

A Complex Bio-networks of the Function Profile of Genes

蔣以仁

Charles C. H. Liu;I-Jen Chiang;Jau-Min Wong³;Ginni Hsiang-Chun Ts;T. Y. Lin

Abstract

This paper presents a novel model of concept representation using a multilevel geometric structure, which is called Latent Semantic Networks. Given a set of documents, the associations among frequently co-occurring terms in any of the documents define naturally a geometric complex, which can then be decomposed into connected components at various levels.

This hierarchical model of knowledge representation was validated in the functional profiling of genes. Our approach excelled the traditional approach of vector-based document clustering by the geometrical forms of frequent itemsets generated by the association rules. The biological profiling of genes were a complex of concepts, which could be decomposed into primitive concepts, based on which the relevant literature could be clustered in adequate "resolution" of contexts. The hierarchical representation could be validated with tree-based biomedical ontological frameworks, which had been applied for years, and been recently enriched by the online availability of Unified Medical Language System (UMLS) and Gene Ontology (GO).

Demonstration of the model and the clustering would be performed on the relevant GeneRIF (References into Function) document set of NOD2 gene. Our geometrical model is suitable for representation of bio-logical information, where hierarchical concepts in different complexity could be explored interactively according to the context of application and the various needs of the researchers. An online clustering search engine for use on general purpose and for biomedical use, managing the search results from Google or from PubMed, are constructed based on the methodology (<http://ginni.bme.ntu.edu.tw>). The hierarchical presentation of clustering results and the interactive graphical display of the contents of each cluster shows the merits of our approach.