Machine-Assisted Extraction of Formal Semantics from Domain Specific Semi-Formal Diagrams

Eric Davis[1], Alec Theriault[1], Max Orhai[1], Eddy Westbrook[1], and Ryan Wright[1]

[1]Galois, Inc

Abstract

Analyzing complex systems is a challenging process which requires not only teams of domain experts but often also a multidisciplinary team of data scientists, mathematicians, statisticians, and software engineers in order to support the life cycle of model development, model-based inference, as well as knowledge extraction and synthesis. The models that typically result from this process today are bespoke, lack generalizability, are not performable, lack reusability, and are difficult to synthesize actionable knowledge and policies from their raw outputs. In this paper we describe AMIDOL: the Agile Metamodel Inference using Domain-specific Ontological Languages, a project that aims to reduce the overhead associated with the development, deployment, maintenance, and reuse of models of complex systems. Our technique utilizes a common intermediate representation which is designed to support a number of scientific, physical, social, and hybrid domains by allowing domain experts to define their models in novel ways, using domain specific ontological languages (VDSOLs). The intermediate abstract representation provides formal, executable, meaning to the semi-formal diagrams domain experts normally create, and allow the inference engine to build prognostic queries on associated reward models. AMIDOL binds results from the inference engine to the original ontologies providing more explainability when compared to conventional methods.

1 Introduction

The construction of computational models is an important practice for scientists, engineers, policy makers, and other domain experts as it provides a way to make formal predictions based on a formal expectation of system behavior, and allows practitioners to test hypotheses and explain phenomena in complex systems. The process of building suitable models, however, is a laborious one which requires diverse teams of experts, significant manual effort, and typical results in models with severe limitations, low performability, and little reusability or generalizability. Such models are not only costly to produce, both in time and effort, but are also prone to errors, difficult to verify, and do not utilize best practices in software development.

AMIDOL seeks to improve the problems inherent to modern machine-assisted inference with two high-level goals: 1) improving the ability of domain experts to build and maintain models and 2) improving the explainability and agility of the results of machine-inference. The techniques applied to achieve these goals incorporate abstract intermediate representations, semantic knowledge representation and binding in graph structures, traditional machine learning and model solution techniques, with intuitive model generation giving meaning to the semi-formal diagrams already being generated by domain experts.

Current modeling formalisms are often specified in formal languages which seem arcane and unnatural to domain experts, but have unambigous formal meaning with defined execution semantics. By
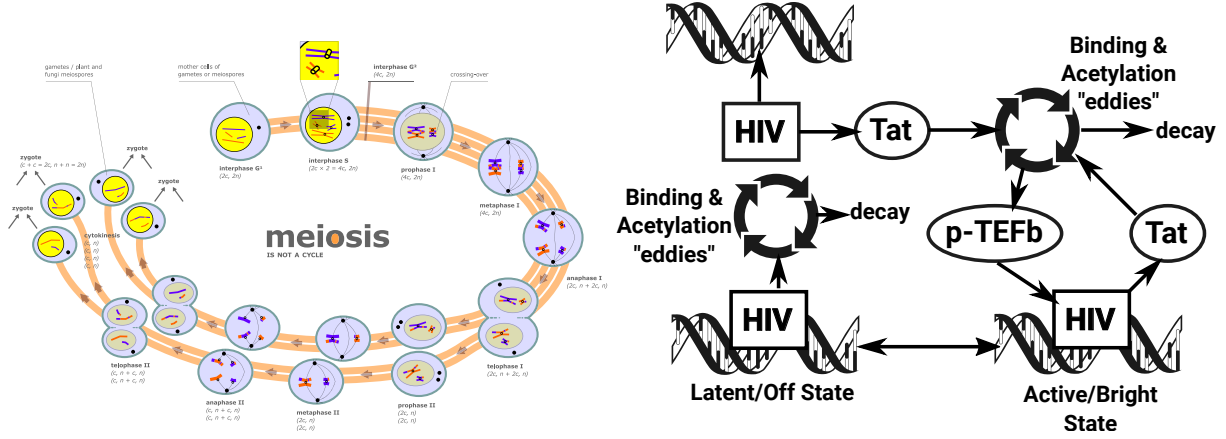
(a) Example of a semi-formal diagram of Meoisis. CC-BY-SA 3.0 Marek Kultys, July 2, 2008.

(b) Example of a semi-formal diagram of the molecular model of the Tat transactivation circuit.

Figure 1: Examples of semi-formal diagrams drawn by domain experts to represent operational semantics and complex system models.

contrast domain experts have naturally developed visual semi-formal ways of describing the systems they study, such as those indicated in Figure 1, but which lack semantic meaning, executability, and are often ambiguous.

$$\frac{d\ tat}{dt} = -dtt \cdot tat \tag{1}$$

$$\frac{d\ nRNA}{dt} = \frac{(b + v \cdot tat)}{k + tat} - ex \cdot nRNA - dr \cdot nRNA \tag{2}$$

$$\frac{d\ cRNA}{dt} = ex \cdot nRNA - dr \cdot cRNA \tag{3}$$

$$\frac{d\ P}{dt} = \frac{vp \cdot cRNA}{kp + cRNA} - dp \cdot P \tag{4}$$

"There should be a concerted programme ... which will require both the development of the appropriate languages to describe information processing in biological systems and the generation of more effective methods to translate biochemical descriptions into the functioning of the logic circuits that underpin biological phenomena."

([14] Paul Nurse (2008))

Abstract machines of systems biology [2]

Significance

2   Related Work

Gene gate modeling in the stochastic pi-calculus [1]
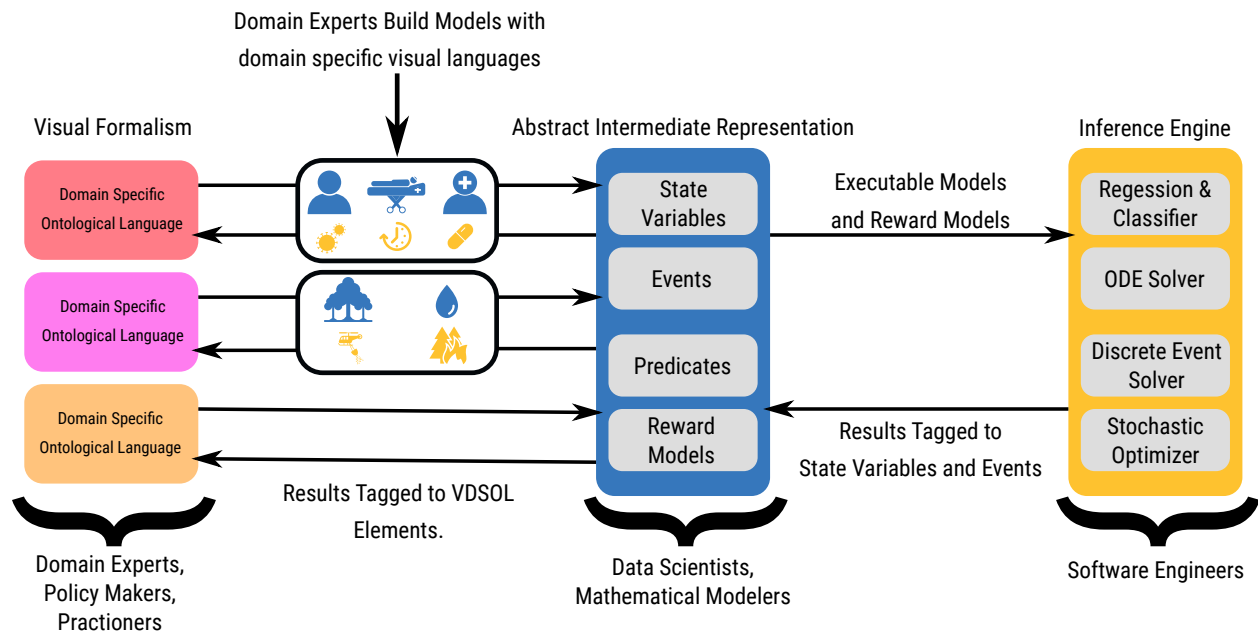
State charts [11]

Pi-calculus [20]

Figure 2: AMIDOL Architecture

Petri-net modeling of biological networks [4]

2.1 Generating Formal Meaning from Informal Diagrams

3 AMIDOL

3.1 Visual Domain Specific Languages

Composition [16, 19]

3.2 Intermediate Representation

Markov models [12]

Petri-nets with inhibitor arcs [5]

Stochastic activity networks [13, 18]

State and reward variables   Reward structures [15, 7, 6, 17]

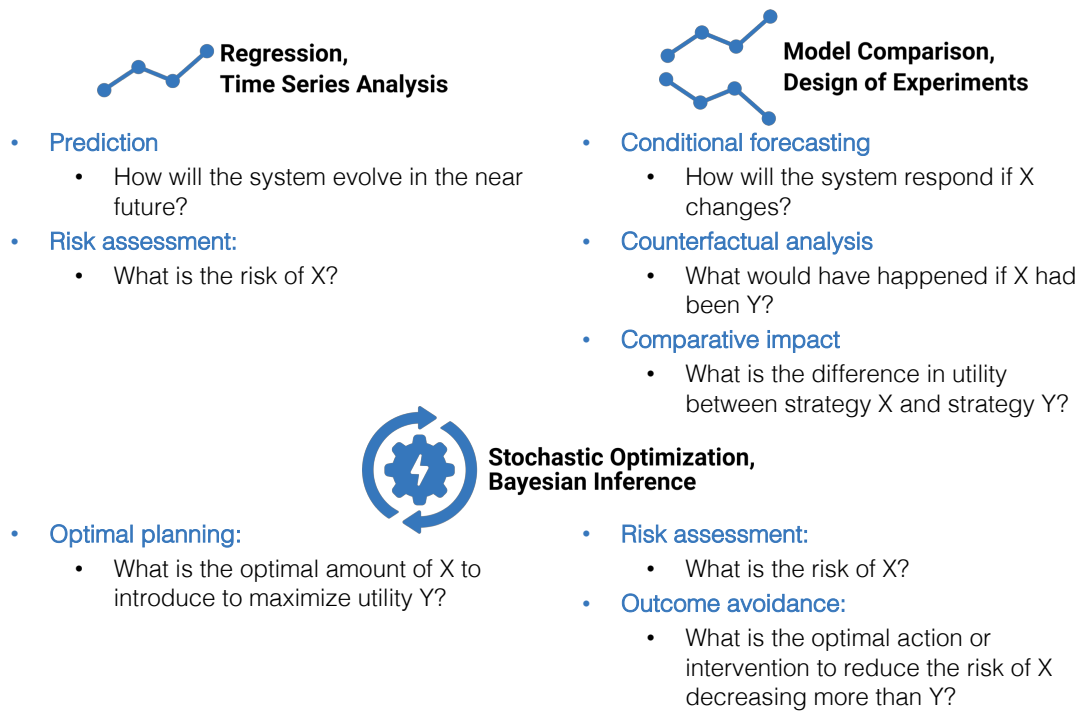Instant of time... [10]

Events

Input and output predicates

**Regression,
Time Series Analysis**

- Prediction
  - How will the system evolve in the near future?
- Risk assessment:
  - What is the risk of X?

**Model Comparison,
Design of Experiments**

- Conditional forecasting
  - How will the system respond if X changes?
- Counterfactual analysis
  - What would have happened if X had been Y?
- Comparative impact
  - What is the difference in utility between strategy X and strategy Y?

**Stochastic Optimization,
Bayesian Inference**

- Optimal planning:
  - What is the optimal amount of X to introduce to maximize utility Y?
- Risk assessment:
  - What is the risk of X?
- Outcome avoidance:
  - What is the optimal action or intervention to reduce the risk of X decreasing more than Y?

Figure 3

3.3   Inference Engine

4   Compartmental Model for Epidemiology

4.1   SIRS Model

H1N1 R_0 importance [9].

Ebola R_0 importance [8]

CDC Data [3]

4.2   Vital Dynamics

5   Conclusions

6   Future Work

7   Acknowledgments

8   Resources, web sites, etc.

MWS seeks to build a community and share resources, so feel free to have a section in your paper that points readers to web sites, github pages, etc.

References

[1] Ralf Blossey, Luca Cardelli, and Andrew Phillips. Compositionality, stochasticity, and cooperativity in dynamic models of gene regulation. *HFSP journal*, 2(1):17–28, 2008.

[2] Luca Cardelli. Abstract machines of systems biology. In *Transactions on Computational Systems Biology III*, pages 145–168. Springer, 2005.

[3] CDC. National, regional, and state level outpatient illness and viral surveillance. https://www.cdc.gov/flu/weekly/fluactivitysurv.htm. Accessed: January 2019.

[4] Claudine Chaouiya. Petri net modelling of biological networks. *Briefings in bioinformatics*, 8(4):210–219, 2007.

[5] Giovanni Chiola, Marco Ajmone Marsan, Gianfranco Balbo, and Gianni Conte. Generalized stochastic petri nets: A definition at the net level and its implications. *IEEE Transactions on software engineering*, 19(2):89–107, 1993.

[6] Gianfranco Ciardo and Robert Zijal. Well-defined stochastic petri nets. In *Modeling, Analysis, and Simulation of Computer and Telecommunication Systems, 1996. MASCOTS'96., Proceedings of the Fourth International Workshop on*, pages 278–284. IEEE, 1996.

[7] Daniel D Deavours and William H Sanders. An efficient well-specified check. In *Petri Nets and Performance Models, 1999. Proceedings. The 8th International Workshop on*, pages 124–133. IEEE, 1999.

[8] David Fisman, Edwin Khoo, and Ashleigh Tuite. Early epidemic dynamics of the west african 2014 ebola outbreak: estimates derived with a simple two-parameter model. *PLoS currents*, 6, 2014.

[9] Christophe Fraser, Christl A Donnelly, Simon Cauchemez, William P Hanage, Maria D Van Kerkhove, T Déirdre Hollingsworth, Jamie Griffin, Rebecca F Baggaley, Helen E Jenkins, Emily J Lyons, et al. Pandemic potential of a strain of influenza a (h1n1): early findings. *science*, 2009.

[10] Roberto Freire. A technique for simulating composed san-based reward models. 1990.

[11] David Harel. Statecharts: A visual formalism for complex systems. *Science of computer programming*, 8(3):231–274, 1987.

[12] Ronald A Howard. *Dynamic probabilistic systems: Markov models*, volume 1. Courier Corporation, 2012.

[13] Ali Movaghar. Performability modeling with stochastic activity networks. 1985.

[14] Paul Nurse. Life, logic and information. *Nature*, 454(7203):424, 2008.

[15] Muhammad A Qureshi, William H Sanders, Aad PA Van Moorsel, and Reinhard German. Algorithms for the generation of state-level representations of stochastic activity networks with general reward structures. *IEEE Transactions on Software Engineering*, 22(9):603–614, 1996.

[16] William H Sanders and Luai M Malhis. Dependability evaluation using composed san-based reward models. *Journal of parallel and distributed computing*, 15(3):238–254, 1992.

[17] William H Sanders and John F Meyer. Reduced base model construction methods for stochastic activity networks. *IEEE Journal on Selected Areas in Communications*, 9(1):25–36, 1991.

[18] William H Sanders and John F Meyer. Stochastic activity networks: Formal definitions and concepts. In *School organized by the European Educational Forum*, pages 315–343. Springer, 2000.

[19] William Harry Sanders. Construction and solution of performability models based on stochastic activity networks. 1988.

[20] Davide Sangiorgi and David Walker. *The pi-calculus: a Theory of Mobile Processes*. Cambridge university press, 2003.