

Bayesian networks: a powerful tool for systems biology study

Xiu-Jie WANG

State Key Laboratory of Plant Genomics, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China

© Higher Education Press and Springer-Verlag Berlin Heidelberg 2010

The wide application of omics research approaches caused a burst of biological data in the past decade, and also promoted the growth of systems biology, a research field that studies biological questions from a genome-wide point of view. One feature of systems biology study is to integrate and identify. Not only experiments are carried out at whole-genome scales, but also data from various resources, such as genomics, transcriptomics, proteomics, and metabolics data, need to be integrated to identify correlations among targeted entities. Therefore, plenty amounts of experimental data, robust statistical methods, and reliable network construction models are indispensable for systems biology study. Among the available network construction models, Bayesian network is considered as one of the most effective methods available so far for biological network predictions (Pe'er, 2005).

Bayesian networks are constructed based on the Bayes' theorem. The Bayes' theorem (often called Bayes' law or Bayes' rule) is well-known in probability analysis. It is named for Rev. Thomas Bayes, a British mathematician. The basis of Bayes' theorem is to show the relation between one conditional probability and its reverse. That is, given that event B happened, the occurrence probability of event A depends not only on the relationship between A and B, but also on the absolute probability of A independent of B as well as the absolute probability of B independent of A.

Bayesian networks can be considered as a mechanism to automatically apply Bayes' theorem to complex problems. A Bayesian network is a probabilistic graphic model that represents a set of random variables and their conditional independencies. In Bayesian networks, random variables are presented by nodes, and the conditional dependencies of variables are shown by directed lines. Therefore,

Bayesian networks are directional, and are also named as "directed acyclic graphical model". Bayesian network is very robust to identify relationships among variables, and to find out hidden dependencies among a subset of variables when the dependencies of other variables are given. Thus, it is ideal for relationship identification among large-scale datasets and has been widely applied in systems biology studies (Campos et al., 2004).

As an outstanding molecular systems biologist, Jing-Dong Jackie Han at the Institute of Genetics & Developmental Biology, Chinese Academy of Sciences, Beijing, China, and her research group have accomplished many pioneer works using Bayesian networks to decipher protein interactomes in various species (Xia et al., 2006; Xue et al., 2007; Xia et al., 2008; Yu et al., 2008). To generalize the application of Bayesian networks, in a review of this issue's *Frontiers in Biology*, Liu and Han (Liu and Han, 2010) introduced some basic concepts of probabilistic models and Bayesian networks, explained the structures of Bayesian networks and their building algorithms, and summarized the appropriate applications of Bayesian networks on biological data. The review provides readers a general picture about Bayesian networks along concise and clear written lines. It is undoubtful that with the increment of omics data of various types, the application of Bayesian networks in systems biology analysis will bring more striking discoveries.

References

- Campos L M, Fernández-Luna J M, Huete J F (2004). Bayesian networks and information retrieval: an introduction to the special issue. *Information Processing & Management*, 40: 727–733
- Liu Y, Han J D J (2010). Application of Bayesian networks on large-scale biological data. *Front Biol*. DOI 10.1007/s11515-010-0023-8
- Pe'er D (2005). Bayesian network analysis of signaling networks: A

- primer. *Sci STKE*, 281: 14
- Xia K, Fu Z, Hou L, Han J D J (2008). Impacts of protein-protein interaction domains on organism and network complexity. *Genome Res*, 18: 1500–1508
- Xia K, Xue H, Dong D, Zhu S, Wang J, Zhang Q, Hou L, Chen H, Tao R, Huang Z, Fu Z, Chen Y G, Han J D J (2006). Identification of the proliferation/differentiation switch in the cellular network of multicellular organisms. *PLoS Comput Biol*, 2: e145
- Xue H, Xian B, Dong D, Xia K, Zhu S, Zhang Z, Hou L, Zhang Q, Zhang Y, Han J D J (2007). A modular network model of aging. *Mol Syst Biol*, 3: 147
- Yu H, Zhu S, Zhou S, Xue H, Han J D J (2008). Inferring causal relationships among different histone modifications and gene expression. *Genome Res*, 18: 1314–1324