode a GO term's semantics (biological meanings) into a numeric value by aggregating the semantic contributions of their ancestor terms (including this specific term) in the GO graph and, in turn, designed an algorithm to measure the semantic similarity of GO terms. Based on the semantic similarities of GO terms used for gene annotation, we designed a new algorithm to measure the functional similarity of genes. The results of using our algorithm to measure the functional similarities of genes in pathways retrieved from the saccharomyces genome database (SGD), and the outcomes of clustering these genes
based on the similarity values obtained by our algorithm are shown to be consistent with human perspectives. Furthermore, we developed a set of online tools for gene similarity measurement and knowledge discovery.

**Availability:** The online tools are available at: [http://bioinformatics.clemson.edu/G-SESAME](http://bioinformatics.clemson.edu/G-SESAME)

**Contact:** jzwang@cs.clemson.edu

**Supplementary information:** [http://bioinformatics.clemson.edu/Publication/Supplement/gsp.htm](http://bioinformatics.clemson.edu/Publication/Supplement/gsp.htm)