Accuracy of Variational Estimates for Random Graph Mixture Models

Steven Gazal, Jean-Jacques Daudin, Stéphane Robin

Accuracy of Variational Estimates for Random Graph Mixture Models

Steven Gazal [a,b] & Jean-Jacques Daudin [a,b] & Stéphane Robin [a,b]

[a] *AgroParisTech, UMR 518, F-75005, Paris,* France
[b] *INRA, UMR 518, F-75005, Paris,* France

L'analyse des réseaux exerce depuis quelques années un attrait croissant. Les données qui sont sous la forme de mesures de relations entre items sont de plus en plus disponibles, et abandonnent la structure usuelle d'un jeu de données de type individus-variables pour une structure de type individus-individus. Ces données "relationnelles" sont très souvent présentées sous la forme d'un graphe, même si cette représentation a ses limites, notamment quand le nombre d'individus dépasse la centaine. La représentation graphique des données des réseaux est alors attractive, mais nécessite un modèle synthétique.
Le modèle de graphe le plus ancien et le plus utilisé est le modèle de mélange d'Erdös-Rényi. C'est un modèle simple dont les propriétés moyennes ou asymptotiques sont connues : distribution des degrés, densité du graphe, coefficient d'agrégation, diamètre moyen... L'écriture littérale de la vraisemblance de ce modèle est très simple, mais son temps de calcul croit de façon exponentielle avec le nombre d'individu. Une utilisation des algorithmes d'estimation usuels comme E-M n'est pas envisageable. Une approche variationnelle a été utilisé comme alternative pour implémenter un algorithme d'estimation des paramètres du modèle, et cela pour des réseaux de très grande taille (Daudin & *al* (2008)).

La méthode variationnelle est une technique d'optimisation qui permet de trouver le maximum d'une fonction en optimisant une borne inférieure (Jaakola (2000)). Les solutions de ces problèmes sont souvent des équations de point fixe.
Les propriétés statistiques des estimateurs produits par cette approche sont cependant mal connues. Gunawardana et Byrne (2005) ont prouvé que l'algorithme variationnel converge vers une solution qui minimise la divergence, mais qui n'est pas un point stationnaire de la vraisemblance, sauf pour les modèles dégénérés. McGrory et Titterington (2007) et (2009) ont étudié les propriétés des estimateurs variationnels en terme de séléction de modèles pour des mélanges gaussiens et des modèles de chaine de Markov cachée. Hormis ces travaux, nous n'avons pas d'autres informations sur les propriétés statistiques des estimateurs variationnels.

Le but de nos travaux est donc d'étudier par simulation la convergence de différents estimateurs variationnels. Seront étudiés l'estimateur variationnel classique (VEM), l'estimateur de la Belief Propagation (BP) (Yedidia & *al* (2003)), ainsi qu'une extension au modèle bayesien (VB) (Beal & Ghahramani (2003)) où les paramètres sont considérés comme des variables cachées.

Tout d'abord nous avons simuler des réseaux aux dimensions "restreintes" pour pouvoir comparer la qualité et la précision des estimateurs EM et variationnels. Puis nous avons fait varier la taille des graphes pour étudier la convergence des estimateurs variationnels uniquement.
Les résultats montrent une bonne convergence des estimateurs variationnels, et une qualité très proche de celle des estimateurs EM.
Nous avons également prouvé théoriquement la consistance de l'estimateur variationnel. La démonstration ne sera pas montré mais les idées principales seront avancées.
Enfin nous illustrerons nos différents estimateurs variationnels VEM et VB sur le réseau de régulation d'E. Coli.

Complex networks are more and more studied in different domains such as social sciences and biology. The network representation of the data is graphically attractive, but there is clearly a need for a synthetic model, giving a lightning representation of complex networks. Statistical methods have been developed for analyzing complex data such as networks in a way that could reveal underlying data patterns through some form of classification.
The Erdös-Rényi's mixture model is a network model used a lot. It is very simple and its properties are well known. The likelihood can be written easily, but the big size of the graph forbids the use of the traditional EM algorithm. The dependency of all the nodes implies that the E step explores all the configurations of a graph. It is too complex to compute it when the number of nodes is high. A variational method has been used for estimating the parameters in a reasonable time (Daudin & *al* (2008)).

The variational methods refer to a large collection of optimization techniques (Jaakola (2000)). It consists on replacing a complex likelihood by a lower bound of the likelihood that is simpler to compute. In EM algorithm, the estimators that maximize this lower bound have unknown statistical properties. Gunawardana & Byrne (2005) claim that variational methods converge to solutions that minimize divergence, but these are not necessarily stationary points in likelihood. They only converge to stationary points in likelihood in degenerate cases. McGrory and Titterington (2007) and (2009) studied the properties of variational estimates in terms of model selection for Gaussian mixtures and hidden Markov models. Despite these works and others, we still do not have an overall picture of the statistical properties of variational estimates.

We present several variational estimates of the model parameters, and compare their accuracy through a simulation study. We study the variational estimator (VEM), the Belief Propagation estimate (BP) (Yedidia & *al* (2003)) and the variational approach will be extended to the Bayesian setting, the parameters being considered as unobserved

variables (VB) (Beal and Ghahramani (2003)).
First we simulated small networks in order to be able to calculate the EM estimate and to compare its quality with the variational estimators. Then we increased the size of our networks to study the convergence of the variational estimates only.
The results show a good convergence of the variational estimates and a good quality compared to the EM estimate.
We also proved the consistance of the variational estimate. The proof will not be explained but the main ideas will be shown.
We finally illustrate the differences between the variational estimates VEM and VB on the regulatory network of E. coli.

Mots-clés: Modèles graphiques - Méthodes bayésiennes

Bibliographie

[1]Balazsi, G., Barabasi, A.-L., Oltvai, Z. (2005) Topological units of environmental signal processing in the transcriptional network of escherichia coli. *PNAS*, 102(22), 7841-7846.
[2]Beal, M. J., Ghahramani, Z. (2003) The variational Bayesian EM algorithm for incomplete data: with application to scoring graphical model structures. *Bayesian Statistics 7*, Oxford University Press, pp. 543-52.
[3]Besag, J. (1986) On the statistical analysis of dirty pictures. *J. R. Statist. Soc. B*, 48 (3), 259-302.
[4]Daudin, J.-J., Picard, F., Robin, S. (2008) A mixture model for random graphs. *Stat Comput*, 18, 173-183.
[5] Dempster, A. P., Laird, N. M., Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *J. R. Statist. Soc. B*, 39, 1-38.
[6] Gunawardana, A., Byrne, W. (2005) Convergence theorems for generalized alternating minimization procedures. *J. Machine Learn. Res.*, 6, 2049-2073.
[7] Jaakola, T. S. (2000) Tutorial on variational approximation methods. *MIT Press.*
[8] Jordan, M. I., Ghahramani, Z., Jaakkola, T., Saul, L. K. (1999) An introduction to variational methods for graphical models. *Machine Learning*, 37 (2), 183-233.
[9] Latouche, P., Birmel, E., Ambroise, C. (2008) Bayesian methods for graph clustering. *SSB Research Report 17.*
[10] MacKay, D. J. (2003) Information Theory, Inference, and Learning Algorithms. *Cambridge Univer- sity Press.*
[11] McGrory, C., A., Titterington, D., M. (2009) Variational Bayesian analysis for hidden Markov models. *Austr. & New Zeal. J. Statist.*, 51 (2), 227-44.
[12] McGrory, C. A., Titterington, D. M. (2007) Variational approximations in Bayesian model selection for finite mixture distributions. *Comput. Statist. and Data Analysis*, 51, 5332-67.
[13] McLachlan, G., Peel, D. (2000) *Finite Mixture Models.* Wiley.

[14] Nowicki, K., Snijders, T. (2001) Estimation and prediction for stochastic block-structures. *J. Amer. Statist. Assoc.*, 96, 1077-87.
[15] Pattison, P. E., Robins, G. L. (2007) Handbook of Probability Theory with Applications. Sage Publication, Ch. Probabilistic Network Theory.
[16] Picard, F., Miele, V., Daudin, J. J., Cottret, L., Robin, S. (2009) Deciphering the connectivity structure of biological networks using MixNet. *BMC Bioinformatics* , 10.
[17] Shen-Orr., R., M., S., M., U., A. (2002) Network motifs in the transcriptional regulation network of escherichia coli. *Nature genetics* , 31, 64-68.
[18] Yedidia, J. S., Freeman, W. T., Weiss, Y. (2003) Understanding belief propagation and its general- izations. *Exploring Artificial Intelligence in the New Millenium*, 8, 239-236.