Towards a Language for Specifying Properties of Simulation Trajectories

Tom Warnke, Danhua Peng, Fiete Haack, Adelinde Uhrmacher

University of Rostock, Institute of Computer Science,
Albert-Einstein-Straße 22, 18059 Rostock, Germany
`{tom.warnke,danhua.peng2,fiete.haack,adelinde.uhrmacher}@uni-rostock.de`
`http://www.uni-rostock.de`

1 Introduction

There is an increasing interest in annotating models with information about
simulation experiments. Unambiguous experiment descriptions facilitate the re-
producibility of simulation results, which is a basic requirement for model publi-
cation [5]. Specifying the crucial observations from experiments regarding model
behavior, together with experiment configurations (e.g., values of model pa-
rameters and stop conditions), provides useful information for model reuse [4].
The dynamic behavioral properties a model exhibits can be described by the
properties of simulation trajectories produced in simulation experiments, e.g.,
oscillation of a protein concentration, and increase of a protein concentration to
a certain peak value followed by a decrease. Often such properties and their tem-
poral order are specified qualitatively with temporal logics like LTL (e.g., [2]).
However sometimes, also quantitative temporal information about the produced
trajectories is of interest.

2 Our approach

In our previous work [4], the domain specific language SESSL (Simulation Exper-
iment Specification via a Scala Layer) was extended to specify Linear Temporal
Logic (LTL) properties and simple probabilistic statements. Based on that, we
now propose a language to describe properties of trajectories. The language cur-
rently allows to specify properties explicitly referring to time points (denoted
with P) and episodes (denoted with E). In both cases, time constraints and an
identifier (which can be used for further constraints) can optionally be added.
The three parts of the pattern, i.e., properties, time constraints, and identifiers,
are separated by semicolons. For example, $P((a > 0); t > 5 \wedge t < 8; t)$ states that
between time point 5 and 8 a time point t exists where the model variable a is
greater than 0. While describing an episode, the start, the end or the length can
be specified, e.g., $E((a > 0); start = 5 \wedge len > 3;)$ states $a > 0$ holds during an
anonymous episode starting from time point 5 and lasting for more than 3 units.
Episodes can be related to each other using Allen's interval algebra [1], e.g., X
before Y means episode X ends some time before episode Y begins. As primitive

properties we consider increasing, decreasing, peaking, stability and oscillation of a variable as well as being in a steady state. As we are aiming at stochastic models whose results may be noisy, we define these properties as trends over episodes. For example, an experiment with the Wnt model [3] with an assumed peak of the beta-catenin concentration in a parameter range can be specified (and executed) in SESSL as follows:

```scala
val exp = new Experiment with Observation with Hypotheses {
  model = "file-mlrj:./WntAxinModel.mlrj"
  scan("kAxin_P" <~ range(0.1, 0.1, 5)) // corresponds to kA -> kAp
  scan("kAxin_uP" <~ range(0.01, 0.01, 0.5)) // corresponds to kAp -> kA
  stopTime = 720
  observe("Cell/Nuc/Bcat()") // beta-catenin is observed
  observeAt(range(1, 10, 720))
  // peak between time point 50 and 150 with a value between 0.1 and 2.4
  assume(Pr(P(Peak("Cell/Nuc/Bcat()") = v; t ≤ 50 ∧ t ≥ 150; t) ∧
    v > 0.1 ∧ v < 2.4) >= 0.8)
}
```

Another key feature of the language is a measurement of similarity to other experiment results (e.g., wet-lab data): given a set of timed data points with error bars, a distance metric and an error threshold, it determines whether a trajectory is sufficiently similar to the given data points. With further discussion needed referring to the language's expressiveness and complexity, the proposed constructs are currently tested in cell biological applications.

References

1. J. F. Allen. Maintaining knowledge about temporal intervals. *Communication of ACM*, page 832843, 1983.
2. F. Fages and A. Rizk. On the analysis of numerical data time series in temporal logic. In *Computational Methods in Systems Biology*, 2007.
3. O. Mazemondet, M. John, S. Leye, A. Rolfs, and A. M. Uhrmacher. Elucidating the sources of beta-catenin dynamics in human neural progenitor cells. *Plos One*, 7(8):e42792–e42792, 2012.
4. D. Peng, R. Ewald, and A. M. Uhrmacher. Towards semantic model composition via experiments. In *Proceedings of the 2nd ACM SIGSIM/PADS conference on Principles of advanced discrete simulation*, pages 151–162. ACM, 2014.
5. D. Waltemath, R. Adams, F. Bergmann, et al. Reproducible computational biology experiments with SED-ML - the simulation experiment description markup language. *BMC Systems Biology*, 5:198, 2011.