低ランク行列のベイズモデリング

中央大学大学院理工学研究科　黒澤 大樹　　　中央大学理工学部　酒折 文武

欠測を含む低ランク行列のモデリングは，ノイズを含む画像の補完や推薦アルゴリズムなど，様々な応用分野で活用されている．そして，低ランクモデリングをベイズ的に表現することで，より柔軟なモデリングや，欠測部分の予測分布に基づく推論などが可能となる．

いま，$n \times p$ データ行列を X とする．この行列には観測されない欠測部分が含まれているとする．X は潜在的に低ランク構造をもつと仮定し，その低ランク行列を復元することが目的である．低ランク行列のベイズモデリングとしては，Ruslan & Andriy (2008) により提案された Bayesian Probabilistic Matrix Factorization (BPMF) がある．BPMF は，行列のランクにあたる $r \ll \min(n, p)$ を所与として $X = AB^T + \varepsilon$ という構造を仮定し，ギブスサンプリングにより行列復元を行う手法である．この手法において r の選択は本質的な問題であるが，適切な方法は提案されていない．

一方，Mazumder *et al*. (2010) では，

$$\text{minimize} \quad \|L\|_*, \quad \text{subject to} \quad \sum_{(i,j)\in\Omega} (X_{ij} - L_{ij})^2 \le \delta \tag{1}$$

という凸最適化問題により，低ランク行列を復元する方法を提案している．ここで，$\Omega \subset \{1, \ldots, n\} \times \{1, \ldots, p\}$ は X の観測要素のインデックス集合，$\|L\|_*$ は L の特異値の和である核ノルムであり，$\delta \ge 0$ は復元誤差の許容限界を制御するパラメータである．この最適化問題は，以下のラグランジュ型の目的関数の最小化問題

$$\underset{L}{\text{minimize}} \quad \frac{1}{2}\|\mathcal{P}_\Omega(X) - \mathcal{P}_\Omega(L)\|_F^2 + \lambda\|L\|_* \tag{2}$$

で表現される．ここで，$\mathcal{P}_\Omega, \mathcal{P}_\Omega^\perp$ はそれぞれ Ω, Ω^c の上への射影

$$\mathcal{P}_\Omega(Y)_{ij} = \begin{cases} Y_{ij} & (i,j) \in \Omega \\ 0 & (i,j) \in \Omega^c \end{cases}, \quad \mathcal{P}_\Omega^\perp(Y)_{ij} = \begin{cases} Y_{ij} & (i,j) \in \Omega^c \\ 0 & (i,j) \in \Omega \end{cases},$$

$\lambda \ge 0$ は核ノルムを制御するチューニングパラメータ，$\|A\|_F$ はフロベニウスノルムである．λ を所与として，(2) はアルゴリズム soft-impute により反復的に解かれ，特異値をスパースにすることで低ランク行列を推定することができる．この推定量はまた，$X_{ij} \sim N(L_{ij}, \sigma^2)$ *i.i.d.*, $d_k \sim \mathrm{EXP}(\lambda)$ とし，EM アルゴリズムと MAP 推定による推定量として表現することが可能である．ここで d_k $(k = 1, \ldots, r)$ は L の第 k 特異値である．さらに，階層的ベイズモデルに拡張し，より適応的なモデルに拡張することもできる．

本研究では，こうした低ランクベイズモデリングにおけるハイパーパラメータおよび事前分布の選択法について提案し，比較・考察する．適応的なモデルの 1 つとして，次のモデルを提案する．

$$\underset{L}{\text{minimize}} \quad \frac{1}{2}\|\mathcal{P}_\Omega(X) - \mathcal{P}_\Omega(L)\|_F^2 + \lambda\sum_{k=1}^{r} w_k d_k, \quad w_k = \left(\frac{1}{d_k^*}\right)^\gamma, \quad \gamma > 0.$$

また，CV や情報量規準などにより適切なチューニングパラメータを選択することを検討する．さらに実データに応用し，その有効性を検証する．結果は当日報告する．

参考文献

[1] Mazumder R., Hastie T. & Tibshirani R. (2010). Spectral Regularization Algorithms for Learning Large Incomplete Matrices, *Journal of Machine Learning Research 11*, 2287-2322.

[2] Ruslan S. & Andriy M. (2008). Bayesian Probabilistic Matrix Factorization using Markov Chain Monte Carlo, *Proceedings of the 25th international conference on Machine learning*, 880-887.