Inferring Genetic Networks via Information Theory

Mathukumalli Vidyasagar, University of Texas at Dallas

Biography:

Mathukumalli Vidyasagar was born in Guntur, India on September 29, 1947. He received the B.S., M.S. and Ph.D. degrees in electrical engineering from the University of Wisconsin in Madison, in 1965, 1967 and 1969 respectively. Between 1969 and 1989, he was a Professor of Electrical Engineering at various universities in the USA and Canada. His last overseas job was with the University of Waterloo, Waterloo, ON, Canada, where he served between 1980 and 1989. In 1989 he returned to India as the Director of the newly created Centre for Artificial Intelligence and Robotics (CAIR) in Bangalore, under the Ministry of Defence, Government of India. Between 1989 and 2000, he built up CAIR into a leading research laboratory with about 40 scientists and a total of about 85 persons, working in areas such as flight control, robotics, neural networks, and image processing. In 2000 he moved to the Indian private sector as an Executive Vice President of India's largest software company, Tata Consultancy Services. In the city of Hyderabad, he created the Advanced Technology Center, an industrial R&D laboratory of around 80 engineers, working in areas such as computational biology, quantitative finance, e-security, identity management, and open source software to support Indian languages. In 2009 he retired from TCS at the age of 62, and joined the Erik Jonsson School of Engineering & Computer Science at the University of Texas at Dallas, as a Cecil & Ida Green Professor of Systems Biology Science. In March 2010 he was named as the Founding Head of the newly created Bioengineering Department. His current research interests are in the application of stochastic processes and stochastic modeling to problems in computational biology, control systems and quantitative finance.

Abstract:

The human body consists of about 22,000 genes, which interact in highly complex ways. Microarray experiments generate data that consists of the "expression levels" of thousands of genes simultaneously, for dozens of humans. From this data, one would like to infer which genes "regulate" which other genes. As stated the problem is hopeless so researchers focus instead on small subnetworks of perhaps 15 or 20 genes and try to infer their regulatory networks. One algorithm that has found some favor in the biology world is called ARACNE. It is based on a powerful theorem in Markov random fields called the Hammersley-Clifford theorem, which states necessary and sufficient conditions under which the joint distribution of n random variables (in this case the gene expression levels) can be factored as a product of joint distributions involving only neighboring random variables. However, this theorem is applicable ONLY to UNDIRECTED graphs. In the biology context, this translates to the assumption that if gene A regulates gene B, then gene B also regulates gene A in precisely the same manner. This is PALPABLY UNTRUE in biology. So we need to prove some version of the Hammersley-Clifford theorem for directed graphs. In this talk, I will review all of the above background, and also show how the concept of conditional relative entropy overcomes the need for having symmetric interactions (undirected graphs). This work is preliminary but if carried through to completion, it will have tremendous impact on computational biology.