

A Mutually information method for initialization of K2 algorithms in Bayesian Gen Regularity Network

Sajjad nazmi*

Department of Computer science, Tarbiat Modares University, Tehran, Iran

*altin.nazmi@gmail.com

Abstract: Currently, understanding gene regulatory networks is an important problem in computational biology and has applications in a variety of contexts including discovering new drugs, reducing side effects, therapeutic methods, full genetic disease identification. Various approaches have been used in modeling of gene regulatory networks [2], which the Bayesian network is one of the bests. This approach includes two main structure and parameter learning steps. Here structure learning finds the structure of the network (relation between genes) and parameter learning finds the parameter of the adjusted distributions [1]. So the structure learning is much more important than parameter learning. Once the structure learning of the Bayesian network is determined, the parameter learning comes up. Methods for learning the structure of the Bayesian network are divided into constraint and score-based methods. A greedy k2 algorithm is the important and standard score-based method [3]. This method uses the posterior probability of the network as scores function. This method is very sensitive to the input initial parent candidates. In this paper, we propose a novel method based on mutual information to choose and suitable subsets of the nodes as an input of the K2 algorithm. The experimental results confirm the quality of this method in comparison with the other state the art methods. The steps of the proposed method could be summarized as follows: In the first step mutually information between all pairs of nodes is computed. Then in the second step, we determine the directed edge between the nodes based on obtained mutually information. For both X_i, X_j variables, if the entropy between X_j, X_i (i.e., from X_j to X_i) is greater than X_i, X_j , the direction of the edge between these two variables is from X_i to X_j , and vice versa. If they were equal, however, the edge between vertices cannot be directed using this method and we need to use other methods to direct the rest of the edges. In the third step, we find the initial order for the algorithm using a spanning tree.

Keywords: Gene regulation network; Structure learning; Constraint-based methods; K2 algorithm

References

- [1] Adabor, Emmanuel S., George K. Acquah-Mensah, and Francis T. Oduro. "SAGA: a hybrid search algorithm for Bayesian Network structure learning of transcriptional regulatory networks." *Journal of biomedical informatics* 53 (2015)
- [2] Wang, Shuqiang, et al. "Modeling and analysis of gene regulatory networks with a Bayesian-driven approach." *Communications and Information Technologies (ISCIT), 2014 14th International Symposium on. IEEE, 2014.*
- [3] Cooper, Gregory F., and Edward Herskovits. "A Bayesian method for the induction of probabilistic networks from data." *Machine learning* 9.4 (1992): 309-347).