A Probablistic Model for Characterizing Regulatory Targets of Transcription Factors from ChIP-Seq and ChIP-chip Binding Profiles

Chao Cheng
Renqiang Min
Mark Gerstein
To be submitted to Bioinformatics very soon

A Probablistic Model for Characterizing Regulatory Targets of Transcription Factors from ChIP-Seq and ChIP-chip Binding Profiles

- Suppose that we have a set of genes $\{g\}$, a set of Transcription Factors (TF) $\{t\}$, and the binding profile of each TF t on each gene g , that is, S_g^t , we want to calculate the posterior probability of TF *t*'s targeting gene *g* ,

$$p(t_g = 1 \mid S_g^t) = \frac{p(t_g = 1, S_g^t)}{p(S_g^t)} \propto p(t_g = 1, S_g^t) \approx \Sigma_{i=1}^{i=n} p(t_{gi} = 1, S_{gi}^t)$$

$$= \Sigma_{i=1}^{i=n} p(t_{gi} = 1) p(S_{gi}^t \mid t_{gi} = 1)$$

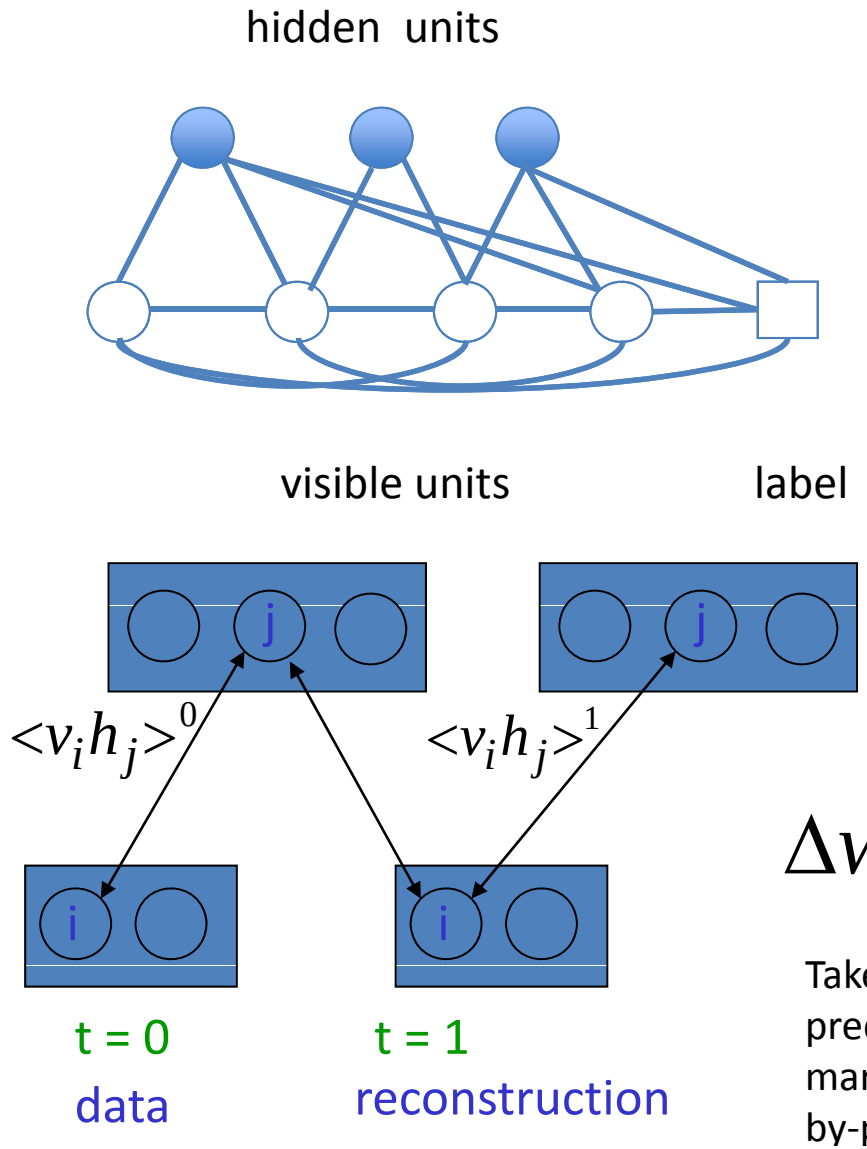
$$\propto \Sigma_{i=1}^{i=n} w_{gi} \times f(S_{gi}) = \Sigma_{i=1}^{i=n} w_i \times f(S_{gi})$$

w_{gi} is the prior prob. that TF ***t*** targets position ***i*** of gene ***g*** , and we assume that all the genes share the same position-specific prior probabilities w_i , which is estimated by Chao by the proportion of the total amount of reads or peak heights covered at position *i* over all genes

Discriminative Semi-RBMs for Modeling Interactions between Histone Modifications and TF Bindings to Predict Gene Expression

Renqiang Min
Chao Cheng
Mark Gerstein
To be submitted to NIPS 2011 soon

Semi-RBMs with Lateral Connections



$$E(v,h) = -\sum_{i,j} v_i h_j W_{ij} - \sum_{y,j} l_y h_j W_{yj} - \sum_{y} l_y v_i v_{i'} L_{ii'} - \sum_i v_i b_i - \sum_j h_j c_j - \sum_y l_y d_y$$

$$p(v,h) \propto e^{-E(v,h)}$$

$$\Delta w_{ij} = \varepsilon \left(<v_i h_j>^0 - <v_i h_j>^1 \right)$$

Take home message: given expression level, the model predicts that, a small number of highly interacted histone mark features are essential, and all the others are redundant by-products.