

383d Inferring Transcription Factor Activities with a Probabilistic Framework of Hidden Markov Model

Zheng Li, Stephen M. Shaw, Matthew J. Yedwabnick, and Christina Chan

Recent advances in high throughput technologies have generated a tremendous amount of biological information, such as gene expression, protein-protein interaction, and metabolic data. These various types of data capture different levels of cellular response to environmental factors and contain within it information about the underlying regulatory network structure. Coupling genome-wide protein-DNA and protein-protein interaction information with gene expression data permits the reconstruction of gene regulatory [1] and signal networks [2]. In a gene regulatory network, genes are typically regulated by transcription factors. The activity of transcription factors, i.e. the fraction of transcription factor that binds to DNA is more difficult to measure as compared to gene expression. Models such as Network component analysis (NCA) have been applied to extract this information from expression data [3,4]. However, these approaches are limited by the types of network motifs that may be incorporated in the analysis. In order to overcome this limitation, we present a Hidden Markov Model (HMM) that represents gene networks with many different regulatory motifs, such as feed-back and autoregulation, making HMM a useful complement to existing approaches. In this paper, a gene regulatory network incorporating motifs such as feed-forward, auto-regulation, and multiple input was constructed as well as simulated with an HMM model. The simulated gene expression data was used to infer the transcription factor activities. The ability of HMM to infer transcription factor activities was evaluated by comparing the inferred and simulated transcription factor activity profiles. Second, HMM was applied to gene expression data obtained from *E. Coli* and *Saccharomyces cerevisiae* cell cycle. The results were compared to that of NCA and further validated based upon the binding motifs of the transcription factors. In summary, the HMM model provides a probabilistic framework to simulate gene regulatory networks and to infer activity profiles of hidden variables.

References:

1. Lee, T.I., et al., (2002) Transcriptional Regulatory Networks in *Saccharomyces cerevisiae*. *Science*, 298(5594): 799-804.
2. Steffen, M., et al., Automated modelling of signal transduction networks. *BMC Bioinformatics* [online computer file], 2002. 3: p. No pp given.
3. Kao, K.C., Yang, Y.L., Boscolo, R., Sabatti, C., Roychowdhury, V.P., Liao, J.C., (2004) Transcriptome-based determination of multiple transcription regulator activities in *Escherichia coli* by using network component analysis. *PNAS*, 101(2): 641-646.
4. Liao, J.C., Boscolo, R., Yang, Y.L., Tran, L.M., Sabatti, C., Roychowdhury, V.P., (2003), Network component analysis: Reconstruction of regulatory signals in biological systems. *PNAS*, 100(26):15522-15527.