JOBIM Sattelite's Day ModGraph June 08, 2009 Polytech'Nantes

IL-15 Bayesian Networks EDA

Differential study of the GRNs in the cancerous pathology & therapeutic response:

An **EVOLUTIONARY** approach based on the **BAYESIAN NETWORKS**

PhD Student: NGUYEN, H.T. : LINA-COD
Supervisor: RAMSTEIN, G. : LINA-COD
Co-Director: JACQUES, Y. : INSERM-U892
Director: LERAY, P. : LINA-COD

KEYWORDS: Genes Regulation Networks, Cytokine IL-15, Bayesian Networks, Estimation of Distribution Algorithm

JOBIM Sattelite's Day ModGraph June 08, 2009 Polytech'Nantes

Foreword

IL-15 Bayesian Networks EDA

This is a research project combining :

Bioinformatics, Machine learning and Statistics

The main goal of this presentation is to introduce in general :

1. a **machine learning** approach for **bioinformatics**
   - That is Bayesian networks (BN) for gene regulatory network (GRN) reconstruction
   - in order to infer the differentiation of the cytokine implication in **different experimental conditions**
2. an evolutionary algorithm (EA) for BN structure learning
   - Estimation of Distribution Algorithm (EDA)

JOBIM Sattelite's Day ModGraph June 08, 2009 Polytech'Nantes

Outline

IL-15 Bayesian Networks EDA

JOBIM Sattelite's Day ModGraph June 08, 2009 Polytech'Nantes

Outline

IL-15 Bayesian Networks EDA

JOBIM Sattelite's Day ModGraph June 08, 2009 Polytech'Nantes

Context

What do the biologists need?

IL-15 Bayesian Networks EDA

Biology

- Found **IL-15** in recent years [Arena et al. 2000]
- This cytokine plays a critical role in the immune system
- It has **the similar action** to the others cytokines in this system

Question:

$\longrightarrow$ How is the **implication** of IL-15 in the **different experiments**?

JOBIM Sattelite's Day ModGraph June 08, 2009 Polytech'Nantes

Context

What can the bioinformaticians respond?

IL-15 Bayesian Networks EDA

Bioinformatics & Machine Learning

- **Microarray** allows to measure simultaneously the **expression level** of **thousands** of genes
- **Gene regulatory networks** (GRNs) allow to achieve the **regulation of gene expression**
- There are various **machine learning** methods proposed to reconstruct the GRNs
  - **Bayesian networks** (BNs) can solve major problems of this reconstruction

Question:

$\longrightarrow$ How can we use the **BNs** to infer the **implication** of IL-15 in the **different experiments**?

JOBIM Sattelite's Day ModGraph June 08, 2009 Polytech'Nantes

Problems & Solutions

"If there aren't any problems, we have only the solution"

IL-15 Bayesian Networks EDA

Problem: **Data**

- **Heterogeneous** and **noisy**
- A **massive number** of variables (over 25.000 genes)
- But a **small number** of samples (dozen experiments)

Solution: **Data**

- **Public** database : to increase the number of samples
- **Normalization** : to normalize heterogeneous data
- **Bayesian networks** : to deal with noisy data, the massive number of variables

JOBIM Sattelite's Day ModGraph June 08, 2009 Polytech'Nantes

Outline

IL-15 Bayesian Networks EDA

JOBIM Sattelite's Day ModGraph June 08, 2009 Polytech'Nantes

Machine learning aproaches
for the reconstruction of the GRNs

IL-15 Bayesian Networks EDA

Overview:

1. **Clustering** [MacQueen, 1967]
2. **Boolean Networks** [Kauffman, 1969]
3. **Bayesian Networks**
   - Using BNs to Analyze Expression Data [Freidman et al., 2000]
   - Inferring Subnetworks from Expression Profiles [Pe'er et al., 2001]
   - **Our approach** : Using **a set of BNs** to infer the differential study of the cytokine in different experimental conditions

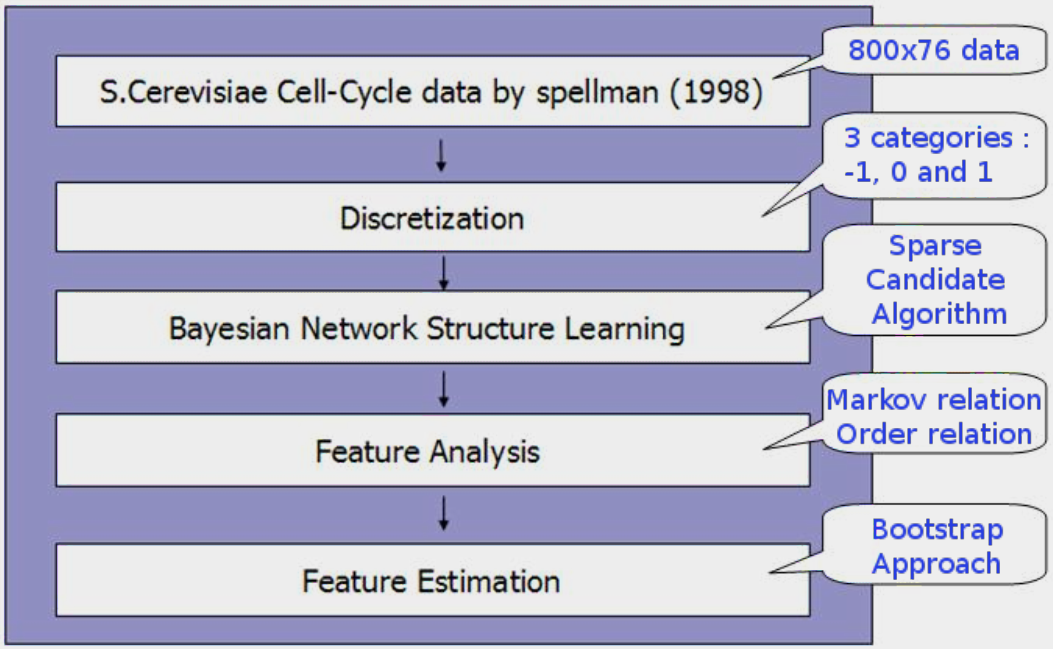
JOBIM Sattelite's Day ModGraph June 08, 2009 Polytech'Nantes

Bayesian networks

Using BNs to Analyze Expression Data [Freidman et al., 2000]

IL-15 Bayesian Networks EDA

Architecture of system proposed by Freidman et al., 2000



Open problems:

Continuous data, discretization method, temporal expression data, causal patterns, biological knowledge

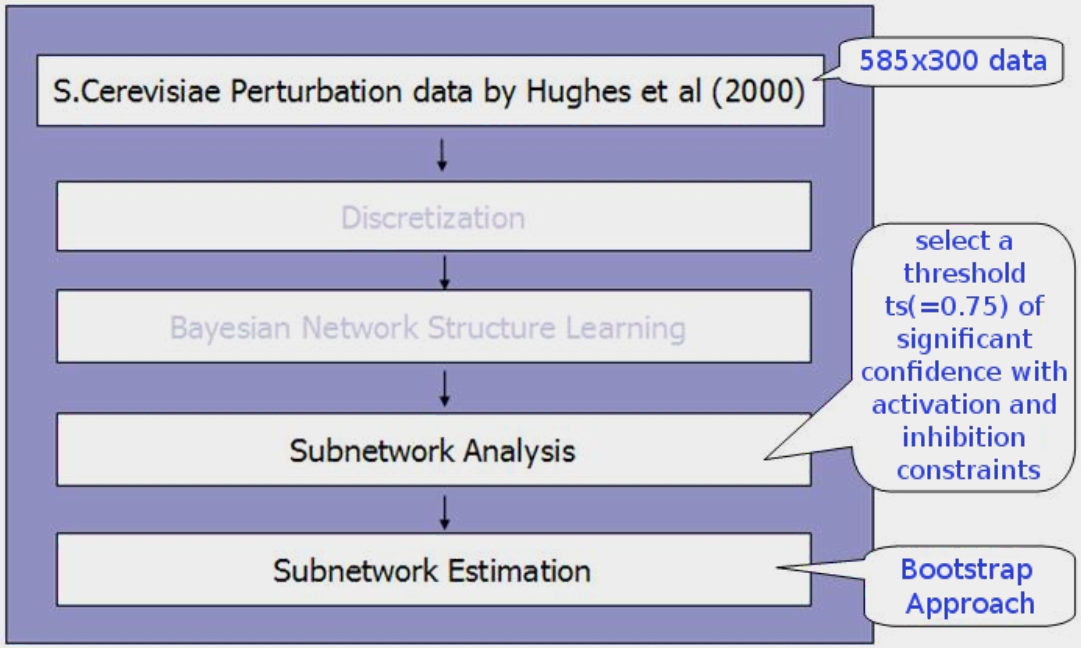
JOBIM Sattelite's Day ModGraph June 08, 2009 Polytech'Nantes

Bayesian networks

Inferring Subnetworks from Expression Profiles [Pe'er et al., 2001]

IL-15 Bayesian Networks EDA

Architecture of system proposed by Pe'er et al., 2001



Open problems:

Latent factors that interact with several observed genes, biological knowledge

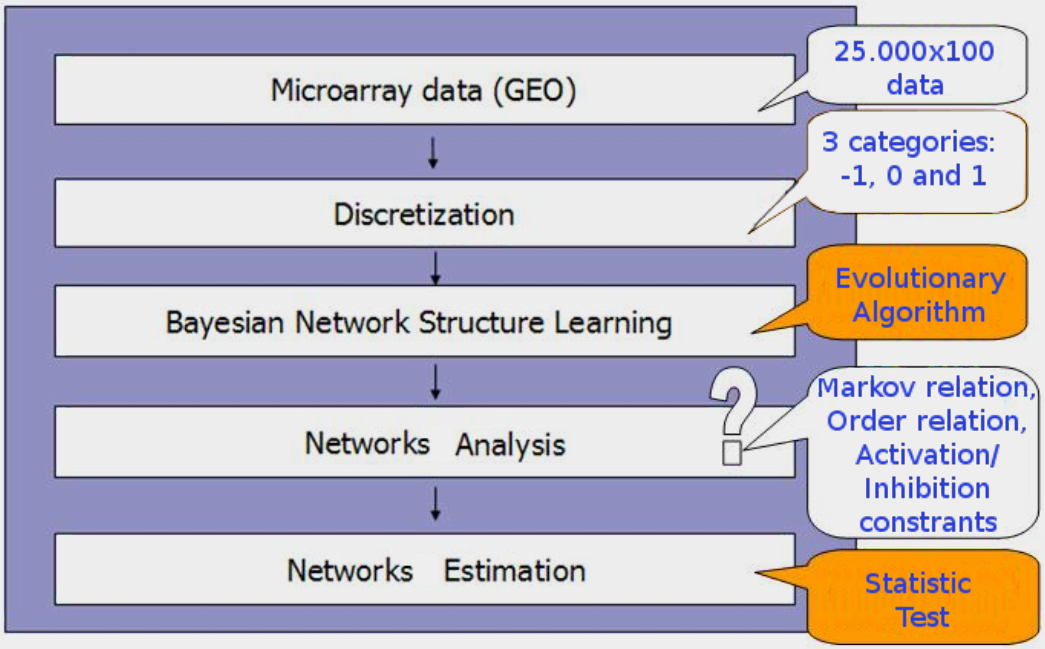
JOBIM Sattelite's Day ModGraph June 08, 2009 Polytech'Nantes

Bayesian networks

Our approach

IL-15 Bayesian Networks EDA

Our architecture of system (overview) :



Open problems:

More detail in **"Future Works"** section

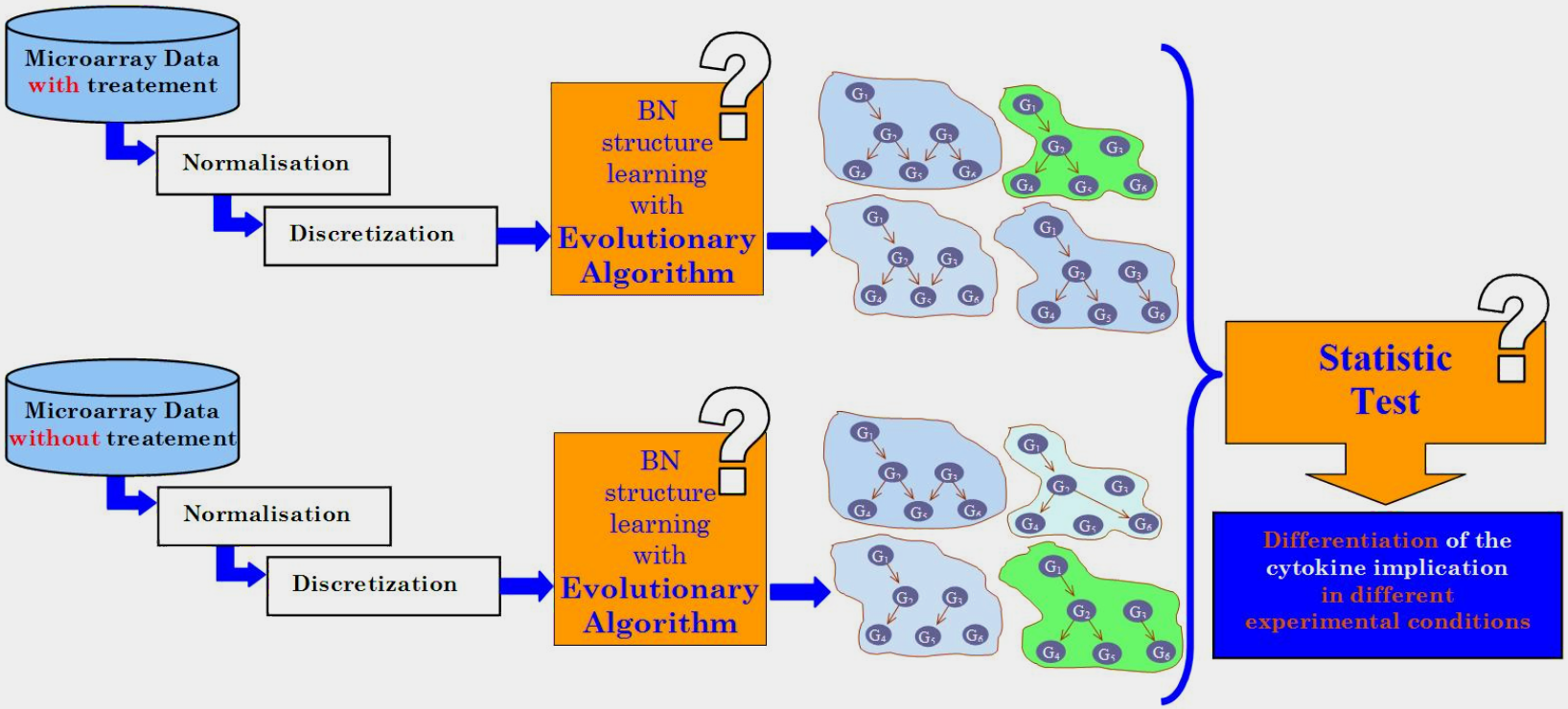
JOBIM Sattelite's Day ModGraph June 08, 2009 Polytech'Nantes

Bayesian networks

Our approach

IL-15 Bayesian Networks EDA

Our architecture of system (detail)



JOBIM Sattelite's Day ModGraph June 08, 2009 Polytech'Nantes

Outline

IL-15 Bayesian Networks EDA

JOBIM Sattelite's Day ModGraph June 08, 2009 Polytech'Nantes

BN structure learning

Different approaches

IL-15 Bayesian Networks EDA

Different approaches to learn the structure of BNs

- From an **expert**
- **Constraint-based**: find in the data for conditional independence relations, then construct graphical structures for these relations
- **Search-Scoring** : search in the space of legal structures for the BNs that maximize the score
  - **Evolutionary Algorithm**:
    $\longrightarrow$ Genetic Algorithm (GA)
    $\longrightarrow$ **Estimation of Distribution Algorithm** (EDA)

JOBIM Sattelite's Day ModGraph June 08, 2009 Polytech'Nantes

Evolutionary Algorithm

and its representative EDA (Estimation of Distribution Algorithm)

IL-15 Bayesian Networks EDA

EA : What and why?

- "EA is a subset of evolutionary computation, a **population-based** heuristic optimization algorithm"
- EA allow to maintain **a set of interesting solutions**

EDA : What and why?

- "EDA is an **outgrowth** of genetic algorithm. In EDA a population may be approximated with **a probability distribution** and new candidate solutions can be obtained by sampling this distribution"
- EDA allows to maintain **a set of interesting solutions** with **the good probabilistic distributions**
  $\longrightarrow$ This could be useful for a **statistic test** after

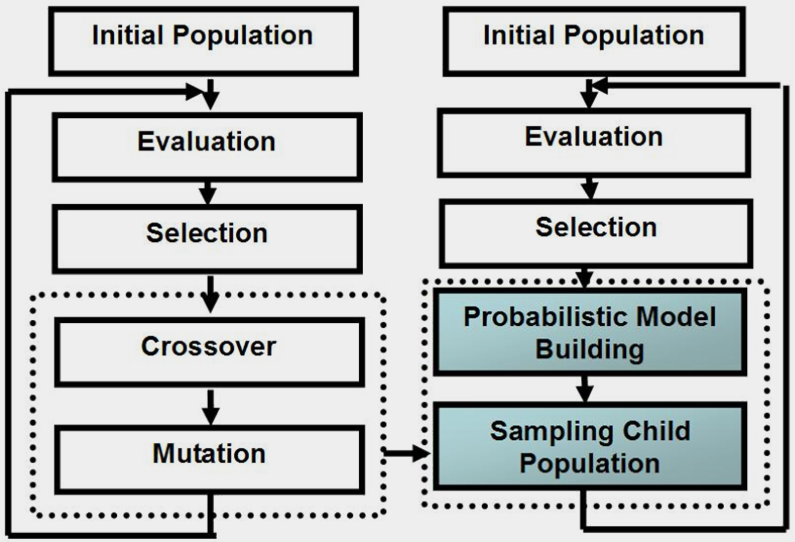
JOBIM Sattelite's Day ModGraph June 08, 2009 Polytech'Nantes

EDA

How does it work ?

IL-15 Bayesian Networks EDA

EDA & GA : **Case study for BN structure learning**



→ Binary chromosome for each candidate
↳ For BNs → **Connectivity Binary String**

Fitness function
↳ For BNs → **Scoring function (BIC, BDe...)**

Open problem

JOBIM Sattelite's Day ModGraph June 08, 2009 Polytech'Nantes

EDA

Different names and approaches

IL-15 Bayesian Networks EDA

Other Names of EDA :

- **PMBGAs** (Probabilistic Model Building Genetic Algorithms)
- **DEAs** (Distribution Estimation Algorithms)
- **IDEAs** (Iterated Density Estimation Algorithms)

EDA is categorised into three groups :

1. **Univariate EDA**
   - **PBIL** (Population Based Incremental Learning) **[Baluja, 1994]**
   - **UMDA** (Univariate Marginal Distribution Algorithm) **[Mühlenbein et al., 1996]**
   - **cGA** (Compact Genetic Algorithm) **[Hariket et al., 1998]**
2. **Bivariate EDA**
   - **MMIC** (Mutual Information Maximization for Input Clustering) **[Bonet et al., 1997]**
   - **COMIT** (Combining Optimizers with Mutual Information Trees) **[Baluja, 1997]**
   - **BMDA** (Bivariate Marginal Distribution Algorithm) **[Pelikan et al., 1999]**
3. **Multivariate EDA**
   - **ECGA** (Extended Compact Genetic Algorithm) **[Harik, 1999]**
   - **EBNA** (Estimation of Bayesian Networks Algorithm) **[Etxeberria et al., 1999]**
   - **FDA** (Factorized Distribution Algorithm) **[Mühlenbein et al., 1999]**
   - **LFDA** (Learning Factorized Distribution Algorithm) **[Mühlenbein et al., 1999]**
   - **BOA** (Bayesian Optimization Algorithm) **[Pelikan et al., 2000]**

JOBIM Sattelite's Day ModGraph June 08, 2009 Polytech'Nantes

Outline

IL-15 Bayesian Networks EDA

JOBIM Sattelite's Day ModGraph June 08, 2009 Polytech'Nantes

Conclusions and Future Works

Conclusions

IL-15 Bayesian Networks EDA

Conclusions :

- The goal of this work is **differential analysis** :
  1. BNs reconstruct GRNs
  2. EAs maintain a set of good BNs
  3. Comparison the obtained BNs in different experiments

Future Works :

- **Theory** : Which type of EDA can be used? May the statistic test be a good approach?
- **Implementation** : A module of BN structure learning by EA with **ProBT(C)** and **EO library** (Evolving Objects)
- **Experimentation** : Test this module with **GEO** (Gene Expression Omnibus) data

JOBIM Sattelite's Day ModGraph June 08, 2009 Polytech'Nantes

Acknowledgements

IL-15 Bayesian Networks EDA

Researchers

RAMSTEIN Gerard
Laboratoire d'Informatique de Nantes-Atlantique, France
Equipe « **COnnaissances et Decisions** »

LERAY Philippe
Laboratoire d'Informatique de Nantes-Atlantique, France
Equipe « **COnnaissances et Decisions** »

JACQUES Yannick
Centre de Recherche en Cancerologie Nantes/Angers, France
Equipe « **Cytokines et Recepteurs** »

HUYNH Xuan-Hiep
Laboratoire d'Informatique de Université de Cantho, Vietnam
Equipe « **Génie Logiciel** »

Questions ?

NGUYEN H.T.

nguyen-ht@univ-nantes.fr

JOBIM Sattelite's Day ModGraph June 08, 2009 Polytech'Nantes

Appendix

Representation for a BN in the evolutionary methods

IL-15 Bayesian Networks EDA

Ideas:

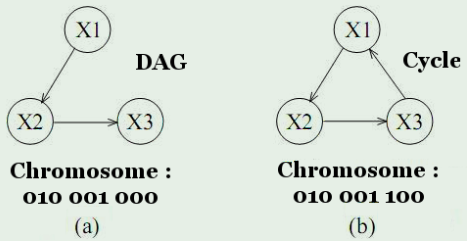
Each possible candidate of BNs is represented by an *nxn* connectivity string C_{ij}

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

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

$$c_{11}c_{21}\dots c_{n1}\ c_{12}c_{22}\dots c_{n2}\ \dots\ c_{1n}c_{2n}\dots c_{nn}$$

Example



JOBIM Sattelite's Day ModGraph June 08, 2009 Polytech'Nantes

Appendix

Example of a simple EDA for the BN structure learning

IL-15 Bayesian Networks EDA

Example