Theoretical Computer Science 412 (2011) 2798–2799

Contents lists available at ScienceDirect

Theoretical Computer Science

journal homepage: www.elsevier.com/locate/tcs

Preface

Foundations of formal reconstruction of biochemical networks

This special issue of the journal Theoretical Computer Science, Section C (TCS-C) is devoted to a challenging research field of systems biology — "Foundations of formal reconstruction of biochemical networks".

It has been inspired by the stimulating contributions of the 6th Conference on Computational Methods in Systems Biology (CMSB), held in Rostock/Warnemünde in October 2008. This special issue comprises three substantially extended versions of former conference papers, which are complemented by three additionally invited contributions.

The reconstruction of biochemical networks belongs to the core issues in systems biology. Biochemical networks have been designed – as opposed to technical systems – by the organic evolution of living organisms. Network reconstruction aims at understanding of the dynamic interactions between biomolecular components by the generation of explanatory models from observed data. Generated models need to be confirmed by model validation; e.g. model-based predictions on the outcome of perturbed system variants have to be checked by wet-lab experiments.

Obviously, computational methods are crucial for any approach to network reconstruction. Formal approaches to systematically reconstruct biochemical networks do require a rather specific blend of computational methods, which basically have to support three basic steps:

(1) Structure identification.

The first step aims at determining those network structures (topologies), which are in conformance with the experimental data; i.e., the outcome are qualitative models. Ideally, an algorithmic approach generates a complete list of all structures using the minimal number of nodes required to reproduce the experimental data, and accepts user-defined constraints to preclude structures contradicting expert knowledge.

(2) Parameter estimation.

The next step consists of refining a qualitative model into a quantitative one; i.e. the outcome is a stochastic, continuous or hybrid model. For this purpose, kinetic rate laws and their kinetic rate constants (usually called parameters) need to be determined. Model fitting may also involve estimation of the required initial mass.

(3) Model validation.

Finally, the behaviour of the derived computational model needs to be thoroughly validated against the behaviour of the actual experimental system. This requires behaviour checking, which might be better known in computer science as model checking. Depending on the kind of quantitative modelling style applied, analytical and/or simulative model checking techniques might be appropriate.

In practice, a family of such computational methods will be helpful; it will generally depend on the peculiarities of the system in hand, which combination of methods is the most suitable one. We can only scratch the surface in this special issue. Thus, the variety of methods of each step are just illustrated by two papers, with one being based on a former CMSB 2008 paper and one being additionally invited. The contributions in this special issue are ordered according to the three basic steps sketched above.

We start off with "An algorithmic framework for network reconstruction" by Durzinsky, Wagler, and Weismantel. It uses the modelling paradigm of Petri nets to introduce a combinatorial approach to structure identification. The proposed algorithmic framework generates all possible Petri nets fitting the given experimental data, under the condition that the data have a sufficient quality. The number of generated networks may exponentially grow. Thus, a refined reconstruction approach is outlined for so-called monotone data and some further suitable assumptions, which together substantially bound the number of solutions generated.

The combinatorial approach is demonstrated by reconstructing the sporulation network of physarum polycephalum plasmodia.

On the contrary, the second paper, "Parameter estimation for Boolean models of biological networks" by Dimitrova et al., follows an algebraic approach to structure identification within the modelling paradigm of Boolean networks, a popular modelling technique to study the dynamics of gene regulator networks. Observed phenomena are modelled in terms of a finite-state polynomial model, which allows the use of algebraic tools, like Gröbner bases, to find the set of all solutions. This paper describes a software package, Polynome, which integrates several algorithms for parameter estimation and model

0304-3975/$ – see front matter
doi:10.1016/j.tcs.2011.02.001

simulation. It can also easily be used to reconstruct Boolean network models based on experimental data and biological input. For this purpose, the Boolean functions of the Boolean network are represented as general polynomials, with undetermined (0/1) coefficients. Inference methods use experimental time course data to estimate functions that will result in a model that fits the data. This principle might be considered as a discrete analogue of parameter estimation.

To illustrate its performance, the package is applied to the well-known lac operon, the network that regulates lactose metabolism E. coli.

The third paper, "Continuous valuations of temporal logic specifications with applications to parameter optimization and robustness measures" by Rizk et al., opens the section on parameter estimation. It defines a continuous violation degree that quantifies how far from satisfaction an LTL formula is in a given model. The model behaviour is assumed to be characterized by its numerical traces, which allows to apply the approach to a wide variety of modelling paradigms. The paper gives two examples demonstrating the use of such a satisfaction measure to reason about numerical traces (such as continuous, stochastic or hybrid simulation traces). First, it can be deployed as a fitness function with evolutionary optimization methods to find kinetic parameter values satisfying a set of biological properties formalized in temporal logic. Second, it can be used to estimate the robustness of a biological model with respect to its temporal specification.

These methods are evaluated on models of the cell cycle and of the MAPK signalling cascade.

The next paper, "Component-based construction of bio-pathway models: The parameter estimation problem" by Koh, Hsu and Thiagarajan, aims at alleviating the high-dimensional search space problem in parameter estimation by proposing an approach which exploits the network structure to identify network components. At first, the components are considered separately, before being assembled and adjusted to each other in a global network model. To capture the structure, models are represented as Reaction Network Graphs (RNG), which could also have been called continuous Petri nets (CPN), and each component is represented as a factor graph. RNG and CPN are bipartite graph; they enjoy an unambiguous structure and uniquely define systems of ordinary equations. Factor graphs are standard probabilistic graphical models, supporting a standard probabilistic inference technique — the belief propagation, which is done over the combined factor graph of all components. The results are optimized parameter values that are globally consistent.

The approach is validated on a synthetic network model based on the Akt-MAPK signalling pathways.

The final section devoted to model validation starts with the paper "CTRL: Extension of CTL with regular expressions and fairness operators to verify genetic regulatory networks" by Mateescua et al. It provides a temporal specification language that allows expressing properties of biological interest and tries to find a suitable compromise between expressive power, user-friendliness, and complexity of model checking. For this purpose, it defines Computation Tree Regular Logic (CTRL), which subsumes both CTL and LTL. A particular strength of this logic is the convenient specification of multistability and oscillation properties. CTRL formulae are translated into Hennessy–Milner Logic with Recursion (HmlR), which allows reusing the verification technology available in the CADP toolbox.

The use of CTRL and its on-the-fly model checker are demonstrated by analysing the genetic regulatory networks controlling the carbon starvation response of Escherichia coli.

We conclude this special issue with the paper "IDD-based model validation of biochemical networks" by Schwarick and Tovchigrechko. It presents efficient techniques for the qualitative and quantitative analysis of biochemical networks, which are modelled as qualitative or stochastic Petri nets. The supported analyses cover standard Petri net properties as well as model checking of Computation Tree Logic (CTL) and Continuous Stochastic Logic (CSL). There are two crucial points for the reported efficiency gain: the sophisticated data structure – the interval decision diagrams (IDD), which generalize the basic principles of binary decision diagrams, and the efficient algorithms, which exploit the network structure and one of the basic principles of Petri nets – locality. Further key buzzwords are saturation-based reachability analysis and state space construction and on-the-fly (i.e. matrix-free) transient and steady state analysis of continuous-time Markov chains (CTMC). The techniques are implemented in the tool Marcie (formerly called IDD-MC), which outperforms comparable tools, such as SMART and PRISM, for typical benchmarks of biochemical networks.

Many more promising approaches can be found in the literature; some of them are referenced in the papers of this special issue.

We wish all readers of this issue an enjoyable journey through the challenging field of formal reconstruction of biochemical networks. A rigorous automation of the network reconstruction process will make the modelling outcome reproducible, more transparent and easier to understand. However, there is still a long way to go.

Monika Heiner*
Institute of Computer Science, Brandenburg University of Technology, D-03013 Cottbus, Germany
E-mail address: monika.heiner@informatik.tu-cottbus.de.

Adelinde Uhrmacher
Department of Computer Science, Universität Rostock, D-18059 Rostock, Germany
E-mail address: adelinde.uhrmacher@uni-rostock.de.

* Corresponding editor. Tel.: +49 355 69 3884; fax: +49 355 69 3587.