Supplementary material for manuscript: "NETGEM: Network Embedded analysis of Temporal Gene Expression Model"

Vinay Jethava

January 21, 2010

This document presents the supplemental material for the manuscript "NETGEM: Network Embedded analysis of Temporal Gene Expression Model".

0.1 Original Model Description

0.1.1 Observation model

The observed gene expression levels, $\mathbf{x}^s(t)$, for an strain s at time t are modeled as an Ising system [Song *et al.*, 2009]:

$$P\left(\mathbf{x}^s(t)|\mathbf{w}^s(t)\right) = \frac{1}{Z(t)} \exp\left(-\sum_{(i,j)\in E} w^s_{(i,j)}(t) x^s_i(t) x^s_j(t)\right) \tag{1}$$

where $Z(t)$ is the normalization constant.

0.1.2 Evolution model

We assume that the weights evolve according to the Markov chain, i.e.,

$$P(\mathbf{w}(t+1) = \mathbf{w}_{t+1}|\mathbf{w}(t) = \mathbf{w}_t) = \mathbf{Q}(\mathbf{w}_t, \mathbf{w}_{t+1}) \tag{2}$$

where $\mathbf{Q}(\mathbf{w}_t, \mathbf{w}_{t+1})$ is the probability of the transition from state $\mathbf{w}_t$ at time t to state $\mathbf{w}_{t+1}$ at time $(t+1)$. In general, if there are S strains present, then each will have corresponding transition probability matrix $\mathbf{Q}_s$.

1 NETGEM model derivation

This section presents the derivation for the inference and parameter learning in the generative model, NETGEM. The inference is done over the hidden variables $\{y_e(t), w_e(t)\}$ based on the Markovian structure of the problem.

We assume that the data for strain, s, is independently generated based on the Ising model with the weights that are damped versions of the weights in the original strain. This leads to the observation model specified as:

$$
\begin{aligned}
o_e^t(l) &= P(x_e^{1:S}(t)|w_e(t) = w_l) && (3)\\
&= \frac{1}{Z}\prod_{s=1}^{S} P(x_e^s(t)|w_e(t) = w_l) && (4)\\
&= \frac{\exp\left\{-w_l(\sum_{s=1}^{S} x_i^s(t)x_j^s(t)\Gamma^s(i,j))\right\}}{\sum_{l=1}^{\mathcal{W}} \exp\left\{-w_l(\sum_{s=1}^{S} x_i^s(t)x_j^s(t)\Gamma^s(i,j))\right\}} && (5)
\end{aligned}
$$

The forward iterates, $f_e^t(l,h)$ and backward iterates, $b_e^t(l,h)$ can be computed as follows:

$$
\begin{aligned}
f_e^t(m,h) &= P(x_e^{1:S}(1:t), w_e^t = w_m, y_e^t = h|\Psi_e^{(n)}) && (6)\\
&= P(x_e^{1:S}(t)|w_m)\sum_{w_l}\sum_{h'}\Big[P(y_e^t = h|\alpha^{(n)}) &&\\
&\times\ P(w_m|w_e^{t-1} = w_l, y_e^{t-1} = h') \times f_e^{t-1}(l,h')\Big] && (7)\\
&= o_e^t(m)\sum_{l=1}^{\mathcal{W}}\sum_{h'=1}^{H} f_e^{t-1}(l,h')\alpha_h^{(n)} q_{h'}^{(n)}(l,m) && (8)\\
b_e^t(m,h) &= P(x_e^{1:S}((t+1):T)|w_e(t) = w_m, y_e^t = h, \Psi_e^{(n)}) && (9)\\
&= \sum_{w_l}\sum_{h'}\Big[P(x_e^{1:S}(t+1)|w_e^{t+1} = w_l)b_e^{t+1}(l,h') &&\\
&\times\ P(w_e^{t+1} = w_l|w_m, y_e^t = h)P(y_e^{t+1} = h'|\alpha^{(n)})\Big] && (10)\\
&= \sum_{m=1}^{\mathcal{W}}\sum_{h'=1}^{H} q_h^{(n)}(m,l)o_e^{t+1}(l)\alpha_{h'}^{(n)} b_e^{t+1}(l,h') && (11)
\end{aligned}
$$

The conditional probability $P(\Omega_e^t = (w_l,h), \Omega_e^{t+1} = (w_m,h')|\mathbf{x}_e^{1:S}(1:T), \Psi^{(n)})$ denoted by $\xi_e^t(l,m,h,h')$ can be computed as

$$
\xi_e^t(l,m,h,h') \propto f_e^t(l,h)\alpha_{h'}^{(n)} q_h^{(n)}(l,m)o_e^{t+1}(m)b_e^{t+1}(m,h') \quad (12)
$$

The likelihood term, $\mathcal{L}(\Psi;\Psi^{(n)})$, in (**??**) can be expressed in terms of the conditioned edge probabilities, ξ_e^t, in (12) as

$$\mathcal{L}(\Psi;\Psi^{(n)}) = \sum_{e\in E}\sum_{t=1}^{T-1}\mathbf{E}_{\xi_e^t}[\ln q_h(l,m) + \ln \alpha_{e,h'}] \tag{13}$$

subject to the constraints

$$\sum_m q_h(l,m) = 1 \quad \forall h \tag{14}$$

$$\sum_h \alpha_{e,h} = 1 \quad \forall e \tag{15}$$

2 Discussion on NETGEM vs Naive HMM

This section presents a study of the naive HMM vs the NETGEM model.

One can model the dynamics of the simple model defined by (1) and (2), using a simple HMM. The quantity to be estimated is the transition probability $\mathbf{Q}$. However, the exponential state space makes such an approach impractical.

For the purpose of study, we explore the relation between our model and the simple HMM in the following two sections. We also present a comparison of the results obtained using the independent weights evolution assumption vs the results using a standard HMM implementation in section **??**.

2.0.3 Evaluation of independent weights dynamics

We now discuss the relationship between the results obtained using the independent weights evolution assumption and the original problem.

Lemma 2.1 *The independently evolving weights assumption gives a rank-1 tensor approximation [Horn & Johnson, 1990] to the global weight transition probability.*

For example, if the interaction network consists of two edges having transition probabilities $A = \{a_{ij}\}$ and B respectively, the approximation to the original matrix, $\hat{Q}$ is given as the Kronecker product

$$\hat{Q} = A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B \\ a_{21}B & a_{22}B \end{bmatrix}$$

We note that if the weights indeed evolve independently of each other and we are given the edge transition probabilities Q_e, the overall system transition probability, $\mathbf{Q}$, is given as

$$\mathbf{Q} = Q_1 \otimes Q_2 \otimes \ldots \otimes Q_E$$

In general, if there are E edges with estimated transition probability matrices, $\hat{Q}_1, \ldots, \hat{Q}_E$, we obtain the approximation, $\hat{\mathbf{Q}}$

$$\hat{\mathbf{Q}} \simeq Q_1 \otimes Q_2 \otimes \ldots \otimes Q_E$$

We discuss the quality of approximation in case the original probability matrix Q is a higher order tensor elsewhere.

2.0.4 Number of parameters to be learnt

We note that the generative model, NETGEM, does an order reduction the number of parameters to be learnt. Specifically, if there are E edges in the network, with each edge taking on W denotes the number of possible interaction strengths, and H is the number of functional classesfrom $O(W^{2*E})$ to $O(HW^2 + HE)$, where . Typical values are $E \sim 6000$, $H \sim 260$ and $W \sim 5$.

3 Independent weights dynamics

This section presents a derivation of the forward and backward probabilities for the independent weights evolution model in the presence of multiple strains. The observation model, $o_e^t(l) = P(x_e^{1:S}(t)|w_e^t = w_l)$, remains unchanged as in (3)-(5).

The update equations for computing the forward and backward probability distributions are given as:

$$\begin{aligned}
f_e^{t+1}(m) &= P(x_e^{1:S}(1:t), w_e(t+1) = w_m | Q_e^{(n)}) && (16)\\
&= o_e^{t+1}(m) \sum_{l=1}^{\mathcal{W}} f_e^t(l) q_e^{(n)}(l,m) && (17)\\
b_e^t(l) &= P(x_e^{1:S}((t+1):T) | w_e(t) = w_l, Q_e^{(n)}) && (18)\\
&= \sum_{m=1}^{\mathcal{W}} q_e^{(n)}(l,m) o_e^{t+1}(m) b_e^{t+1}(m) && (19)
\end{aligned}$$

and the joint probability as

$$\begin{aligned}
\xi_e^t(l,m) &= P(w_e(t,t+1) = (w_l, w_m) | x_e^{1:S}(1:T), Q_e^{(n)}) && (20)\\
&\propto f_e^t(l) q_e^{(n)}(l,m) o_e^{t+1}(m) b^{t+1}(m) && (21)
\end{aligned}$$

We use the independent weights dynamics where the parameter to be learnt is the transition matrix Q_e for each edge, $e = (i, j) \in E$. We solve the Expectation Maximization (EM)[Dempster *et al.*, 1977] problem for each edge, e,

$$\begin{array}{llcl} \text{E-step:} & \mathcal{L}(Q_e; Q_e^{(n)}) & = & E_{w_e^{|}}[\ln P(\mathbf{x}_e^{1:S}(1:T), \mathbf{w}_e(1:T)|Q_e)] \\ \text{M-step:} & \hat{Q}_e^{(n+1)} & = & \arg\max_{Q_e}(\ln P(Q_e) + \mathcal{L}(Q_e; Q_e^{(n)})) \end{array} \tag{22}$$

where $W_e^{|}$ is the conditioned variable, $w_e(1:T)|\mathbf{x}_e^{1:S}(1:T), Q_e^{(n)}$ and $Q_e^{(n)}$ is the MAP estimate for the transition probability, Q_e, at the n^{th} iteration of the algorithm.

This leads to the update equation for the MAP estimate for transition probabilities, $q(l, m)$, obtained by the maximization step in (22) as

$$q_e^{(n+1)}(l, m) = \frac{(\theta_{lm} - 1) + \sum_{t=1}^{T-1} \xi_e^t(l, m)}{\sum_m (\theta_{lm} - 1) + \sum_{t=1}^{T-1} \sum_{m=1}^{\mathcal{W}} \xi_e^t(l, m)} \tag{23}$$

References

[Dempster *et al.*, 1977] Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977) Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological),* **39** (1), 1–38.

[Horn & Johnson, 1990] Horn, R. A. & Johnson, C. R. (1990) *Matrix Analysis.* Cambridge University Press.

[Song *et al.*, 2009] Song, L., Kolar, M. & Xing, E. P. (2009) Keller: estimating time-varying interactions between genes. *Bioinformatics,* **25** (12).