Poster presentation

Similarity-based virtual screening using bayesian inference network

A Abdo* and N Salim

Address: Universiti Teknologi Malaysia, FSKSM, D07, 81310 UTM Skudai, Malysia

* Corresponding author

This abstract is available from: http://www.journal.chemistrycentral.com/content/3/S1/P44

Many methods have been developed to capture the biological similarity between two compounds for used in drug discovery. A variety of similarity metrics have been introduced, the Tanimoto coefficient being the most prominent [1][2]. Recent research in information retrieval has proved that retrieval models based on Bayesian inference networks give significant improvements in retrieval performance compare to conventional models [3][4].

One of the disadvantages in conventional 2D similarity searching is that molecular features or descriptors that are not related to the biological activity carry the same weights as the important ones. To overcome this limitation, we introducing a novel an inference network model for chemical similarity searching where the features carry different statistical weights. Features that are statistically less relevant are being deprioritized. In this study, we look at similarity searching problem using inference or evidential reasoning and decision making under uncertainty.

The network model consists of two component networks: a compound network and a query network. The compound network characterises the compounds in the database that is to be searched. The compound network is built once for a given database and its structure does not change during query processing. The query network consists of a single node which represents the user's activity requirement and one or more query node representations. A query network is built for each activity required and is modified during query processing as the query is refined or added more queries in an attempt to better characterise the activity requirement. Similarity searching is then carried out by combining the two networks and then propagates the information toward the node represent the activity required. This process of propagation is known as inference.

An important characteristic of the network model is that permits the encoding of different types of fingerprint, similarity coefficient and queries. Our experiments demonstrate that similarity approach based on network model is outperform the Tanimoto similarity search with reasonable improvement and offer a promising alternative to existing similarity search approaches. In addition, Bayesian inference network method is more efficient than Tanimoto similarity method.

References

1. Willett P, Barnard JM, Downs GM: **Chemical Similarity Searching.** *J Chem Inf Comput Sci* 1998, **38:**983.
2. Salim N, Holliday J, Willett P: **Combination of Fingerprint-Based Similarity Coefficients Using Data Fusion.** *J Chem Inf Comput Sci* 2003, **43:**435.
3. Berthier ANR, Richard M: *Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval ACM, Zurich, Switzerland*; 1996.
4. Howard RT, Croft WB: **Evaluation of an inference network-based retrieval model.** *ACM Trans Inf Syst* 1991, **9:**187.