Approche bayésienne non paramétrique pour la factorisation de matrice binaire à faible rang avec loi de puissance

Adrien Todeschini [1] & François Caron [2]

[1] *INRIA - IMB - Univ. Bordeaux, 33405 Talence*
Adrien.Todeschini@inria.fr
[2] *Dept. of Statistics, Univ. Oxford, OX1 3TG, Oxford UK*
Francois.Caron@stats.ox.ac.uk

Résumé. Nous proposons un modèle bayésien non paramétrique (BNP) à faible rang pour les graphes bipartis. Récemment, Caron (2012) a proposé un modèle BNP où chaque élément possède son propre paramètre de sociabilité permettant de capturer le comportement en loi de puissance observé dans les graphes bipartis réels. Ce modèle peut être considéré comme une factorisation non négative de rang un de la matrice d'adjacence. En nous appuyant sur les mesures composées aléatoires récemment introduites par Griffin et Leisen (2014), nous dérivons une généralisation de rang p de ce modèle où chaque élément est à présent associé à un vecteur p-dimensionnel de paramètres de sociabilité représentant plusieurs dimensions latentes. Tout en préservant les propriétés désirées d'interprétabilité, de passage à l'échelle et de comportement en loi de puissance, notre modèle est plus flexible et offre de meilleures performances prédictives comme illustré sur plusieurs jeux de données.

Mots-clés. méthodes bayésiennes non paramétriques, factorisation de rang faible, MCMC, matrices binaires, graphes bipartis, filtrage collaboratif

Abstract. We introduce a low-rank Bayesian nonparametric (BNP) model for bipartite graphs. Recently, Caron (2012) proposed a BNP model where each node is given its own sociability parameter allowing to capture the power-law behavior of real world bipartite graphs. This model can be considered as a rank one nonnegative factorization of the adjacency matrix. Building on the compound random measures recently introduced by Griffin and Leisen (2014), we derive a rank p generalization of this model where each node is associated with a p-dimensional vector of sociability parameters accounting for several latent dimensions. While preserving the desired properties of interpretability, scalability and power-law behavior, our model is more flexible and provides better predictive performance as illustrated on several datasets.

Keywords. Bayesian nonparametrics, low-rank factorization, MCMC, binary matrices, bipartite graphs, collaborative filtering

1 Introduction

Nous nous intéressons aux réseaux bipartis, aussi appelés réseaux d'affiliation ou de collaboration. Dans ce type de réseaux, les éléments sont divisés en deux types A et B, et seules les connexions entre les éléments de types différents sont autorisées. Des exemples de ce genre peuvent être des acteurs de cinéma jouant dans le même film, des scientifiques co-auteurs d'un article, des internautes postant un message sur le même forum, des personnes qui lisent le même livre ou écoutent la même chanson, etc. Nous reprenons ici l'analogie utilisée par Caron (2012), *i.e.* les éléments de type A sont appelés lecteurs et les éléments de type B sont appelés livres.

Les méthodes bayésiennes non paramétriques (BNP) offrent une façon très élégante et utile de modéliser les relations entre deux types d'entités. Il est en effet raisonnable de considérer que l'ensemble des livres disponibles $\{\theta_j\}$ n'est pas fixé à l'avance, mais qu'il peut augmenter à mesure que de nouveaux lecteurs sont ajoutés, sa taille étant potentiellement infinie. Par ailleurs, les modèles BNP permettent de capturer les propriétés en loi de puissance de telles données.

Nous représentons l'ensemble des livres lus par le lecteur i par le processus ponctuel

$$Z_i = \sum_{j=1}^{\infty} z_{ij}\delta_{\theta_j}$$

où $z_{ij} = 1$ si le lecteur i a lu le livre θ_j, 0 sinon. La collection de mesures binaires $(Z_1, ..., Z_n)$ définit l'ensemble des relations entre les lecteurs et les livres.

Généralisant l'extension du processus du buffet indien (IBP) (Griffiths et Ghahramani, 2005, 2011) appelée IBP stable (Teh et Görür, 2009), Caron (2012) propose le modèle

$$z_{ij}|\gamma_i, w_j \sim \text{Ber}\left(1 - \exp(-\gamma_i w_j)\right)$$

où les (w_j, θ_j), $w_j > 0$ sont issus d'une mesure complètement aléatoire (CRM, voir section 2) et où chaque lecteur possède son propre paramètre d'intérêt pour la lecture $\gamma_i > 0$.

Ce modèle plus flexible permet une distribution des degrés des lecteurs non Poissonnienne, tout en conservant les propriétés de conjugaison et un processus génératif similaire à l'IBP (stable). Cependant, ce modèle est limité à une approximation de rang 1 de la collection $(Z_1, ..., Z_n)$: chaque lecteur n'a qu'un seul paramètre qui régle la quantité de livres qu'il lira et chaque livre ne possède qu'un seul paramètre qui règle sa popularité.

2 Mesures complètement aléatoires

Les CRMs (Kingman, 1967, 1993) sont des outils standards pour la construction de modèles BNP. Une CRM est une mesure aléatoire G telle que pour toute collection de sous-ensembles disjoints $A_1, ..., A_n$ d'un espace mesurable Θ, les masses aléatoires des sous-ensembles $G(A_1), ..., G(A_n)$ sont indépendantes. On considère ici les CRMs constituées

uniquement de masses aléatoires $w_j > 0$ et de localisations aléatoires $\theta_j \in \Theta$, prenant la forme

$$G = \sum_{j=1}^{\infty} w_j \delta_{\theta_j}$$

La loi de G peut être caractérisée par un processus de Poisson sur l'ensemble des points $\{(w_j, \theta_j)_{j=1,2,...}\} \subset \mathbb{R}^+ \times \Theta$ de mesure intensité ν satisfaisant $\int_0^\infty \int_\Theta (1-\exp(-w))\nu(dw, d\theta) < \infty$. On considère ici que la mesure de Lévy s'écrit $\nu(dw, d\theta) = \rho(dw)H(d\theta)$ où H est une distribution à densité. Cela implique que les localisations sont indépendantes des masses et sont i.i.d. selon H, tandis que les masses sont issues d'un processus de Poisson sur $\mathbb{R}^+$ avec pour mesure de Lévy ρ.

On s'intéresse au cas où ρ est la mesure de Lévy associée au processus de gamma généralisé (GGP) (Brix, 1999) pour sa flexibilité, son interprétabilité et ses propriétés de conjugaison

$$\rho(dw) = \frac{\alpha}{\Gamma(1-\sigma)} w^{-1-\sigma} \exp(-w\tau) dw$$

où $\alpha > 0$, $\sigma < 1$ et $\tau \geq 0$. Dans le cas où $\sigma \in (0, 1)$, la CRM est à activité infinie, *i.e.* $\int_0^\infty \rho(dw) = \infty$, et le modèle présente un comportement en loi de puissance.

3 Modèle statistique

Nous généralisons le modèle de rang 1 de Caron (2012) à un modèle de rang p : on suppose à présent l'existence de p features latentes et qu'un lecteur choisira un livre si ses intérêts pour certaines features correspondent à la popularité du livre dans ces features. Plus formellement le modèle est défini par

$$z_{ij} | \gamma_i, w_j \sim \text{Ber}\left(1 - \exp\left(-\sum_{k=1}^{p} \gamma_{ik} w_{jk}\right)\right)$$

où $\gamma_{ik} > 0$ représente l'intérêt du lecteur i pour la feature k, et $w_{jk} > 0$ est la pertinence du livre j dans la feature k. Cette représentation peut être considérée comme une factorisation non négative de rang p de la collection $(Z_1, ..., Z_n)$ capable de capturer des dépendances plus complexes dans les données. On suppose que les poids (w_{jk}) sont issus du vecteur de CRMs $(W_1, ..., W_p)$ que nous modélisons par une extension mutivariée des CRMs appelée CRM composée (CCRM), introduite par Griffin et Leisen (2014).

On montre que ce modèle possède un comportement en loi de puissance sur le nombre J de livres ayant au moins un lecteur et sur la distribution des degrés des livres.

4 Inférence

Par l'introduction d'un jeu de variables latentes convenablement choisi, on montre que la vraisemblance complète prend une forme très avantageuse pour la conjugaison

des lois *a priori* considérées permettant de dériver un algorithme de Gibbs efficace pour échantillonner selon la loi *a posteriori*. Nous complétons le modèle en supposant un *a priori* vague sur les hyper-paramètres qui sont mis à jour par une étape de Métropolis-Hastings.

Lorsque les données sont grandes, la complexité d'une itération de l'algorithme est en $\mathcal{O}\left([n+J]\times p+\sum_{i,j} z_{ij}\right)$. Néanmoins les variables de grande dimension peuvent être échantillonnées indépendemment en parallèle et seuls les hyper-paramètres de faible dimension ont une étape d'acceptation rejet.

Bibliographie

Brix, A. (1999). Generalized gamma measures and shot-noise Cox processes. *Advances in Applied Probability*, 31(4):929–953.

Caron, F. (2012). Bayesian nonparametric models for bipartite graphs. *Dans Advances in Neural Information Processing Systems 25*, éditeurs : Pereira, F., Burges, C., Bottou, L. et Weinberger, K., pages 2051–2059. Curran Associates, Inc.

Griffin, J. E. et Leisen, F. (2014). Compound random measures and their use in Bayesian nonparametrics. *arXiv preprint arXiv :1410.0611*.

Griffiths, T. et Ghahramani, Z. (2005). Infinite latent feature models and the Indian buffet process. *Dans NIPS*.

Griffiths, T. et Ghahramani, Z. (2011). The Indian buffet process : an introduction and review. *Journal of Machine Learning Research*, 12(April):1185–1224.

Kingman, J. (1967). Completely random measures. *Pacific Journal of Mathematics*, 21(1):59–78.

Kingman, J. (1993). *Poisson processes*, volume 3. Oxford University Press, USA.

Teh, Y. et Görür, D. (2009). Indian buffet processes with power-law behavior. *Dans NIPS*.