Temporal Logic for Systems Biology

Chris Banks

PEPA club

18th Feb, 2011

Introduction

- Temporal Logic (TL) has long been used in Computer Science for the specification and verification of systems.
- In the emerging field of Computational Systems Biology (CSB) computational theoretic techniques have been applied to problems in Systems Biology.
- TL, therefore, has been a natural development for specifying and verifying models of biological systems.

Part I

A (very) brief overview...

Inception

- One of the first uses of TL for CSB was Pathway Logic:
  - Eker et al. (2002) at SRI International [6]
  - Used the Maude specification language and toolset:
    - Model was an algebraic specification with a rewrite theory.
    - Model checking followed naturally from this.
    - Used LTL over the computed transition system.
  - Maude could also run simulations.
    - Including backward simulation - "from what initial state(s) can we get some desired (or known) final state?"

Branching Time

- Chabrier and Fages (2003 - Projet Containtes at INRIA) proposed the use of branching time logic CTL: [4]
  - as a query language for biological processes.
  - With symbolic model checking for qualitative models
  - and constraint based model checking for quantitative models.
- Concurrent Transition Systems as the qualitative model.
  - Model checking in NuSMV.
- Euler's method used to generate a deterministic transition system from ODEs as the quantitative model.
  - Renders the universe quantifiers (**A**,**E**) meaningless.
  - However the constraint based model checking used applies to infinite state Kripke structures - e.g. with variables ranging over continuous numerical domains.
- Applied queries to Kohn's mammalian cell cycle model.

TL as a specification

- Often model parameters (e.g. reaction rates) are not available and difficult to obtain by experimental methods.
- This is often avoided by sticking to qualitative methods.
- Bernot et al. (2004 - J.Theo.Bio.) introduce the idea of finding suitable models/parameters by expressing desired behaviour as a set of CTL formulae. [1]
  - Given a set of models, find the subset that satisfy the specification.
  - A set of models could be one structure with a range of possible parameters.

Stochastic models

- In a 2006 paper Calder et al. formulate a signalling pathway model as a CTMC. [2]
- The PRISM model checker and Continuous Stochastic Logic was used to analyse the model.
- This laid the foundations for stochastic process algebra models, e.g. Sπ and BioPEPA, to be analysed by model checking.
  - This had been proposed, but not detailed, in the setting of PEPA in a 2004 paper by Calder, Gilmour, Hillston.

Part II

Bringing it all together...

BioCHAM

"...a software environment for for modelling complex cell processes, making simulations (i.e. "in silico experiments"), formalizing the biological properties of a system known from real experiments, checking them, and using them as specification when refining a model."

- Developed by Fages et al. at INRIA. [5, 7]

BioCHAM

- A simple rule-based language:

$$
\begin{aligned}
\mathit{object} &= \mathit{mol} \mid \mathit{mol} :: \mathit{location} \\
\mathit{mol} &= \mathit{name} \mid \mathit{mol} - \mathit{mol} \mid \mathit{mol} \sim \{\mathit{name}_1, ..., \mathit{name}_i\} \\
\mathit{reaction} &= \mathit{solution} \Rightarrow \mathit{solution} \mid \mathit{kinetics}\ \texttt{for}\ \mathit{solution} \Rightarrow \mathit{solution} \\
\mathit{solution} &= \_ \mid \mathit{object} \mid \mathit{number} * \mathit{object} \mid \mathit{solution} + \mathit{solution}
\end{aligned}
$$

- Examples:
  - $A + B \Rightarrow A - B$
    complexation of A and B.
  - $A + B \Rightarrow A \sim \{p\} + B$
    phosphorylation of A by B.
  - $A \Rightarrow \_$
    degradation of A.
  - $A :: L_1 \Rightarrow A :: L_2$
    transport of A from L_1 to L_2.

BioCHAM

The language then has three interpretations:

- Boolean semantics
  - Objects are only represented by a boolean variable denoting their presence or absence.
  - Represented by a Kripke structure.
  - CTL model checking is easily done over this structure and the NuSMV symbolic model checker is used.
  - Includes commands for commonly used CTL formulae.

BioCHAM

- Population semantics
  - Every object has an associated integer denoting the number of molecules in the system.
  - Represented by a CTMC.
  - PLTL($\mathbb{Z}$) formulae with numerical constraints are evaluated with their probability.
    - $([A] = 0) \implies \mathbf{P}_{\geq 1}\mathbf{F}([A] > 0)$
      given that we start with no A, is it 100% likely that we produce some?

BioCHAM

- Concentration semantics
  - Every object has an associated real number denoting the concentration in the system.
  - Represented by a system of ODEs.
  - LTL($\mathbb{R}$) formulae can be checked over numerical traces of the ODE solutions.
    - $\mathbf{G}([A] \leq [B])$
      the concentration of B is always greater than the concentration of A.
    - $\mathbf{F}(d[A]/dt > 0)$
      eventually the concentration of A increases.

BioCHAM

BioCHAM also uses TL specifications for rule inference and parameter fitting. [3]

- CTL specifications can be used to infer rules in the boolean semantics.
  - A *reaction pattern* is supplied and permutations are attempted until a rule addition (or deletion) causes the specification to be satisfied.
  - Multiple additions/deletions can be inferred at once.
    - The performance of this is improved by the theorem:
      $s \not\models_{K'} \phi_{ECTL} \implies s \not\models_{K} \phi_{ECTL}$
      $s \not\models_{K} \phi_{ACTL} \implies s \not\models_{K'} \phi_{ACTL}$
      where $K = (S, R); K' = (S, R'); R \subseteq R'$.
- A similar method is used to fit parameters in the concentration and population semantics - given a range of parameters and a constraint LTL specification.
  - However, because of the cost of probabilistic model checking highly non-deterministic systems (as we often get in biology), parameter fitting in the population semantics is not practical for large models.

Part III

Not everything in biology is black and white...

Degree of satisfaction

- It is useful to know whether a model satisfies a TL specification.
- However, if a model does not satisfy a formula then it might be nice to know how far off satisfaction we were.
- Rizk et al. (2008) do just this. [8]
- Has obvious applications in guiding parameter search.

Degree of satisfaction

- $LTL(\mathbb{R}) \mapsto QFLTL$
- QFLTL is the quantifier free fragment, with real constraints $(c_1 \ldots c_n)$ abstracted as variables $(x_1 \ldots x_n)$.
- For an $LTL(\mathbb{R})$ formula ϕ and a QFLTL abstraction ϕ_{qf} the objective $obj(\phi)$ is the single point in the space $\mathbb{R}^n$ of ϕ_{qf} where $x_i = c_i$ for all $1 \leq i \leq n$.
- The algorithm computes the exact domain of validity $D_{\phi_{qf}}(T) \subset \mathbb{R}^n$ for the trace T.
- The violation degree is then simply the Euclidean distance between $D_{\phi_{qf}}(T)$ and $obj(\phi)$.

Degree of satisfaction

- Used in optimization techniques to guide parameter search.
- Also used to define a degree of robustness
  - Inverse of the average violation degree over a set of perturbations.

Part IV

In the abstract for this talk I wrote:

> *"I will then conclude by commenting (tentatively!) on my proposed PhD topic and how I could offer potential improvement."*

The response I received from a certain predecessor of mine:

> *"I'd suggest a haircut."* - *Dr. Marek Kwiatkowski*

I do have some *slightly* better ideas. . .

$c\pi$ model checking

Direct symbolic model checking of $c\pi$ models.

- Performance?
- Branching time?
- Modularity?

If that is done, there is this whole body of ideas which can be applied:

- Parameter fitting / rule inference?
- Quantitative robustness analysis.
- Loads of ideas for biological applications.

References I

[1] Gilles Bernot, Jean-Paul Comet, Adrien Richard, and Janine Guespin.
Application of formal methods to biological regulatory networks: extending Thomas' asynchronous logical approach with temporal logic.
Journal of theoretical biology, 229(3):339–47, August 2004.

[2] Muffy Calder, Vladislav Vyshemirsky, David Gilbert, and Richard Orton.
Analysis of signalling pathways using continuous time Markov chains.
Transactions on Computational Systems Biology VI, pages 44–67, 2006.

[3] Laurence Calzone, N. Chabrier-Rivier, F. Fages, and S. Soliman.
Machine learning biochemical networks from temporal logic properties.
Transactions on Computational Systems Biology VI, pages 68–94, 2006.

References II

[4] Nathalie Chabrier and F. Fages.
Symbolic model checking of biochemical networks.
In *Computational Methods in Systems Biology*, pages 149–162. Springer, 2003.

[5] N. Chabrier-Rivier, F. Fages, and S. Soliman.
The biochemical abstract machine BIOCHAM.
In *Computational Methods in Systems Biology*, pages 172–191. Springer, 2005.

[6] S. Eker, M. Knapp, K. Laderoute, P. Lincoln, Jose Meseguer, and Kemal Sonmez.
Pathway logic: Symbolic analysis of biological signaling.
In *Pacific Symposium on Biocomputing 2002: Kauai, Hawaii, 3-7 January 2002*, page 400. World Scientific Pub Co Inc, 2002.

References III

[7] F. Fages.
Temporal logic constraints in the biochemical abstract machine biocham.
Logic Based Program Synthesis and Transformation, pages 1–5, 2006.

[8] A. Rizk, G. Batt, F. Fages, and Sylvain Soliman.
On a continuous degree of satisfaction of temporal logic formulae with applications to systems biology.
In *Computational Methods in Systems Biology*, pages 251–268. Springer, 2008.