

イジングマシンを用いた行列因子分解による協調フィルタリング

理学専攻・情報科学コース 2240659 照井 雪乃

1 はじめに

近年、EC サイトや動画配信サービスなどの推薦システムの普及が益々進んでいる。協調フィルタリングは推薦アルゴリズムとして最も広く使用されており、ユーザーと似た嗜好をもつ他のユーザーのデータを元に、新しいコンテンツを推薦する。各ユーザーの好みの傾向を示す嗜好データは、ユーザーに評価が与えられているアイテムが少ないため、スパース（疎）な状態である。よって嗜好データからユーザーの嗜好を推定することは、データ不足の観点で困難である。これに対処するモデルとして、非負値行列因子分解 (Nonnegative Matrix Factorization; NMF) [1] が広く用いられている。本研究は NMF を拡張した非負値二値行列因子分解 (Nonnegative/Binary Matrix Factorization; NBMF) を協調フィルタリングに適用する。NBMF はイジングマシンを活用した新たな機械学習モデルであり、先行研究では画像解析に適用されている [2, 3]。画像解析で扱われるデータは密であり、NBMF が嗜好データなど疎なデータの学習に活用できるかは未だ不明である。本研究は NBMF をスパースなデータの学習に適したモデルに改良し、その性能について検証する。

2 モデル

2.1 行列因子分解

NBMF は与えられた $n \times m$ の行列 V を、 $n \times k$ の非負の実行列 W と $k \times m$ の二値のバイナリ行列 H に分解する。ここで $k \leq \min(n, m)$ である。

$$V \approx WH \quad (1)$$

式 (1) となるような W, H を求めるため、 $\|V - WH\|_F$ を最小化する。ここで、 $\|\cdot\|_F$ はフロベニウス・ノルムである。 V と行列積 WH の最小化のため、以下のように交互に行列 W, H を更新する。

$$W := \arg \min_{X \in \mathbb{R}^{n \times k}} \|V - XH\|_F^2 + \lambda_1 \|X\|_F^2 \quad (2)$$

$$H := \arg \min_{X \in \{0,1\}^{k \times m}} \|V - WX\|_F^2 + \lambda_2 \|X\|_F^2 \quad (3)$$

ここで、 λ_1, λ_2 はともに正のパラメータである。

行列 W は行毎に更新する。 W の行ベクトルを \mathbf{x} とすると、行毎に最小化する目的関数は以下である。

$$f(\mathbf{x}) = \|\mathbf{v} - H^T \mathbf{x}\|^2 + \lambda_1 \|\mathbf{x}\|^2 \quad (4)$$

ここで \mathbf{v}^T は対応する行列 V の行ベクトルである。式 (4) の最小化には射影勾配法を用い、汎用コンピュータで実行する。一方の行列 H は、列毎に更新する。 H の列ベクトルを $\mathbf{q} \in \{0, 1\}^k$ とすると、列毎に解く目的関数は以下である。

$$f(\mathbf{q}) = \|\mathbf{u} - W\mathbf{q}\|^2 + \lambda_2 \|\mathbf{q}\|^2 \quad (5)$$

ここで \mathbf{u} は対応する行列 V の列ベクトルである。式 (5) は変数が二値であるため、イジングモデルで定式化することができる。よってイジングマシンで解く。

NMF の場合は、行列 W, H がともに非負の実行列であるため、式 (3) の $X \in \{0, 1\}^{k \times m}$ が $X \in \mathbb{R}^{k \times m}$ に置き換わる。従って式 (5) の列ベクトル \mathbf{q} も非負の実行列となり、射影勾配法を用いて汎用コンピュータで解く。

2.2 提案モデル

2.2.1 マスク付き NBMF

協調フィルタリングの文脈における行列 V は嗜好データであり、評価値行列と呼ばれる。 n はユーザー数、 m はアイテム数に対応し、行列 V の (i, j) 成分 v_{ij} はアイテム j に対するユーザー i の評価を表す。一般的な行列因子分解では、 V のうち未評価に対応する要素を 0 の値で埋めて学習を行う。しかしこのように欠損値まで学習すると、データの特徴を上手く捉えられず、予測性能が低下する。そのため本研究で提案するモデルは、欠損値そのものを学習の対象から除外する。このために、行列 V と同じサイズのマスク行列 M を導入する。 M は以下のように定義する。

$$M_{ij} = \begin{cases} 1 & (V_{ij} \neq 0 \text{ の場合}) \\ 0 & (V_{ij} = 0 \text{ の場合}) \end{cases} \quad (6)$$

このマスク行列を用いて、提案モデルでは以下のように行列因子分解を行う。 \circ はアダマール積を表す。

$$M \circ V \approx M \circ (WH) \quad (7)$$

この改変に伴い、式 (4) で定義した行列 W の更新に用いた目的関数を以下のように置き換える。

$$f(\mathbf{x}) = \|\tilde{\mathbf{v}} - \tilde{H}^T \mathbf{x}\|^2 + \lambda_1 \|\mathbf{x}\|^2 \quad (8)$$

ここで i 番目の行を更新する場合、 j 番目の要素は $\tilde{v}_j = M_{ij}V_{ij}$ 、及び $(\tilde{H}^T \mathbf{x})_j = \sum_l M_{il}H_{lj}x_l$ として与える。同様に、式 (5) で定義した行列 H の更新に用いる目的関数は、以下のように置き換わる。

$$f(\mathbf{q}) = \|\tilde{\mathbf{u}} - \tilde{W}\mathbf{q}\|^2 + \lambda_2 \|\mathbf{q}\|^2 \quad (9)$$

ここで j 番目の列を更新する場合、 i 番目の要素は $\tilde{u}_i = M_{ij}V_{ij}$ 、及び $(\tilde{W}\mathbf{q})_i = \sum_l M_{il}W_{il}q_l$ として与える。

2.2.2 組合せ最適化計算

本研究では、シミュレーテッド分岐マシンを使用する [4]。このイジングマシンでは、式 (10) の形式でイジングモデルの入力を受け取る。スピン $s_i \in \{-1, 1\}$ を用いて、式 (5) の目的関数を最小化するイジングモデルを作成すると、係数行列 J_{ij}, h_i はそれぞれ式 (11), (12) で与えられる。

$$H = -\frac{1}{2} \sum_{i=1} \sum_{j=1} J_{ij} s_i s_j + \sum_{i=1} h_i s_i \quad (10)$$

$$J_{ij} = -\frac{1}{2} \sum_{\substack{r \\ i \neq j}} W_{ri} W_{rj}, \quad (11)$$

$$h_i = \frac{1}{2} \left(\sum_r W_{ri} \left(\sum_j W_{rj} - 2u_r \right) + \lambda_2 \right). \quad (12)$$

提案モデルは式 (9) を最小化するため、式 (11) と式 (12) の W_{rl} が $(\widetilde{W})_{rl} = M_{r,j} W_{rl}$ に置き換わる。

3 数値実験

3.1 実験設定

数値実験には、映画の嗜好データとして MovieLens 1M Dataset [5] を使用した。これを使用して要素の 20% に評価値が存在する評価値行列 V_0 を用意した。さらに評価が存在する要素を 5 つの異なるセットに分け、一度に 1 つのセットを隠した行列を学習データとして、5 回の試行を行った。 V_0 の評価が存在する要素についてのみ、学習して得られた行列との間で二乗平均平方根誤差 (Root Mean Squared Error; RMSE) を計算する。5 回の試行の平均値を求めることで、モデルの学習精度を評価する。

3.2 既存手法との比較

提案手法 NBMF と従来手法 NMF の学習精度を比較した。どちらの手法もマスクを付け、欠損値は学習しない。10 エポックの学習における RMSE と計算時間の推移を表した結果が図 1 である。

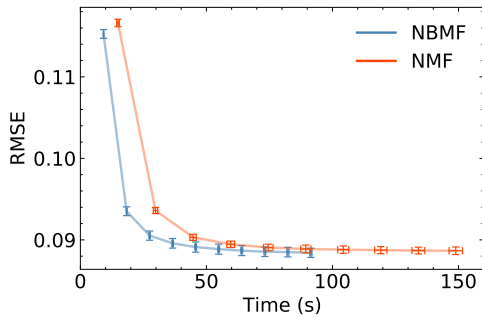


図 1: エポック毎の RMSE と計算時間の推移. 5 回の試行の平均. エラーバーは標準偏差を表す。

両手法におけるエポック毎の RMSE の差は僅かであるが、減衰する様子は NMF よりも NBMF の方が急速であることが分かる。10 エポックの学習における RMSE は NBMF が僅かに良く、計算時間は NBMF が 1.63 倍高速であった。

3.3 充填率依存性

評価値行列 V_0 の充填率を変え、10 エポックの学習を行った結果を図 2 に示す。

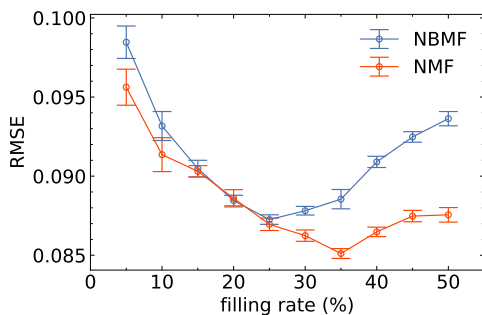


図 2: V_0 の充填率を変化させた結果. 5 回の試行の平均. エラーバーは標準偏差を表す。

充填率が極端に低いまたは高い場合、学習精度が悪化する傾向がある。その原因として、充填率が低い場合は学習データが不足し予測が難しくなるため、また充填率が高い場合は最適化する目的関数が複雑化するためと考えられる。

NMF は分解における解像度の高さから、NBMF よりも RMSE が低いことが期待される。しかしその期待に反し、20%未満の充填率では NBMF と NMF の RMSE の差は殆ど見られない。このことは、NMF が充填率が低い場合に過学習を起こし、学習能力を低下させる可能性があること、また NBMF が過学習に対して比較的頑健であることを示している。

3.4 マスキング処理による効果

本研究で提案した NBMF は、従来型の NBMF アルゴリズムにマスキング処理を追加した。マスキングの有無による結果の比較を表 1 に示す。処理が有る場合の結果は図 1 と同一の結果を使用している。

	計算時間 (s)	RMSE
マスキング有	91.4	0.088
マスキング無	84.0	0.348

表 1: 10 エポックの学習における最終的な RMSE と計算時間の平均。

マスキング処理が有る場合は、無い場合に比べて目的関数が複雑化するため、計算時間が長くなる。しかし RMSE は 4 分の 1 程度に減少し、精度が大幅に向上する。この結果は、マスキング処理が協調フィルタリングにおいて重要な役目を果たすことを示している。

4 まとめ

協調フィルタリングに使用される嗜好データはスパースであることが課題であり、本研究はスパースなデータの学習に適したマスク付き NBMF を考案した。従来手法の NMF と比較すると、提案手法は計算が高速かつ過学習にも比較的頑強であることから、スパースなデータの学習に有用であると言える。本研究により、イジングマシンを利用した NBMF の新しい応用可能性が示された。

参考文献

- [1] C.-J. Lin, *Neural Comput.* **19**, 2756–2779 (2007).
- [2] D. O’Malley, V. V. Vesselinov, B. S. Alexandrov and L. B. Alexandrov, *PLOS ONE* **13**, 1 (2018).
- [3] H. Asaoka and K. Kudo, *Journal of the Physical Society of Japan* **89**, 085001 (2020).
- [4] H. Goto, K. Endo, M. Suzuki, Y. Sakai, T. Kanao, Y. Hamakawa, R. Hidaka, M. Yamasaki and K. Tatsumura, *Science Advances* **7**, eabe7953 (2021).
- [5] F. M. Harper, J. A. Konstan and J. A., *ACM Trans. Interact. Intell. Syst.* **5**, 1 (2016).