A SEMANTIC SIMILARITY BASED METHOD FOR DETECTING FUNCTIONAL MODULES USING ANNOTATION PROFILES

Cheng-Wei Cheng¹, I-Jen Chiang¹² and Jau-Min Wong¹

¹Institute of Biomedical Engineering, National Taiwan University, Taipei 106, Taiwan, R.O.C.
²Institute of Medical Informatics, Taipei Medical University, Taipei 110, Taiwan, R.O.C.

D94548016@ntu.edu.tw

The functional module is defined as a group of genes that participate in one specific biological function; most genes from the module would be annotated with this function by the Gene Ontology (GO) terms. However, different genes may not completely be annotated by the identical terms but the related ones; it was apparent that different terms probably have similar semantic meanings. For this reason, we proposed a novel method to retrieve the modules taking advantage of GO annotation and the characteristic of the semantic similarity between terms.

Our approach contains three steps. First, we extract one set of genes from literatures and collect their annotated GO terms to build a preliminary annotation profile which is represented by a matrix of numeric values standing for the presence of the terms. Next, we calculate the semantic similarity values between terms from the profile and integrate those values to assemble a complete one. Lastly, we use this profile to identify the functional modules by means of hierarchical clustering method.

Eventually, we tested our method on a set of genes related with MAPK pathway and detected functional modules including transcription cofactor and protamine kinase. Furthermore, we found the property of annotation amplification of this approach could successfully separate two different biological processes which are MAPK and TGF pathway, respectively. As a result, the semantic similarity between annotated terms not only improves the data completeness but also enriches the functional signal within the module.