Reconstructing Formal Temporal Logic Models of Cellular Events using the GO Process Ontology

Naren Ramakrishnan, Marco Antoniotti, and Bud Mishra
Department of Computer Science, Virginia Tech, Blacksburg, VA 24061
Courant Institute of Mathematical Sciences, New York University, NY 10012
NYU School of Medicine, New York, NY 10016

ABSTRACT

We present an approach (GOALIE) to use the GO process ontology to reconstruct formal temporal logic models of cellular systems. The reconstructed models are expressed as Kripke structures and support various query, inference, and reasoning operations. This application highlights how the use of an ontology can help describe complex cellular dynamics in the vernacular of propositional temporal logic.

Introduction

The GO process ontology spans a wide range of biological events, from intra-nuclear processes such as DNA transcription, to organism-wide processes such as aging. The traditional use of such a vocabulary is in functional enrichment analysis of gene sets, as a driver for automated annotation of hypothetical proteins, or for model management in biological databases. Such applications essentially exploit only the taxonomical properties (e.g., membership, set containment) of the ontology but do not otherwise use its process-oriented nature to present dynamical perspectives on biological systems. In this paper, we present an approach (GOALIE; Gene Ontology Algorithmic Logic for Invariant Extraction) that uses the GO process ontology to reconstuct formal temporal logic models of cellular events.

The models reconstructed by GOALIE are formally referred to as *Kripke models* in the model checking literature [2]. For our purposes, a Kripke model is simply a directed graph whose nodes encode possible transcriptional states, edges indicate state transitions, and where the nodes are labeled by propositions that hold true in that state. By choosing these propositions from the set of 8517 possible GO process ontology terms, we ensure that any inferences made (e.g., a temporal invariant) on the resulting Kripke structures are interpretable as biologically relevant patterns and hypotheses. For instance, from Fig. 1 we see that *all* state transitions from a state where q is true to a state where r is true *must* pass through a state where p is true. This shows that cell size serves as an effective checkpoint in the transition into the DNA synthesis phase. The biologist can similarly pose other interesting queries about the satisfaction (or refutability) of temporal logic formulae, in the reconstructed model, under given conditions, obtaining affirmative or impossibility answers. Needless to say, a Kripke model is a powerful mechanism to reason about process happenings in a biological context.

GOALIE

We recover Kripke structures by utilizing the GO process ontology in conjunction with time course microarray datasets. We define the states of the Kripke structures as clusters obtained by partitioning (e.g., by a k-means algorithm) overlapping time windows of the time course dataset. These clusters are then *labeled* with the GO process ontology term using an empirical Bayes approach. Labeled clusters are then "chased" to yield transitions to clusters in neighboring time windows. The basis for relating clusters across time windows is the commonality of labelings as revealed by the previous step. The above stages are then repeated, as necessary, in an iterative fashion to refine the initial clusterings (in response to the identified state transitions) or to adjust the transitions (to reflect new cluster assignments). Since the propositions are taken from a controlled vocabulary, we can combine these propositions to create formulate in a propositional temporal logic (CTL), useful in describing complex cellular dynamics. For more details, see [1].

EXPERIMENTAL RESULTS

Fig. 2 depicts a screenshot of the GOALIE software for use in reconstructing a temporal logic model of cell cycle regulation in *S. cerevisiae* (dataset of [3]). GOALIE allows the user to interatively explore chains of GO labelings across time windows and track the validity of temporal formulae, to see if they change state. The system provides hyperlinks to external websites (e.g., related to definitions of GO categories, public repositories of experimental datasets) as well as visualization and query interfaces. Fig. 1 (right) shows the Kripke structure itself; the correspondence with the idealized diagram of Fig. 1 (left) is readily seen.

GOALIE is now being employed in many different case studies, including studying host-pathogen interactions, the dynamics of cancer progression, and the life cycle of the malaria parasite. We are building fast inference algorithms to answer interesting biological queries over large Kripke structures. Our aim is to develop GOALIE into a general framework for reasoning in any suitable vocabulary, not just of temporal processes as done here, but also other multi-modal logics that can encompass richer abstractions of space, control, and variation.

References

[1] M. Antoniotti, N. Ramakrishnan, D. Kumar, M. Spivak, and B. Mishra, Remembrance of Experiments Past: Analyzing Time Course Datasets to Discover Complex Temporal Invariants, *Technical Report TR2005-858*, New York University, Feb 2005.

[2] E.M. Clarke Jr., O. Grumberg, and D.A. Peled, Model Checking, MIT Press, Jan 2000.

[3] P.T. Spellman, G. Sherlock, M.Q. Zhang, V.R. Iyer, K. Anders, M.B. Eisen, P.O. Brown, D. Botstein, and B. Futcher, Comprehensive Identification of Cell Cycle Regulated Genes of the Yeast *Saccharomyces Cerevisiae* by Microarray Hybridization, *Molecular Biology of the Cell*, 9(12), pages 3273-3297, Dec 1998.

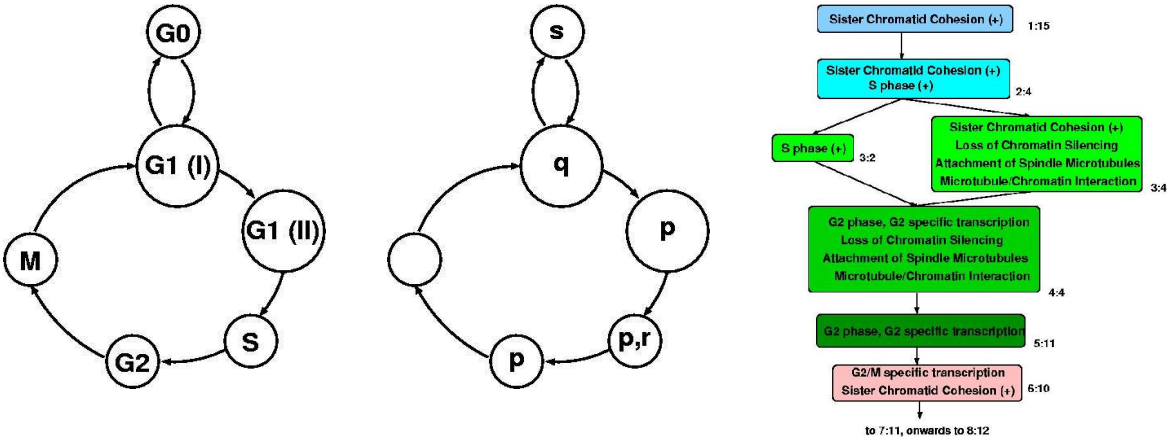

Fig. 1(left) State transition diagram depicting key stages of cell cycle regulation in *S. cerevisiae*. The nodes are labeled with the names of the stages – M(itosis), Gaps, and S(ynthesis). (middle) Kripke diagram of cell cycle regulation obtained manually. States in a Kripke diagram are labeled by the propositions that hold in them. Here, the propositions *p, q, r* and *s* denote "cell size large enough for division," "cytokinesis takes place," "DNA replication takes place," and "cell is in quiescence." For ease of illustration, not all states are labeled. (right) Kripke diagram of cell cycle regulation, obtained automatically by GOALIE. The nodes are identified by cluster numbers (arbitrary) in given time course windows and labeled by GO process ontology terms.

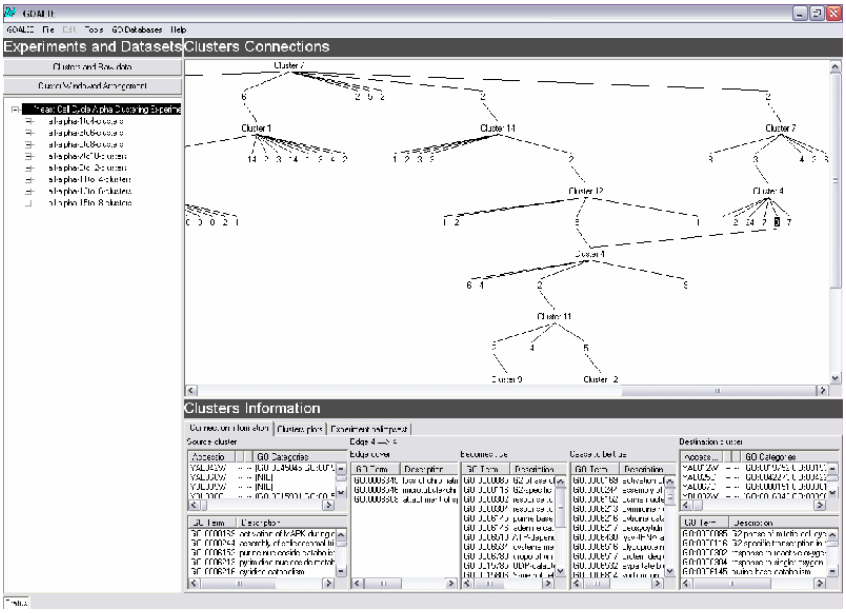

Fig. 2: A screesnshot of GOALIE.The left part shows the various time slices utilized in this study. The top right displays a snapshot of interactive exploration and chasing of clusters. The bottom right part identifies propositions that remain true when going from a source cluster to a destination cluster as well as propositions that become true and those that cease to be true. Notice that cluster 7 in the first time window has been "chased" to yield a chain through successive time windows (clusters 7, 4, 4, 11, and 12 respectively). The links between clusters are labeled with the cardinality of GO terms in common. For instance, the first edge in this chain involves 2 common GO terms, the second involves 3 common GO terms, and so on.