

Summary of Discussion for
Content-rich biological network constructed by mining PubMed abstracts
by Hao Chen and Burt M. Sharp
BMC Bioinformatics 5(147) October 2004

INLS 279: Bioinformatics Research Review
Presented by Christopher Maier
2005-02-08

This paper described a text-mining system, CHILIBOT, that can mine interaction networks from PubMed abstracts. A user's query (in the form of several proteins of interest) is processed by the system and the result is a graph that shows any literature-supported relationships among the query terms in an easy-to-understand fashion.

We found several advantages to the CHILIBOT system. The main appeal of the system is its rich graphical output. By representing interactions as a directed graph it is easy to grasp complex networks. Encoded in the graph are several important data, including the direction of the interaction, the strength of the interaction, the character (inhibitory, stimulatory, neutral, etc.) of the interaction, and the support in the mined literature for the interaction. The graph conveys all this information in a concise and accessible way.

Another strength is the synonym database that backs the system. By expanding queries to include known synonyms of query terms, the coverage of the search is increased. We felt it was very good to present this synonym list back to the user *before* executing the query. In this way, the user can manually prune the synonym list and hopefully arrive at an optimal query.

The ability to easily see the sentences from the abstracts that support a given interaction was also a big plus. This information should be easily accessible to the end-user, and in CHILIBOT it is.

There were some questions we had, though. In constructing a query, CHILIBOT allows users to input a fold increase / decrease for each protein in the query. How is this ratio calculated? The article gives no mention.

Further elaboration of the crafting of the relationship rules used by the system would have been helpful. The authors give an example of classifying a relationship if two nouns do not have a verb between them, but do have a particular 'signal' word -- one signal word is "bind" which is a verb! They talk of determining the directionality of a relationship by determining which of the active or passive voices is present, but do not say how this is determined. Presumably by verb form, but it is not clear.

We would have liked to see further discussion regarding the extent that human curation plays in the system. As one of the most important components in the system, the curation of the synonym database has a large effect on system performance. In terms of system performance, the decision to replace all non-alphanumeric characters with spaces could adversely affect the system, as punctuation and Greek letters are important orthographic features for protein names.

The essential neglect of negation was one of the more troubling aspects of the system. By ignoring negation in determining directionality of interactions, the final output is suspect. How can confident can we be that a detected relationship actually possesses the proposed directionality if negation is not incorporated into the classification?

On the whole, we felt that the CHILIBOT system is a promising development in the NLP bioinformatics field. The authors also get points for making their experimental data available online.