D2-9

学生発表

打ち切り行列の補完：天井効果への処方箋

東京大学 THE UNIVERSITY OF TOKYO

手嶋 毅志 (東大・理研)　Miao Xu (理研)　佐藤 一誠 (東大・理研)　杉山 将 (理研・東大)

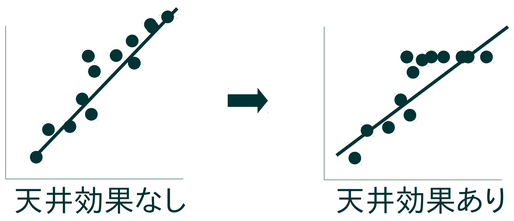
RIKEN

要約

- 「打ち切られた低ランク行列の補完」：新問題設定
  - 打ち切り＝よく見られる情報欠損なので重要
- しかし，既存の行列補完手法は，打ち切られた観測値からの行列補完には失敗する．
- 本研究の内容：
  - そもそも「打ち切り行列補完」は可解なのか？
    →確率的な完全復元のための十分条件を提示
  - 実践的にはどのような手法で復元すればよいか？
    →二乗ヒンジ損失関数を用いる補完手法を提案
  - 本問題に適した正則化を提案し，
    その下での推定誤差の確率的上界を提示
  - 人工データと実データを用いて実験的性能を検証

背景①：天井効果

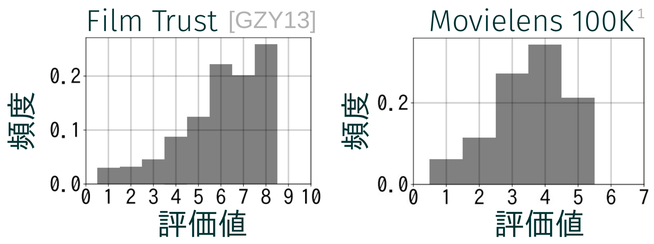
- 自然科学・社会科学では，データを取る際に，「天井効果」が普遍的にみられる　[AB03]
- 天井効果＝データの観測値が打ち切られる現象
- 統計解析に悪影響を及ぼす



天井効果なし　　天井効果あり

- 天井効果はどこで見られるのか？
  - 各種科学分野で普遍的に見られる
    1) 質問紙調査（例. 国勢調査）
    2) 標準化試験（例. IQ テスト）
    3) 生物学（例. 細胞内 ATP レベルの計測）　[YKT+14]
  - 映画推薦システムのベンチマークデータ
    1) 利用者は映画を 5 段階／8 段階で評価
    2) 評価値の頻度分布は右から打ち切られた形をしている

Film Trust [GZY13]　　Movielens 100K[1]



[1] http://grouplens.org/datasets/movielens/100k/

背景②：行列補完

- 行列補完：一部が欠損した行列データが与えられ，欠損した成分を補完するというタスク [CR12]
- 例：映画推薦……映画の評価値を行列の形に並べる

	5		4
5		1	5
5	1	3	3

（未視聴の箇所が欠損値になる）

→穴埋めによって，未視聴の映画への評価を予測

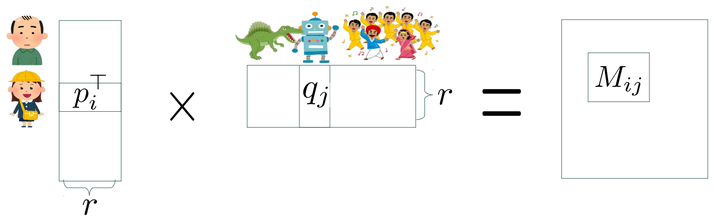
- ほかにも，マルチタスク学習，距離行列推定，etc.
- そのままでは問題として不完全

	5		4
5		1	5
5	1	3	3

単に値を埋めるだけなら何を入れてもよいから

- 意味のある値で埋めるには何かしら仮定が必要
  →真の行列が低ランクだと仮定（行列特有の仮定）

ユーザー特徴量　　映画特徴量　　評価値行列



本研究：打ち切り行列補完

打ち切り行列の推定問題（行列の形状はいずれも $n_1 \times n_2$）
真の行列 $\mathbf{M}$ が上限 C で打ち切られて出来た $\mathbf{M}^{\mathrm{c}}$
$(\mathbf{M}^{\mathrm{c}} = \min(\mathbf{M}, C))$ の，一部の添字 Ω が観測される．
$\{M^{\mathrm{c}}_{ij}\}_{ij\in\Omega}, \Omega, C$ から真の行列 $\mathbf{M}$ を精度よく推定せよ．

真値（未知）　　観測データ（上限 10）　　補完後

既存法の限界

- 打ち切りに対し，既存の（ノイズや欠測の）行列補完は
  (1)情報欠損が真値に依存→理論保証の適用対象外
  (2)観測値と真値の乖離が激しい→学習時に撹乱される

問題の可解性

- メッセージ：打ち切り行列補完は可解である
- 記号　$\mathcal{C} := \{(i,j) : M^{\mathrm{c}}_{ij} = C\}, \mathcal{C}^* := \{(i,j) : M_{ij} < C\}$
- 核ノルム最小化を考える（$\mathcal{P}_{\mathcal{S}}$ は $\mathcal{S}$ 外の成分を 0 にする作用素）

$$\arg\min_{\mathbf{X}\in\mathbb{R}^{n_1\times n_2}} \|\mathbf{X}\|_{\mathrm{tr}} \ \text{s.t.} \ \begin{cases} \mathcal{P}_{\Omega\setminus\mathcal{C}}(\mathbf{X}) = \mathcal{P}_{\Omega\setminus\mathcal{C}}(\mathbf{M}^{\mathrm{c}}), \\ \mathcal{P}_{\mathcal{C}}(\mathbf{M}^{\mathrm{c}}) \le \mathcal{P}_{\mathcal{C}}(\mathbf{X}). \end{cases} \quad (1)$$

定理（完全復元の十分条件）

- 行列 $\mathbf{M}$ の特徴量　$\mathbf{X} = \tilde{\mathbf{U}}\tilde{\mathbf{\Sigma}}\tilde{\mathbf{V}}^\top$：特異値分解　$r = \mathrm{rank}(\mathbf{M})$

$$\mu^{\mathrm{U}}(\mathbf{X}) := \max_{i\in[n_1]} \|\tilde{\mathbf{U}}_{i,\cdot}\|^2, \quad \mu^{\mathrm{V}}(\mathbf{X}) := \max_{j\in[n_2]} \|\tilde{\mathbf{V}}_{j,\cdot}\|^2$$

$$\mu_0 := \max\left\{\tfrac{n_1}{r}\mu^{\mathrm{U}}(\mathbf{M}), \tfrac{n_2}{r}\mu^{\mathrm{V}}(\mathbf{M})\right\} \quad \pi_0 := \sqrt{\tfrac{n_1 n_2}{r}}\|\mathbf{U}\mathbf{V}^\top\|_\infty$$

- 情報損失を特徴づける量

$$(\mathcal{P}^*(\mathbf{Z}))_{ij} := \mathbb{1}\{M_{ij} < C\}Z_{ij} + \mathbb{1}\{M_{ij} = C\}(Z_{ij})_+$$

$$A := \{u_k y^\top : k \in [r], y \in \mathbb{R}^{n_1}\} \quad T := \mathrm{span}(A \cup B)$$

$$B := \{x v_k^\top : k \in [r], x \in \mathbb{R}^{n_2}\} \quad \nu_{\mathcal{C}^*} := \|\mathcal{P}_{\mathrm{T}}\mathcal{P}_{\mathcal{C}^*}\mathcal{P}_{\mathrm{T}} - \mathcal{P}_{\mathrm{T}}\|_{\mathrm{op}}$$

$$\rho_{\mathrm{F}} := \sup_{\mathbf{Z}\in T\setminus\{\mathbf{O}\}:\|\mathbf{Z}\|_{\mathrm{F}}\le\|\mathbf{U}\mathbf{V}^\top\|_{\mathrm{F}}} \left\{\frac{\|\mathcal{P}_{\mathrm{T}}\mathcal{P}^*(\mathbf{Z})-\mathbf{Z}\|_{\mathrm{F}}}{\|\mathbf{Z}\|_{\mathrm{F}}}\right\}$$

$$\rho_{\infty} := \sup_{\mathbf{Z}\in T\setminus\{\mathbf{O}\}:\|\mathbf{Z}\|_{\infty}\le\|\mathbf{U}\mathbf{V}^\top\|_{\infty}} \left\{\frac{\|\mathcal{P}_{\mathrm{T}}\mathcal{P}^*(\mathbf{Z})-\mathbf{Z}\|_{\infty}}{\|\mathbf{Z}\|_{\infty}}\right\}$$

$$\rho_{\mathrm{op}} := \sup_{\mathbf{Z}\in T\setminus\{\mathbf{O}\}:\|\mathbf{Z}\|_{\mathrm{op}}\le\sqrt{n_1 n_2}\|\mathbf{U}\mathbf{V}^\top\|_{\mathrm{op}}} \left\{\sqrt{r\pi_0}\frac{\|\mathcal{P}^*(\mathbf{Z})-\mathbf{Z}\|_{\mathrm{op}}}{\|\mathbf{Z}\|_{\mathrm{op}}}\right\}$$

定理

- 仮定：ノイズなし，各成分は確率 p で独立に観測．
  $n_1, n_2 \ge 2$ および $p \ge 1/(n_1 n_2)$ とする．
  さらに $\rho_{\mathrm{F}} < \frac{1}{2}, \rho_{\mathrm{op}} < \frac{1}{4}, \rho_{\infty} < \frac{1}{2}, \nu_{\mathcal{C}^*} < \frac{1}{2}$ とする．
- もし $p \ge \min\left\{1, c_\rho \max(\pi_0^2, \mu_0) r f(n_1, n_2)\right\}$ ならば，
  $1-\delta$ 以上の確率で，(1)の解は　$\mathbf{M}$ に一致する．

$$\mathrm{c}_\rho = \max\left\{\frac{24}{(1/2-\rho_{\mathrm{F}})^2}, \frac{8}{(1/4-\rho_{\mathrm{op}})^2}, \frac{8}{(1/2-\rho_{\infty})^2}, \frac{8}{(1/2-\nu_{\mathcal{C}^*})^2}\right\}$$

$$f(n_1, n_2) = \mathcal{O}\left(\frac{(n_1+n_2)(\log(n_1 n_2))^2}{n_1 n_2}\right) \quad \delta = \mathcal{O}\left(\frac{\log(n_1, n_2)}{(n_1+n_2)^2}\right)$$

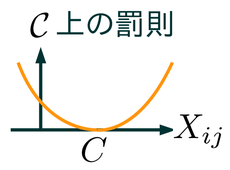
提案手法

方針

欠測とノイズの行列補完（既存法）[TY10]

$$\arg\min_{\mathbf{X}} \frac{1}{2}\sum_{ij\in\Omega}(M^{\mathrm{c}}_{ij} - X_{ij})^2 + \mathcal{R}(\mathbf{X})$$

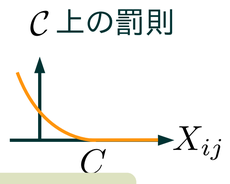
$\mathcal{C}$ 上の罰則



打ち切り行列補完（提案法）

$$\arg\min_{\mathbf{X}} \frac{1}{2}\sum_{ij\in\Omega\setminus\mathcal{C}}(M^{\mathrm{c}}_{ij} - X_{ij})^2 + \frac{1}{2}\sum_{ij\in\mathcal{C}}\max(0, M^{\mathrm{c}}_{ij} - X_{ij})^2 + \mathcal{R}(\mathbf{X})$$

$\mathcal{C}$ 上の罰則



正則化項の設計

二重核ノルム正則化 (DTr-CMC)（今回提案）

$$\mathcal{R}(\mathbf{X}) = \lambda_1\|\mathbf{X}\|_{\mathrm{tr}} + \lambda_2\|\mathrm{Clip}(\mathbf{X})\|_{\mathrm{tr}} \qquad \mathrm{Clip} = \min(\cdot, C)$$

- 真の行列と打ち切り行列の核ノルムが小さい場合に有効
- 最適化は近似的な劣勾配降下法 [AKKS12]

核ノルム正則化 (Tr-CMC) [TY10]

$$\mathcal{R}(\mathbf{X}) = \lambda\|\mathbf{X}\|_{\mathrm{tr}}$$

- 低ランク解を誘導する役割
- 最適化は加速勾配降下法

$$\|\mathbf{X}\|_{\mathrm{tr}} = \sum_{l=1}^{\min(n_1,n_2)}\sigma_l \quad (\sigma_l : \text{第 } l \text{ 特異値})$$

$$\|\mathbf{X}\|_{\mathrm{F}} = \sum_{ij} X_{ij}^2$$

フロベニウスノルム正則化 (Fro-CMC) [JNS13]

$$\mathcal{R}(\mathbf{X}) = \mathcal{R}(\mathbf{P}, \mathbf{Q}) = \frac{\lambda}{2}\left(\|\mathbf{P}\|_{\mathrm{F}}^2 + \|\mathbf{Q}\|_{\mathrm{F}}^2\right)$$

$$\mathbf{X} = \mathbf{P}\mathbf{Q}^\top \quad \mathbf{P} \in \mathbb{R}^{n_1\times k} \quad \mathbf{Q} \in \mathbb{R}^{n_2\times k}$$

- 低ランク解を誘導する役割
- 最適化は近似的な交互最小二乗法

定理（DTr-CMC の推定誤差上界）

- メッセージ：打ち切りがあっても，仮定が満たされる行列では提案法で高精度な推定が可能だと期待される．
- 二重核ノルム正則化による提案法は，

$$\widehat{\mathbf{M}} \in \arg\min_{\mathbf{X}\in G} \|\mathcal{P}_\Omega(\mathbf{M}^{\mathrm{c}} - \mathrm{Clip}(\mathbf{X}))\|_{\mathrm{F}}^2 \quad (2)$$

$G := \left\{\mathbf{X} : \|\mathbf{X}\|_{\mathrm{tr}}^2 \le \beta_1\sqrt{k n_1 n_2}, \|\mathrm{Clip}(\mathbf{X})\|_{\mathrm{tr}}^2 \le \beta_2\sqrt{k n_1 n_2}\right\}$
の目的関数を凸緩和し，制約 G を正則化項に変換した問題．

- 非正規化コヒーレンス　$\mu(\mathbf{X}) := \max\{\mu^{\mathrm{U}}(\mathbf{X}), \mu^{\mathrm{V}}(\mathbf{X})\}$

定理

$\mathbf{M} \in G$ と仮定．$\widehat{\mathbf{M}}$ は (2) の最適解，μ_G は $\sup_{\mathbf{X}\in G}\mu(\mathrm{Clip}(\mathbf{X}))$ を表す．
→ 絶対定数 C_0 および C_1 が存在し，確率 $1 - C_1/(n_1+n_2)$ 以上で

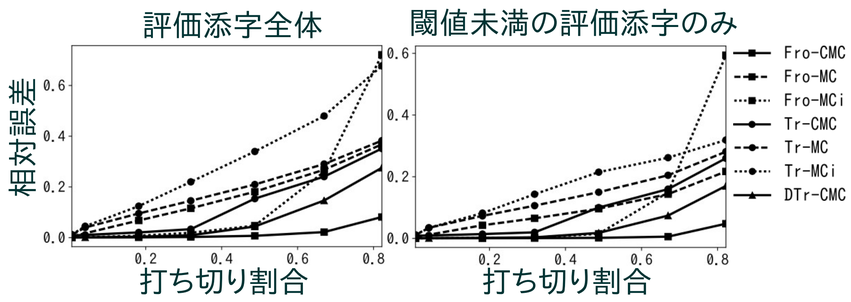
$$\sqrt{\frac{1}{n_1 n_2}\|\widehat{\mathbf{M}} - \mathbf{M}\|_{\mathrm{F}}^2} \le 2(\sqrt{\beta_1} + \sqrt{\beta_2})k^{\frac{1}{4}}(n_1 n_2)^{-\frac{1}{4}} + \sqrt{C_0\frac{2\mu_G{}^2\beta_2}{p}}\left(\frac{pk(n_1+n_2)+k\log(n_1+n_2)}{n_1 n_2}\right)^{\frac{1}{4}}$$

計算機実験

比較した手法

- *-MC：二乗誤差を用いる既存法で補完
- *-MCi：各*-MCで，打ち切られた観測値は欠測扱いにして補完

人工データを用いた補完実験



- 打ち切り閾値を変化→ $\frac{\|\mathcal{P}_{\mathrm{test}}(\widehat{\mathbf{M}}-\mathbf{M})\|_{\mathrm{F}}}{\|\mathcal{P}_{\mathrm{test}}(\mathbf{M})\|_{\mathrm{F}}}$ を評価
- 提案法は70%程度の打ち切りに対しても　10^{-2} のオーダーの誤差で精度良く推定できている

実データによる実験：設定

- タスク設定：
  天井効果を受けた実データの「真値」は得られない．
  →「評価が閾値以上かどうかを予測する」タスクで評価．
- 実験①
  - 実データを人工的に打ち切った行列（★5→★4）から学習．
  - 評価添字を「★5以上」と「★4以下」に分類．
- 実験②
  - 実データ（★1〜★5）の行列から学習．
  - 評価添字を「★5以上」と「★4以下」に分類．
- いずれもF1値で評価．
- ベースラインは，常に+1を出力する識別器．

実データによる実験：結果

実験①	**DTr-CMC**	**Fro-CMC**	Fro-MC	**Tr-CMC**	Tr-MC	(ベースライン)
Film Trust	**0.47** (0.01)	0.35 (0.01)	0.27 (0.01)	0.36 (0.00)	0.22 (0.00)	0.41 (0.00)
Movielens 100K	0.39 (0.00)	**0.41** (0.00)	0.21 (0.01)	0.40 (0.00)	0.12 (0.00)	0.35 (0.00)

- 提案法は打ち切られた成分の真値をより良く推定

実験②	**DTr-CMC**	**Fro-CMC**	Fro-MC	**Tr-CMC**	Tr-MC	(ベースライン)
Film Trust	**0.46** (0.01)	0.40 (0.01)	0.35 (0.01)	0.39 (0.00)	0.35 (0.01)	0.41 (0.00)
Movielens 100K	0.38 (0.00)	**0.41** (0.01)	0.38 (0.01)	0.40 (0.00)	0.38 (0.00)	0.35 (0.00)

5 回の試行の平均（カッコ内は標準誤差）

- 天井効果への頑健性が「高評価」の識別性能を改善

参考文献

[AB03] Austin, P.C., Brunner, L.J., 2003. Type I error inflation in the presence of a ceiling effect. The American Statistician 57.
[AKKS12] Avron, H., Kale, S., Kasiviswanathan, S.P., Sindhwani, V., 2012. Efficient and practical stochastic subgradient descent for nuclear norm regularization. ICML.
[CR12] Candès, E., Recht, B., 2012. Exact matrix completion via convex optimization. Communications of the ACM 55.
[GZY13] Guo, G., Zhang, J., Yorke-Smith, N., 2013. A Novel Bayesian Similarity Measure for Recommender Systems. IJCAI.
[JNS13] Jain, P., Netrapalli, P., Sanghavi, S., 2013. Low-rank matrix completion using alternating minimization. STOC.
[TY10] Toh, K.-C., Yun, S., 2010. An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems. Pacific Journal of Optimization 6.
[YKT+14] Yaginuma, H., Kawai, S., Tabata, K.V., Tomiyama, K., Kakizuka, A., Komatsuzaki, T., Noji, H., Imamura, H., 2014. Diversity in ATP concentrations in a single bacterial cell population revealed by quantitative single-cell imaging. Sci Rep 4.

- 本発表の論文
  Teshima, T., Xu, M., Sato, I., and Sugiyama, M. (2018). Clipped Matrix Completion: a Remedy for Ceiling Effects. ArXiv:1809.04997. (to appear in AAAI 2019).
- 連絡先：teshima@ms.k.u-tokyo.ac.jp