Information and Computation 247 (2016) 235–253

Contents lists available at ScienceDirect

Information and Computation

www.elsevier.com/locate/yinco

Smoothed model checking for uncertain Continuous-Time Markov Chains [☆]

Luca Bortolussi [a,b,c,*], Dimitrios Milios [d], Guido Sanguinetti [d,e]

[a] Department of Mathematics and Geosciences, University of Trieste, Italy
[b] Modelling and Simulation Group, University of Saarland, Germany
[c] CNR/ISTI, Pisa, Italy
[d] School of Informatics, University of Edinburgh, United Kingdom
[e] SynthSys, Centre for Synthetic and Systems Biology, University of Edinburgh, United Kingdom

ARTICLE INFO

Article history:
Received 10 May 2015
Received in revised form 28 August 2015
Available online 12 January 2016

Keywords:
Model checking
Uncertainty
Continuous-Time Markov Chains
Gaussian Processes

ABSTRACT

We consider the problem of computing the satisfaction probability of a formula for stochastic models with parametric uncertainty. We show that this satisfaction probability is a smooth function of the model parameters under mild conditions. This enables us to devise a novel Bayesian statistical algorithm which performs model checking simultaneously for all values of the model parameters from observations of truth values of the formula over individual runs of the model at isolated parameter values. This is achieved by exploiting the smoothness of the satisfaction function: by modelling explicitly correlations through a prior distribution over a space of smooth functions (a Gaussian Process), we can condition on observations at individual parameter values to construct an analytical approximation of the function itself. Extensive experiments on non-trivial case studies show that the approach is accurate and considerably faster than naive parameter exploration with standard statistical model checking methods.

1. Introduction

Computational verification of logical properties by model checking is one of the great success stories of theoretical computer science, with profound practical implications. Robust and mature tools such as PRISM [1] and MRMC [2] implement probabilistic model checking, enabling the quantification of the truth probability of a formula for a wide variety of stochastic models and logics. Recently, the applicability of model checking has been greatly extended by the development of randomized algorithms which leverage the possibility of drawing large number of samples from generative models such as Continuous-Time Markov Chains (CTMCs), by means of stochastic simulation algorithms such as SSA [3]. Statistical model checking (SMC, [4,5]) repeatedly draws independent samples (runs/ trajectories) from the model to estimate satisfaction probabilities as averages of satisfactions on individual runs; by the law of large numbers, these averages will converge to the true probability in the limit when the sample size becomes large, and general asymptotic results permit to bound (probabilistically) the estimation error.

[☆] We thank Botond Cseke for sharing with us his code for Expectation Propagation, Alberto Policriti for highlighting the analogy with Smoothed Complexity Analysis, and Jane Hillston and Vincent Danos for useful conversations. D.M. and G.S. acknowledge support from the ERC under grant MLCS 306999. L.B. acknowledges partial support from EU-FET project QUANTICOL (nr. 600708), by FRA-UniTS, and by the German Research Council (DFG) as part of the Cluster of Excellence on Multimodal Computing and Interaction at Saarland University.

[*] Corresponding author.
E-mail address: luca@dmi.units.it (L. Bortolussi).

http://dx.doi.org/10.1016/j.ic.2016.01.004
0890-5401/

Both analytical and statistical tools for model checking however start from the premises that the underlying mathematical model is fully specified (or at least that a mechanism to draw independent and identically distributed samples exists in the case of SMC). This is both conceptually and practically problematic: models are abstractions of reality informed by domain expertise. Condensing the domain expertise in a single vector of parameter values is at best an approximation. We are particularly interested in modelling biological systems: in this important application area for formal methods, parametric uncertainty is the norm, and reasoning about how the qualitative properties of a system may vary while varying parameters is often of paramount interest. While parameter estimation from measurements has seen considerable progress in recent years [6–10], uncertainty can never be completely eliminated. In such cases, acceptance of the inherent uncertainty is the natural way forward, and alternative semantics such as Interval Markov Chains [11] or Constraint Markov Chains [12] should be preferred. Model checking methodologies for these semantics are however in their infancy and mostly rely on a reduction to continuous-time Markov Decision Processes [13], obtaining upper and lower bounds on the satisfaction probability [11,14]. Other approaches such as [15,16] provide provably correct piecewise constant upper and lower bounds on the satisfaction probability, but rely on a computationally intensive use of numerical transient analysis.

In this paper we define a novel, quantitative approach to model checking uncertain CTMCs. We start by defining the *satisfaction function* of a formula, the natural extension of the concept of satisfaction probability of a formula to the case of CTMCs with parametric uncertainty. We prove that, under mild conditions, the satisfaction function is a smooth function of the uncertain parameters of the CTMC. We then propose a novel statistical model checking approach which leverages this smoothness to transfer information on the satisfaction of the formula at nearby values of the uncertain parameters. We show that the satisfaction function can be approximated arbitrarily well by a sample from a *Gaussian Process* (GP), a non-parametric distribution over spaces of functions, and use the GP approach to obtain an analytical approximation to the satisfaction function. This enables us to predict the value (and related uncertainty) of the satisfaction probability *at all values of the uncertain parameters* from individual model simulations at a finite (and generally rather small) number of distinct parameter values. We term this whole approach *smoothed model checking* in analogy with the recently proposed smoothed complexity analysis, another traditional domain of discrete mathematics where embedding problems in a continuous framework has proved highly valuable [17]. We show on three non-trivial examples that smoothed model checking can provide an accurate estimation of the satisfaction function with considerably fewer simulations than naive exploration by SMC, providing in our examples computational savings of an order of magnitude for equivalent accuracy.

The rest of the paper is organised as follows: in the next section we introduce the mathematical background on formulae and model checking, proving in the following section that the satisfaction function is a smooth function of model parameters. We then introduce our smoothed model checking framework and discuss the statistical tools needed. We demonstrate the power of our approach on three nontrivial examples, and conclude the paper by discussing the broader implications of our approach.

2. Background

In this section, we briefly review continuous-time Markov chains, Metric Interval Temporal Logic, and statistical model checking approaches.

2.1. Continuous-Time Markov Chains

A Continuous-Time Markov Chain (CTMC) $\mathcal{M}$ is a Markovian (i.e. memoryless) stochastic process defined on a finite or countable state space S and evolving in continuous time [18]. A CTMC on S is fully characterised by its infinitesimal generator matrix $Q \in \mathbb{R}^{|S|\times|S|}$ [19]. For any $i, j \in S$, the entry Q_{ij} denotes the rate parameter of an exponential distribution that governs the time of transitioning from i to j.

Instead of representing the generator Q explicitly, we will specifically consider population models of interacting agents [20], which can be easily represented by:

- a vector of population variables $\vec{X} = (X_1, \dots, X_n)$, counting the number of entities of each kind, and taking values in $S \subseteq \mathbb{N}^n$;
- a finite set of reaction rules, describing how the system state can change. Each rule η is a tuple $\eta = (\vec{r}_\eta, \vec{s}_\eta, \tau_\eta)$. $\vec{r}_\eta$ (respectively $\vec{s}_\eta$) is a vector encoding how many agents are consumed (respectively produced) in the reaction, so that $\vec{v}_\eta = \vec{s}_\eta - \vec{r}_\eta$ gives the net change of agents due to the reaction. $\tau_\eta = \tau_\eta(\vec{X}, \theta)$ is the rate function, associating to each reaction the rate of an exponential distribution, depending on the global state of the model and on a d dimensional vector of *model parameters*, θ. Reaction rules are easily visualised in the chemical reaction style, as follows:

  $$r_1 X_1 + \dots r_n X_n \xrightarrow{\tau(\vec{X},\theta)} s_1 X_1 + \dots s_n X_n$$

  For the population models we consider, the entries of the generator matrix can be reconstructed as follows:

  $$Q_{ij} = \sum_{\eta} \tau_\eta(\vec{X}, \theta)$$

  where $i = \vec{X}$ and $j = \vec{X} + \vec{v}_\eta$.