

Applications of semantic similarity measures

Andreas Schlicker, Fidel Ramírez, Jörg Rahnenführer, Carola Huthmacher,
Alejandro Pironti, Francisco S. Domingues, Thomas Lengauer, and Mario Albrecht
Computational Biology and Applied Algorithmics
Max Planck Institute for Informatics
Campus El 4, 66123 Saarbrücken, Germany

There has been much interest in uncovering protein-protein interactions and their underlying domain-domain interactions. Many experimental techniques have been developed, for example yeast-two-hybrid screening and tandem affinity purification. Since it is time consuming and expensive to perform exhaustive experimental screens, *in silico* methods are used for predicting interactions. However, all experimental and computational methods have considerable false positive and false negative rates. Therefore, it is necessary to validate experimentally determined and predicted interactions.

One possibility for the validation of interactions is the comparison of the functions of the proteins or domains. Gene Ontology (GO) is widely accepted as a standard vocabulary for functional terms, and is used for annotating proteins and protein families with biological processes and their molecular functions. This annotation can be used for a functional comparison of interacting proteins or domains using semantic similarity measures.

Another application of semantic similarity measures is the prioritization of disease genes. It is known that functionally similar proteins are often involved in the same or similar diseases. Therefore, functional similarity is used for predicting disease associations of proteins.

In the first part of my talk, I will introduce some semantic and functional similarity measures that can be used for comparison of GO terms and proteins or protein families. Then, I will show their application for determining a confidence threshold for domain-domain interaction predictions. Additionally, I will present FunSimMat (<http://www.funsimmat.de/>), a comprehensive resource of functional similarity values available on the web. In the last part, I will introduce the problem of comparing diseases, and a first attempt to apply functional similarity measures based on GO to this problem.