New tool could unpick complex cancer causes and help sociologists mine Facebook

December 15 2008

Researchers at the University of Warwick's Department of Statistics and Centre for Complexity Science have devised a new research tool that could help unpick the complex cell interactions that lead to cancer and also allow social scientists to mine social networking sites such as Facebook for useful insights.

An approach called "graphical models" can be used by researchers to gain an understanding of a range of systems with multiple interacting factors. These models use mathematical objects called graphs to describe and depict the probability of relationships between each of the components. When used to study molecular biology researchers may be interested in saying something about which molecules influence one another; in the social sciences researchers would use them to understand the relationships between various economic and demographic factors.

However gaining such information from a graphical model can be a very challenging exercise, because of the vast range of possible graphs needed for even a relatively small number of variables. For instance the relatively small network studied by the University of Warwick led team for this research paper had just 14 proteins which were implicated in the development of a form of cancer, but those 14 proteins had a vast number of combinations of possible mutual interactions.

Such tasks would be made much easier if the mathematical tools used to undertake the analysis could somehow embody all the current knowledge of what was likely, and or probable, in the networks they were analysing.

Such a mathematical method could be viewed as mimicking how human researchers learn from data, in effect interpreting new information in light of what is already known.

The Warwick researchers led by Dr Sach Mukherjee of Warwick's Department of Statistics and Centre for Complexity Science have devised just such a method that embeds current knowledge in the mathematical analysis to cut through the vast complexity of this type of analysis using a mechanism called "Informative Priors".

The researchers took the 14 protein network and created a mathematical tool that was able to incorporate all of what the interactions, and limits on interactions, that were likely and/or probable in such a network of these particular proteins. This allowed a rapid and accurate analysis of the probabilities of interactions between each on the 14 proteins. The technique even able to cope with misconceptions in current understanding of particular networks as it the was designed to "overturn" any reject any data included in the "Informative Priors" that was consistently at odds with any observed new data.

Analysis with these network models was much better able to resolve complex interactions than simple, correlation-based methods. Moreover using informative priors, gave much more accurate results than analysis that incorporated no prior understanding of the network (so called "flat priors").

The researchers will now use their new technique to examine the network of proteins behind the development of breast cancer but they are also looking at how the tool could be used in social science to mine a vast amount of useful anonymised data from social networking sites such as Facebook to gain significant new understandings of large scale interactions and relationships in society at large.

The research paper is entitled Network inference using informative priors by Dr Sach Mukherjee of the University of Warwick, Terence P. Speed of the University of California, Berkeley. It has just been published in *PNAS*.

Source: University of Warwick