

Differential study of the cytokine network in the immune system: An evolutionary approach based on the Bayesian networks

Hoai-Tuong Nguyen, Gérard Ramstein, Leray Philippe, Yannick Jacques

▶ **To cite this version:**

Hoai-Tuong Nguyen, Gérard Ramstein, Leray Philippe, Yannick Jacques. Differential study of the cytokine network in the immune system: An evolutionary approach based on the Bayesian networks. The 2nd Asian Conference on Intelligent Information and Database Systems (ACIIDS), Mar 2010, Hue City, Vietnam. pp.?-?. hal-00656723

HAL Id: hal-00656723
https://hal.science/hal-00656723

Submitted on 5 Jan 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Differential study of the cytokine network in the immune system: An evolutionary approach based on the Bayesian networks

Hoai-Tuong NGUYEN[1], Gérard RAMSTEIN[1], Philippe LERAY[1], and Yannick JACQUES[2]

[1] LINA - Laboratory of Informatic of Nantes-Atlantique, UMR 6241 CNRS
La Chantrerie - Christian Pauc, 44306, Nantes Cedex 03, France
{hoai-tuong.nguyen;gerard.ramstein;philippe.leray}@univ-nantes.fr
http://www.lina.univ-nantes.fr/

[2] CRCNA, Center of Research on Cancerology of Nantes/Angers, UMR 829 INSERM
9, Moncousu, 44093 Nantes Cedex 01, France
{yjacques@nantes.inserm.fr}
http://www.crcna.univ-nantes.fr/

Abstract. In this paper, we present a Bayesian networks (BNs) approach in order to infer the differentiation of the cytokine implication in different experimental conditions. We introduce an evolutionary method for BNs structure learning that maintains a set of the best learned networks. Each of them will be tested by a statistic test with two populations of patient data: one with treatment (drugs), other without treatment. The answer of question *"how does the treatment influence to gene regulation?"* is expected.

Key words: Gene expression, microarray, gene regulation networks, Bayesian networks, genetic algorithm, estimation of distribution algorithm, statistic test.

1 Introduction

Interleukin 15 (IL-15) have been found in the recent years [1]. This cytokine plays a critical role in the immune system. Moreover, it has the similar action to the others cytokines. Thus, we would like to know how is the implication of IL-15 between them in the different experiments? What can the bioinformaticians respond this question?

Nowadays, microarray allows to measure simultaneously the expression level of thousands of genes. Furthermore, the gene regulatory networks (GRNs) allow to achieve the regulation of gene expression. Additionally, the inference of gene regulatory networks from high-throughput microarray data is a central problem of biological research. In fact, there are various machine learning and statistical methods that have been proposed to reconstruct this kind of networks. Compared

to others, BNs can solve almost principle problems of this reconstruction: (1) complex interactions involving many genes inferred from sparse and noisy data; (2) massive number of variables (over 30.000 genes), but small number of samples (dozens experiments); (3) computational complexity of structures and statistical significance betweens variables in learned networks. In this paper, we would like to introduce an evolutionary approach in order to obtain a set of the best BNs from microarray data. This allows a comparison of different obtained results of these networks in different kinds of experimental data by a statistic test. In the other words, we would like to answer the question "how can we use the BNs to infer the implication of IL-15 in the different experiments?".

2 Methods

2.1 Machine learning for gene regulatory networks (GRNs) reconstruction

The problem of GRNs reconstruction is well-known. And there are various propositions for this problem, for example: clustering [4], BNs [7], [11], [3], [14], Graphical Gaussian Models [12]... Each model has its advantages and also its open problems. In the our work, we investigate a research to improve the inference accuracy of BNs in the reconstruction of GRNs.

The first model in the literature is proposed by Freidman et al. in 2000 [7] :

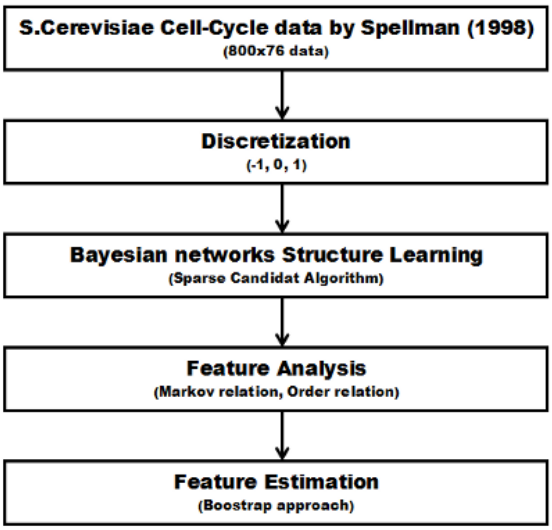


Fig. 1. "Using BNs to Analyze Expression Data", Freidman et al. in 2000

It is one of the most reference for the articles on GRNs reconstruction based on Bayesian networks. In the first work, these authors used a medium size data

and the simple methods for discretization and structure learning. Then, they presented perspectively some typical problems: small number of samples, continuous data, discretization method, temporal expression data, causal patterns and also biological knowledge.

Another well-known model is presented by Peer et al. in 2001 [11]:

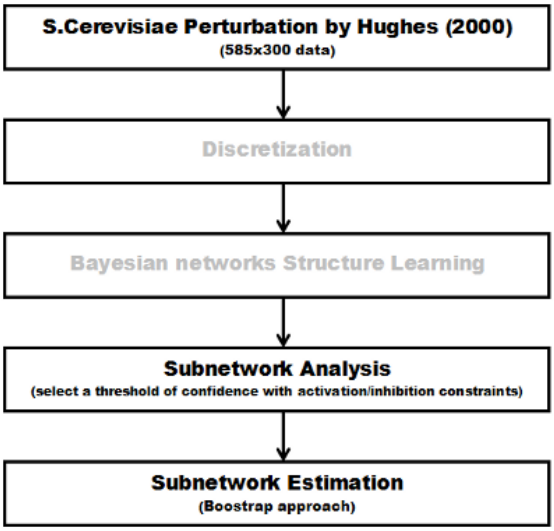


Fig. 2. "Inferring Subnetworks from Expression Profiles", Pe'er et al. in 2001

Pe'er et al. used more samples of experimental data and concentrate their research on the subnetwork analysis by using a threshold of significant confidence with activation/inhibition constraints between variables. They have tested on the continuous data and haven't learned the structure of the network. They were also interested in the latent factors that interact with several observed genes, biological knowledge in their future work.

In our work, we learn the structure of Bayesian networks from the microarray data by the evolutionary method (Figure 3). One of the advantages of microarray is the capacity of simultaneous measurement for the expression level of thousands of genes. Moreover, microarray is also available on the public server where the contribution of many biological laboratories is always appreciated and verified (GEO - Gene Expression Omnibus, Array Express, Oncomine... for example).

In the first phase, we use the evolutionary approach (more detail in the next section) to generate a set of the best BNs according to their scores that were calculated from the experimental data.

Nextly, depending on the nature of each kind of experimental conditions, in which we will test these networks by the statistic test (Figure 4). More precisely, we use a hypothesis testing with two populations of patient data, one with treat-

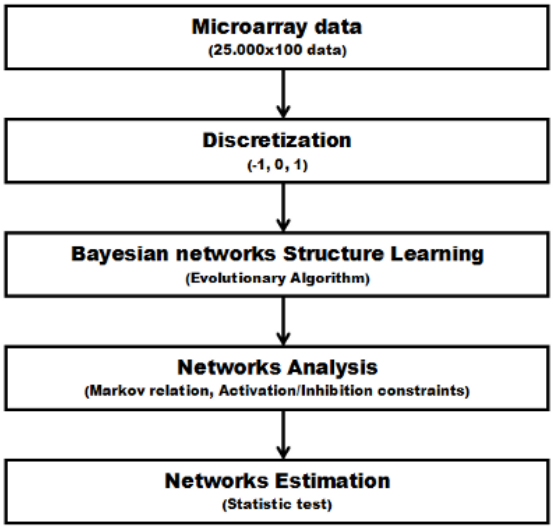


Fig. 3. Our architecture in point of view of two propositions above

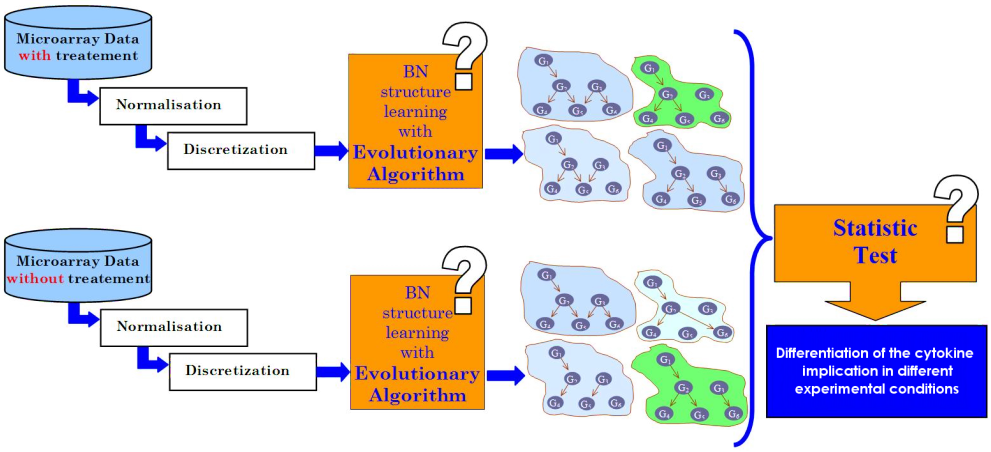


Fig. 4. Our architecture in detail

ment (drugs), other without treatment, to test the result of the best networks produced by the learning methods in the first step. The result of this test will be the answer of question "How does the treatment influence to gene regulation?".

2.2 Bayesian networks (BNs) structure learning

BNs are directed graphical models for representing probabilistic independence relations between multiple interacting entities. Formally, BNs are directed acyclic

graphs (DAGs - a network without any directed cycles) modelling probabilistic dependencies among variables. The graphical structure G of a BN consists of a set of nodes and a set of directed edges.

In the study of reconstruction of gene regulation networks, we use a gene to represent a node and direct influence/interaction between genes to represent an edge. If there is an edge from node A to another node B, then variable B depends directly on variable A (gene A regulates gene B), and A is called a parent of B. In a BN every variable is conditionally independent of its non-descendants given its parents (Markov condition). In the other words, the conditional distribution of a variable A given its parents pa_A in the graph G is $P = P(A|pa_A)$ (parameter of BN, Figure 1). With this simple condition, can infer how well a particular network explains the observed data. For example, in the BN below, the joint distribution decomposes nicely:

$$P(G_1, G_2, G_3, G_4, G_5, G_6) = P(G_1).P(G_3).P(G_2|G_1).P(G_4|G_2).P(G_5|G_2, G_3)$$

In the simplest case, a BN is specified by an expert and then, it is used to perform inference. However, the task of defining the network is too complex for humans. So, the network structure and the parameters of the local distributions must be learned from data. We call this task is BNs learning.

Learning a BN from data requires both identifying the model structure G (*structure learning*) and identifying the corresponding set of model parameter values (*parameter learning*). More simply, given a fixed structure, however, it is straightforward to estimate the parameter values.

To learn the BNs parameter, the common approach is to introduce a statistically motivated scoring function that evaluates each network with respect to the training data, then search for the optimal network according to this score. The most used score is BIC (Bayesian Information Criterion).

To learn the BNs structure, there are two types of methods: (1) *Constraint-based methods* search a database for conditional independence relations and then, construct graphical structures called "patterns" which represent a class of statistically indistinguishable directed DAGs; (2) *Search-and-score methods* perform a search in the space of legal structures. Search-Scoring methods have the advantage of being able to flexibly incorporate prior knowledge and dealing with incomplete data [6]. GA, EDA are the Evolutionary Algorithms that are used as a effect heuristic search engine in the BNs structure learning problem [13], [2].

Which BNs learning algorithms for inferring gene regulatory networks?

In a recent years, there are various researches concentrating motivationally on this problem [8], [10], [9], [2], [5]. For each work, the authors propose their own effective methods to improve the accuracy of the inference of gene regulation networks for a specific type of microarray experiments data. Especially, we are interested in the work of C.Auliac [2] thesis that described perspectively an interesting advantage of BNs structure learning by the evolutionary algorithm. We present more detail this approach in the next section.

2.3 Evolutionary Algorithm (EA) for BN struture learning

Evolutionary algorithm is a subset of evolutionary computation, a population-based heuristic optimization algorithm. EA allows to maintain a set of interesting solutions. One of the most famous representative of EA is genetic algorithm (GA). Recently, an outgrowth of genetic algorithm that are talking about in the EA research is the estimation of distribution algorithm (EDA). With EDA, a population may be approximated with a probability distribution and new candidate solutions can be obtained by sampling this distribution. More precisely, the operation of crossover and mutation of GA are replaced by the probability model building and sampling child population in EDA (see Figure 5). It allows to maintain a set of interesting solutions with the good probabilistic distributions. This could be useful for a statistic test after. It is one of the important goals of our work. Additionally, the way to find a good probability distribution is still a open problem. In fact, there are various version of EDA in order to remedy this problem, such as EBNA (Estimation of Bayesian networks Algorihtm), FDA (Factorized Distribution Algorithm), LFDA (Learning Factorized Distribution Algorithm), BOA (Bayesian Optimization Algorithm). Thus, our work will be also continuing in this interesting topic.

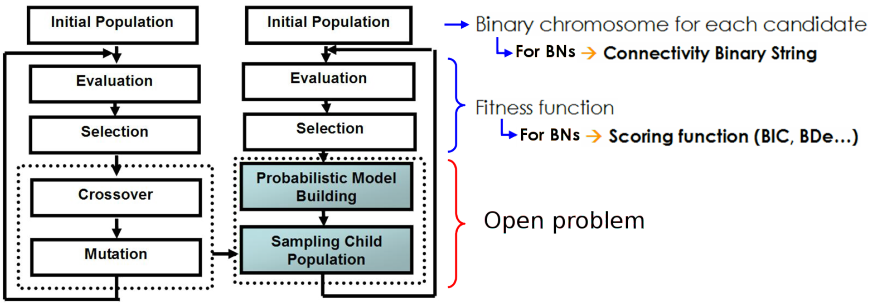


Fig. 5. Comparison between GA and EDA

Application to the case of BNs structure learning, each possible candidate of BNs is represented by an $n \times n$ connectivity string C_{ij}:

$$C_{ij} = \begin{cases} 1 & \text{if } j \text{ is a parent of } i \\ 0 & \text{if otherwise} \end{cases}$$

For each chromosome, we represent an individual of the population by the string :

$$c_{11}c_{21} ::: c_{n1} \quad c_{12}c_{22} ::: c_{n2} \quad \ldots \quad c_{1n}c_{2n} ::: c_{nn}$$

The fitness function in this case is the scoring function calculated from data for each BN. A simple example of EDA computation for BNs structure learning can be found in Figure 7.

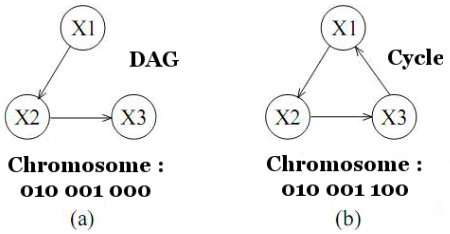


Fig. 6. Representation for a BN in the evolutionary methods

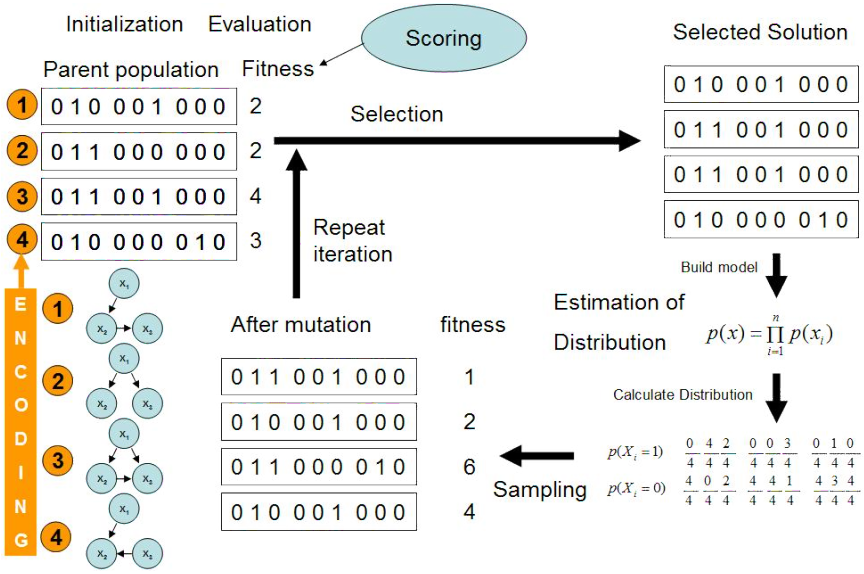


Fig. 7. Example of a simple EDA for the BNs structure learning

3 Conclusion and Future Works

The main goal of this work is the differential analysis using Bayesian networks to reconstruct gene regulatory networks (application to family of cytokine IL15). The evolutionary algorithm is employed in order to maintain a set of good Bayesian networks. After having the best structure, in order to know its real biology performance, we propose to use a hypothesis testing with two populations of patient data, one with treatment (drugs), and other without treatment. Depending on the difference of the result of this test, we can conclude the influence of the treatment on the regulation of the genes. The reconstruction of the gene regulation networks by Bayesian networks will be continuously developed with the bioinformatics research. With the theory as the presentation above, this work is going to nextly presented its implementation and experimentation.

4 Acknowledgements

This research is supported by the BIL project, a regional project of the region Pays de Loire, France.

References

1. A. Arena, RA. Merendino, L. Bonina, D. Iannello, G. Stassi, and P. Mastroeni. The new microbiologica. *Official journal of the Italian Society for Medical, Odontoiatric, and Clinical Microbiology (SIMMOC)*, 23(2), 2000.
2. C. Auliac. Approches évolutionnaires pour la reconstruction de réseaux de régulation génétique par apprentissage de réseaux bayésiens. *PhD Thesis, Université d'Evry-Val d'Essonne, France*, 2008.
3. M. Dejori. Analyzing gene expression data with bayesian networks. *PhD thesis, Technical University of Graz*, 2002.
4. Z. Dongxiao, O. H. Alfred, C. Hong, K. Ritu, and Anand S. Network constrained clustering for gene microarray data. *Bioinformatics*, 2005.
5. S.F. Emmert and M. Dehmer. Analysis of microarray data: A network-based approach. *Wiley-VCH Publishing*, pages 307–329, 2008.
6. O. François. Réseaux bayésiens: de l'identification de structure à la reconnaissance des formes à partir d'informations complètes ou incomplètes. *PhD thesis, INSA Rouen, France*, 2006.
7. N. Friedman, M. Linial, I. Nachman, and D. Pe'er. Using bayesian networks to analyze expression data. *Computer Biology 7(3-4)*, pages 601–620, 2000.
8. F. Geier, T. Jens, and F. Christian. Reconstructing gene-regulatory networks from time series knock-out data, and prior knowledge. *BMC Systems Biology, 1(1):11*, 2007.
9. Y. Huang, J. Wang, Zhang J., Sanchez M., and Y. Wang. Bayesian inference of genetic regulatorynetworks from time series microarray data using dynamic bayesian networks. *Bioinformatics, 2:46-56*, 2007.
10. P. Li, Z. Chaoyang, P. Edward, G. Ping, and Youping D. Comparison of probabilistic boolean network and dynamic bayesian network approaches for inferring gene regulatory networks. *BMC Bioinformatics, 8(Suppl 7):S13*, 2007.
11. D. Pe'er, A. Regev, G. Elidan, and N. Friedman. Inferring subnetworks from perturbed expression profiles. *Bioinformatics (Oxford, England)*, 17(1), 2001.
12. J. Schferand and K. Strimmer. Learning large-scale graphical gaussian models from genomic data. *J. F. Mendes. (Ed.). Proceedings of CNET*, 2005.
13. G. Thibault, S. Bonnevay, and A. Aussem. Learning bayesian network structures by estimation of distribution algorithms: An experimental analysis. *IEEE International Conference on Digital Information Management (ICDIM 07), Lyon, France*, 2007.
14. L. Tiefei. Learning gene network using bayesian network framework. *PhD thesis, National University of Singapore*, 2005.