Stochastic Semantics of Rule-based Models

Jerome Feret[1], Heinz Koeppl[2], and Tatjana Petrov[3]

Rule-based models give a compact description of protein-protein interactions in spite of combinatorial number of protein complexes interacting in the cellular system. Still, solving the Master equation by observing fully defined species remains hard or unfeasible. We instead detect *fragments*, clusters of species which can describe the stochastic behavior self-consistently and without error. In biological context, fragments arise through statistical independence of some binding and modification events. The procedure for detecting fragments is automatic and is of linear complexity in the size of the rule set.

I. Motivation

Rule-based models are recognized as a valuable modeling tool in explaining signal transduction networks [1,2]. Defined as a graph-rewriting language, they allow a compact description of protein-protein interactions in spite of combinatorial number of protein complexes interacting in the cellular system [3]. However, the combinatorial number of species leaves the computation of both ordinary differential and stochastic semantics hard. People have proposed different methods to reduce the differential semantics of biochemical systems [4,5,6]. In particular, the authors in [6] propose the reduction of differential semantics starting from the rule-based specification: they detect a set of partially defined species-fragments which allow description of the semantics in a self-consistent manner. Case studies therein show a significant reduction from the number of species to the number of fragments. We on the other hand focus on the problem of reducing the *stochastic semantics* of rule-based models. We follow the abstract framework used in [6]. The fragmentation will be more refined than in the differential case, since we can omit the dependency between the species only if they behave identically in each stochastic trajectory. It is of utmost importance that our method avoids working on the fully defined species, because any aggregation method that relies on the enumeration of the state space or even on the enumeration of all reachable species is doomed due to the combinatorial blow-up.

II. Results

We first confirm that the abstraction which gives sound differential semantics cannot be directly applied to the stochastic setting, by experimental illustration on a simple case study. We offer the characterization of fragments which allow sound stochastic simulations. More concretely, we observe how a given fragmentation reflects as an aggregation on the underling continuous-time Markov chain, and we define the sound abstraction as a form of backward bisimulation between labeled transition systems [7], also familiar as weak lumpability on Markov chains [8]. We use the theory of abstract interpretation to formally prove the correctness of our approach. We instantiate this generic framework with a particular specification language, the rule-based language Kappa, and we propose an efficient procedure for computing stochastically sound fragments by static analysis of the rule set and the accompanying initial conditions. The procedure exploits and extends the notion of a contact map [6] of a rule set.

III. Conclusion

We proposed an efficient procedure for computing stochastically sound fragments by static analysis of the rule set and the accompanying initial conditions. A practical aspect is that the fragments represent abstract species and as such they can be fed into any general purpose stochastic simulator for chemical kinetics. We tested our algorithm on a case study of a cross-talk between early EGF and insulin pathway [4]. With originally 2899 species involved, the system is reduced to 247 fragments. In the future work, we are interested in developing metrics which will evaluate the error when using the given non-sound fragmentation. In the perspective, this would allow the modeler to choose how to balance the amount of computational effort and precision.

References

[1] Vincent Danos and Cosimo Laneve (2003): Core formal molecular biology. Theoretical Computer Science, 325:69–110.

[2] Michael L. Blinov, James R. Faeder, and William S. Hlavacek. BioNetGen (2004): Software for rule-based modeling of signal transduction based on the interactions of molecular domains. Bioinformatics, 20:3289–3292.

[3] William S. Hlavacek, James R. Faeder, Michael L. Blinov, Richard G. Posner, Michael Hucka, and Walter Fontana (2006): Rules for Modeling Signal-Transduction Systems. Science's STKE, 344.

[4] Holger Conzelmann, Dirk Fey, and Ernst D. Gilles (2008): Exact model reduction of combinatorial reaction. BMC Syst Biol, 2(78):342–351.

[5] Borisov NM, Markevich NI, Hoek JB, Kholodenko BN. (2006): Trading the micro-world of combinatorial complexity for the macro-world of protein interaction domains. BioSyst. 83:152–66.

[6] Jerome Feret, Vincent Danos, Jean Krivine, Russ Harmer, and Walter Fontana (2009): Internal coarse-graining of molecular systems. Proceedings of the National Academy of Sciences, 106(16):6453–6458.

[7] Peter Bucholz (2007): Bisimulation relations for weighted automata, Theoretical Computer Science, Volume 393, Issues 1-3, Pages 109-123.

[8] J. Kemeny, J. L. Snell.(1960): Finite Markov Chains. Van Nostrand.

[1]LIENS (INRIA/ENS/CNRS), E-mail: feret@ens.fr

[2]Ecole Polytechnique Federale de Lausanne (EPFL) , School of Communication and Computer Sciences, E-mail: heinz.koeppl@epfl.ch

[3]Ecole Polytechnique Federale de Lausanne (EPFL) , School of Communication and Computer Sciences, E-mail: tatjana.petrov@epfl.ch