A Comparison of Inference Methods for Sparse Factor Analysis Models

Oliver Stegle
Cavendish Laboratories
University of Cambridge
os252@cam.ac.uk

Kevin Sharp
School of Computer Science
University of Manchester
sharpk@cs.man.ac.uk

Magnus Rattray
School of Computer Science
University of Manchester
magnus.rattray@manchester.ac.uk

John Winn
Microsoft Research
Cambridge, UK
jwinn@microsoft.com

Factor analysis is a general purpose technique for dimensionality reduction with applications in diverse areas including computer vision, collaborative filtering and computational biology. Sparse factor analysis is a natural extension that can be motivated by the observation that sparse features tend to generalize better, or justified based on *a priori* beliefs about the underlying generative model of the observed data. For example, information processing in cells is mediated by sparsely connected gene regulatory networks. An important aspect of this Systems Biology application, which we wish to exploit, is that one can include available prior knowledge about the connectivity structure of the network. Such knowledge determines how likely it is *a priori* that a latent factor, in this case a transcription factor protein, regulates the expression of a specific target gene.

In this work we consider a sparse factor analysis model with an informative connectivity prior which expands on our previous work on Factor models applied to expression modelling [1] and the inference of transcription factor concentration [2]. We compare four approximate Bayesian inference methods, including sampling and deterministic methods. These include a novel hybrid Expectation Propagation/Variational method which achieves encouraging results particularly when considering the accuracy/efficiency trade-off.

Sparse Factor Analysis

The factor analysis model is described by a matrix product of unobserved factor loadings $\mathbf{W}$ and factor activations $\mathbf{X}$ in addition to a mean contribution $\boldsymbol{\mu}$ and random noise Ψ

$$\mathbf{Y} = \boldsymbol{\mu} + \mathbf{W} \cdot \mathbf{X} + \Psi. \tag{1}$$

A mixture prior on the entries of $\mathbf{W}$ is used to achieve sparseness. Conditioning on indicator variables $c_{g,k}$ we obtain

$$\begin{aligned} p(w_{g,k}|c_{g,k} = \text{false}) =& \mathcal{N}(w_{g,k}|0, \sigma_s) \\ p(w_{g,k}|c_{g,k} = \text{true}) =& \mathcal{N}(w_{g,k}|0, 1), \end{aligned} \tag{2}$$

where σ_s is much smaller than 1 hence forcing the corresponding weight to zero. Existing knowledge about the network structure is encoded as a Bernoulli prior over the indicator variables $c_{g,k}$

$$P_{g,k} = p(c_{g,k} = \text{true}) = \begin{cases} \pi_0 & \text{no link} \\ 1 - \pi_1 & \text{link} \end{cases}, \tag{3}$$

where π_0 denotes the false positive rate (FPR) and π_1 the false negative rate (FNR) respectively. The Bayesian network corresponding to this model is depicted in Figure 1.

We compare four approximate Bayesian inference methods on this model. Two Gibbs samplers, a standard sampler (Gibbs) and a collapsed sampler (Gibbs collapsed), are contrasted against two deterministic approximations. The first is a Variational approximation (VB). The second approximation is a novel hybrid approach where parts of model are implemented in VB and other using Expectation Propagation (VB/EP).

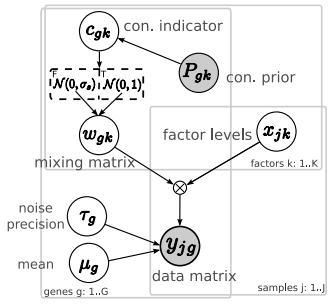

Figure 1: The Bayesian network of the sparse FA model. The indicators $c_{g,k}$ determine the state of the gate either switching the corresponding mixing weight off ($w_{g,k} \sim \mathcal{N}(0, \sigma_s)$) or on ($w_{g,k} \sim \mathcal{N}(0, 1)$). *A priori* knowledge about the connectivity structure is introduced as a prior on $P_{g,k}$.

Results

All four inference methods were compared on simulated and real datasets. On simulated data we evaluated the predictive accuracy of the recovered network structure and factor activations. On real data we evaluated the fill-in performance. As an example, Figure 2 shows the accuracy of the network reconstruction as a function of CPU time. In this small example sampling methods can be run to convergence. Our study demonstrates the utility of sampling algorithms as approximations long before they have reached convergence. Furthermore our results indicate that the new VB/EP hybrid is a promising alternative with run time properties similar to a VB while being significantly more accurate.

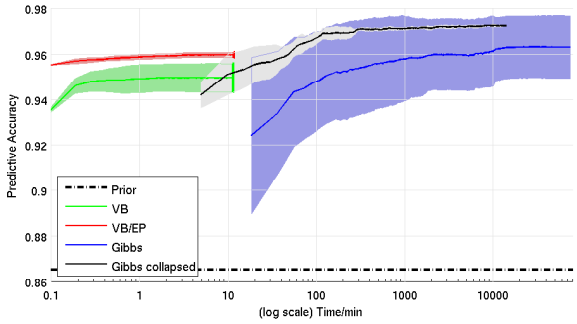

Figure 2: Predictive accuracy of latent network structure $\mathbf{C}$ on a small simulated datasets (543 genes, 26 factors) as function of the CPU time as given by the fraction of correctly predicted links (threshold 0.5). Plots show the mean performance and ± 1 std. dev. error bars (from 5 Gibbs chains and 5 random restarts of the det. methods). As a baseline the accuracy using the links implied the network prior $P_{d,k}$ is included as horizontal line.

References

[1] O. Stegle, A. Kannan, R. Durbin, and J. Winn. Accounting for non-genetic factors improves the power of eqtl studies. In *RECOMB*, pages 411–422, 2008.

[2] G. Sanguinetti, N. D. Lawrence, and M. Rattray. Probabilistic inference of transcription factor concentrations and gene-specific regulatory activities. *Bioinformatics*, 22:2775, 2006.