Composition of Signaling Pathway Models and its Application to Parameter Estimation

Geoffrey Koh[1], David Hsu[2], P.S Thiagarajan[2]

1 Introduction

The functioning of biological pathways depends on the interactions among their constituent elements: genes, proteins, and other molecular species. To gain a systems-level understanding of these complex pathways, we need quantitative models that capture the evolution of such interactions over time. Our focus here is on constructing and, in particular, *composing* dynamic models of signaling pathways.

A biological pathway can be viewed as a network of biochemical reactions. To build a pathway model, we need both the network structure and the parameters – kinetic rate constants, initial conditions, *etc.* – that govern the individual biochemical reactions. Parameter estimation of a biological pathway model is a challenging problem, due to the high-dimensional search space involved and the lack of accurate data. Furthermore, model construction is an incremental process, due to new players being discovered and additional experimental data on the known players of the pathway becoming available. It is thus important to develop methods for building pathway models that can be easily *refined* and *expanded*.

Conventional parameter estimation algorithms [5] fit pathway parameters to *all* available experimental data. When new data becomes available, the entire procedure is repeated afresh, using both the new and the old data. This wastes significant computation time. More importantly, the old data may not be systematically archived and easily accessible.

We propose to use a probabilistic model known as *factor graphs* [3] to address the above issues. By capturing the local interactions, the factor graph model drastically reduces the search space for parameter estimation. Being a probabilistic model, it also naturally handles noise in the data. Most importantly, it contains multiple parameter estimates encoded as probability distributions rather than a single best estimate. In addition, new experimental data and pathway players can be integrated into the factor graph incrementally.

Both model refinement and expansion rely on a probabilistic inference technique called *belief propagation* [6]. Using this technique, one can propagate local constraints through the entire network and obtain a globally consistent model. Factor graphs have been used to model biological systems [1], but in this earlier work, the main goal is to study the functional correlations among the elements in the pathway rather than the dynamics.

2 Factor Graph Models of Pathway Dynamics

A signaling pathway is a network of biochemical reactions where the reactions are often mediated by enzymes. The dynamics of the pathway is described by a system of ordinary differential equations (ODEs). The ith equation has the form $\dot{\mathbf{x}}_i = f_i(\mathbf{x}(t), \mathbf{p})$, where $\mathbf{x}(t)$ is a vector-valued function describing the concentration levels of molecular species at time t and $\mathbf{p}$ is the set of pathway parameters.

We build a factor graph model for a given system of ODEs. A factor graph is an undirected bipartite graph consisting of *variable nodes* and *factor nodes*. Each variable node corresponds to an unknown parameter or enzyme, and each factor node corresponds to the ODE. The edges of a factor graph represent the dependencies of the reaction rates on the parameter values and the enzyme concentration levels.

We represent each parameter as a probability distribution and associate it with a variable node of the factor graph. For completely unknown parameters, their initial distributions are assumed to be uniform. Other parameters have *a priori* distributions that reflect prior knowledge. These distributions are updated as new data becomes available. Each factor node is associated with a joint probability distribution that captures the dependencies of the factor node on the variable nodes, as specified in the ODEs. We build this distribution by *sampling* the values of the parameters corresponding to the variable nodes involved. For each set of sampled parameter values, we simulate the system of ODEs and get a score that is the weighted mean squared difference between the simulated and experimental time-series data. The scores are then normalized to obtain a probability distribution.

[1]NUS Graduate School for Integrative Sciences and Engineering. E-mail: g0306427@nus.edu.sg
[2]Department of Computer Science, National University of Singapore. E-mail: {dyhsu, thiagu}@comp.nus.edu.sg

3 Pathway Composition and Data Integration

Pathway components can arise in several ways. For instance, in our earlier work [4], we tackled the parameter estimation problem for large pathway models by decomposing them into smaller components. Multiple components can also arise when different pathways – elucidated independently – are linked together. In either case, each pathway component can be represented by its own factor graph. Composing the components then involves "fusing" the corresponding factor graphs at their common variable nodes to form a composite factor graph.

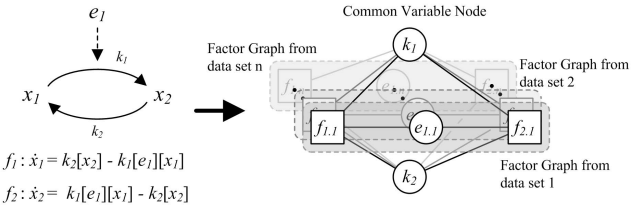

Figure 1: An enzyme mediated reaction and its factor graph representation. A factor graph is constructed for each new dataset, and they are fused at their common variable nodes (k_1 and k_2) to form a composite factor graph.

Similarly, we can integrate new data into an existing pathway model represented as a factor graph. We first sample the part of the pathway relevant to the new data and build a new factor graph for this part. We then combine the new and the existing factor graphs by fusing their common variable nodes. This idea is illustrated in Figure 1.

One key issue in composition is to ensure that the local dependencies and constraints in the components are all captured in the composite factor graph and that they are consistent. To achieve this, we use belief propagation to propagate local constraints globally [6]. Upon convergence, the variable nodes of the factor graph contains the *maximum a posteriori* distributions of the parameters.

4 Results and Discussion

We tested this approach on a simplified model of the Akt-MAPK signaling pathway [4]. Using four sets of experimental data synthesized on the Akt-MAPK model through simulation, we performed parameter estimation on the model incrementally by adding one data set at a time and applying our composition method. For comparison, we applied two other methods implemented in the modeling software COPASI [2]. All the methods were allocated equal amounts of time for the four data sets. Preliminary results (Figure 2) indicate that our method achieved substantially better estimates.

	BP	SRES	GA
1 Dataset	**0.412**	0.483	61.96
2 Datasets	**1.548**	1.356	17.38
3 Datasets	**1.250**	3.020	263.55
4 Datasets	**0.203**	2.040	46.76

Figure 2: Performance comparison of three methods on parameter estimation. BP is our method based on belief propagation. SRES and GA are two methods based on evolutionary strategies with stochastic ranking and genetic algorithms, respectively. The scores are the weighted mean squared difference between simulated and experimental data. Smaller scores are better.

We are currently extending this work in two directions. Recent experimental developments suggest that cross-talks are common between signaling pathways. By systematically composing pathway models, we plan to construct large signaling pathway models that take into account cross-talks between the individual pathway components. Second, we plan to improve the sampling process for building the joint distributions associated with the factor nodes. Currently we sample uniformly over the entire local parameter space. A "guided sampling" approach can improve the results by focusing on the more promising regions of the space.

References

[1] Gat-Viks, I., Tanay, A., Raijman, D. and Shamir, R. 2005 *The factor graph network model for biological systems.* Proceedings of Research in Computational Molecular Biology, LNCS 3500 pp 31–47.

[2] Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N., Singhal, M., Xu, L., Mendes, P. and Kummer, U. 2006. *COPASI - a COmplex PAthway SImulator.* Bioinformatics 22(24) pp 3067–3074.

[3] Kschischang, F.R., Frey, B.J. and Loeliger, H.A. 2001. *Factor graphs and the sum-product algorithm.* IEEE Transactions on Information Theory 42(2) pp 498–519.

[4] Koh, G., Teong, H.F.C., Clèment, M.V., Hsu, D. and Thiagarajan, P.S. 2006. *A decompositional approach to parameter estimation in pathway modeling.* Bioinformatics 22(14) pp e271–e280.

[5] Moles, C.G., Mendes, P. and Banga, J.R. 2003. *Parameter estimation in biochemical pathways: A comparison of global optimization methods.* Genome Research 13(11) pp 2467–2474.

[6] Yedidia, J.S., Freeman, W.T. and Weiss, Y. 2002. *Understanding belief propagation and its generalizations.* Exploring Artificial Intelligence in the New Millenium pp 239–269