Thesis project: Logical approaches for studying the dynamics of Boolean networks: application to cellular gene networks

8 avril 2022

1 Scientific direction

LIS supervisor : Belaïd Benhamou (laboratoire LIS, équipe LIRICA [1])
(In collaboration with biologist partners from IBDM and INP)

2 Project description

Various approaches for describing cell behavior have been developed by the scientific community. The "virtual cell" contains a set of software allowing to describe the metabolic organization of an organism, to estimate the distribution of matter in its functioning and to extrapolate its production potential.

From the point of view of artificial intelligence, cells are sources of information. Disease, and cancer in particular, can be seen as a pathology of alterations in cell signaling networks. The study of signaling events is a capital key for research in biology. The propagation of these signals involves a change in the behavior of proteins on three levels : regulation of activity, interaction and expression. Signaling networks in cells have been studied for years using analytical methods based on protein recognition by specific antibodies. Alternatively, DNA chips are widely used to study the co-expression of candidate genes to provide an explanation for the etiology of certain diseases.

On the logical side in computer science, the study of gene networks has been studied in artificial intelligence for several years. Information and knowledge in this context are often incomplete, uncertain, revisable, contradictory and multi-source. Indeed, the logical representation of the signaling pathways is not complete : on the one hand, the biological experiments do not give all the interactions between proteins, on the other hand the conditions and the difficulties of the experiments mean that the data are not precise. Finally, the sources from which this information comes are multiple and can be contradictory. We should then revise some information and restore consistency. This observation is noticed in most of the human activities. You would have to find a way to reason and make decisions. This is the goal of all non-classical uncertain logics, and in particular, non-monotonic logics and formalisms.

One of the fundamental applications of AI in this context is to complete gene networks which are often incomplete. Biological experiments are made to complement them. But these are often long and expensive. From the AI point of view, we are here in the well-known framework of abduction. Briefly, abduction consists in finding the set or the *minimal* sets X of information that must be added to a set of known facts F to deduce a result R that we wants to infer. Here, F could be the network of known genes, R the inhibition of a protein and X the interactions between unknown genes. This technique would therefore make it possible to target the biological experiments that should be done, thus would save time and reduce the cost by eliminating experiments that have little chance of leading to usable results.

The two main questions that arise are on the one hand the formal representation of the interactions between the genes and on the other hand trying to complete this information. We are therefore in a framework of representation of incomplete knowledge and revisable reasoning. Some researchers, already established causal relations between the interactions of the genes that they represented by using non-monotonic logics, such as default logic (Rieter 80) or the modal logic of assumptions (P. Siegel and C. Schwind). But in artificial intelligence, algorithms of reasonable and effective algorithmic complexity should be guaranteed in practice. The default logic is a very rich and expressive theoretical framework but the resolution tools are often unconstructive in the general case and ineffective in practice. The logic of the hypotheses has made it possible to have constructive algorithms, but the deduction tools come up against the problem of efficiency.

1. Logique, interaction, Raisonnement et Inférence, Complexité, Algèbre

In 2012, we developed a new semantics (B. Benhamou and P. Siegel ICTAI-12) for logic programming within the framework of answer set programming (ASP) which made it possible to provide an algorithm for calculating stable models with very good complexity properties. The exploitation of this semantics is partly the subject of the thesis of Mr. Tarek Khaled (ex-doctoarnt under my supervision) who implemented the first version of an ASP solver based on this semantics. This new method has been tested on a variety of problems and the experimental results obtained are very encouraging (Tarek Khaled et al. ICTAI-2018, LPAR-2018). We plan to continue this work and design in the future a very concurrent ASP system that we plan to apply in the context of the cell to express the knowledge of biologists concerning signaling networks. The information will be expressed as ASP logic program rules for which we can use our new semantics and method to efficiently compute stable models that could represent for example stable states or attractors in gene networks. The objective is to develop a system allowing :

— among other things, the modeling of the rules of behavior of a gene of a cell.
— reasoning with incomplete, uncertain and reviewable information
— to efficiently calculate the attractors of large gene networks
— to detect and eliminate symmetric attractors
— to manage preferences when several solutions are proposed for the same situation.
— effective abductive reasoning.
— the addition of new rules without calling into question the already existing rules.

Finally, we intend to apply our approaches on the transcriptomic data of a biologist partners (from IBDM and INP) available from mouse models recapitulating faithfully : 1) healthy tissues, 2) tissues sensitized to neoplasia, 3) preneoplastic lesions. The uniqueness of these data lies in their ability to reflect spontaneous events at the origin of tumorigenesis, corresponding to those occurring in human diseases.

2.1 Keywords

Artificial Intelligence, Answer Set programming, Non-montonic reasoning, gene and regulatory networks, Reachability and attractor detecytion in gene networks.

2.2 PhD's expected profile

It is desirable that the candidate has knowledge and skills in the field of Answer Set Programming (ASP) and logic programming. It is also desirable that the candidate master constraint programming (SAT Solvers), Integer Linear Programming (ILP or MILP) and combinatorial optimization. Knowledge of Boolean networks or Pétri nets used in bioinformatics to model biological systems and notions in regulatory networks would be a plus and are desirable. Finally, the candidate must know how to program in imperative programming languages such as C, C ++, and Python.

3 Bibliography

[1] De Jong, Hidde. "Modeling and simulation of genetic regulatory systems : a literature review." Journal of computational biology 9, no. 1 (2002) : 67-103.

[2] Stuart A Kauffman. Metabolic stability and epigenesis in randomly constructed genetic nets. Journal of theoretical biology, 22(3) :437–467, 1969.

[3]Kauffman, Stuart A. The origins of order : Self-organization and selection in evolution. Oxford University Press, USA, 1993.

[4] Shmulevich, Ilya, Edward R. Dougherty, and Wei Zhang. "From Boolean to probabilistic Boolean networks as models of genetic regulatory networks." Proceedings of the IEEE 90, no. 11 (2002) : 1778-1792.

[5]Paulevé, Loïc, Juraj Kolčák, Thomas Chatain, and Stefan Haar. "Reconciling Qualitative, Abstract, and Scalable Modeling of Biological Networks." bioRxiv (2020).

[6] Thomas, René, and Richard d'Ari. Biological feedback. CRC press, 1990.

[7] Thomas, René, Anne-Marie GATHOYE, and Lucie Lambert. "A complex control circuit : Regulation of immunity in temperate bacteriophages." European Journal of Biochemistry 71, no. 1 (1976) : 211-227.

[8] Belaid Benhamou, Pierre Siegel, A New Semantics for Logic Programs Capturing and Extending the Stable Model Semantics. ICTAI 2012 : 572-579

[9] Tarek Khaled, Belaid Benhamou : An ASP-based Approach for Boolean Networks Representation and Attractor Detection. LPAR 2020 : 317-333

[10] Tarek Khaled, Belaid Benhamou : Dealing with Biology Systems in the Framework of Answer Set Programming. KES 2020 : 450-459