

# 被評価数を考慮した重み付き最小二乗法による EM-NMF アンサンブル手法

EM-NMF ensemble method by weighted least squares method considering the number of evaluations

大堀 祐一<sup>\*1</sup>      山下 遥<sup>\*2</sup>      後藤 正幸<sup>\*1</sup>  
 Yuichi Ohori      Haruka Yamashita      Masayuki Goto

<sup>\*1</sup>早稲田大学院創造理工学研究科経営デザイン専攻

Graduate School of Creative Science and Engineering Major in Management Design, Waseda University

<sup>\*2</sup>上智大学

Sophia University

In recent years, the importance of recommendation system has been increasing from the development of information technology. One of the important technologies for the recommendation is collaborative filtering. In this study, we focus on EM-NMF which is an effective model for the collaborative filtering. The approach is based on matrix decomposition. Generally, evaluation values by users are biased to some number of items. therefore, EM-NMF tends to learn emphatically to items with many evaluations. The prediction accuracy of evaluation for items with a small number of evaluation data tends to be undesirable. In this study, we propose a method to assemble two matrices; (i) predicted evaluation matrix based on the approach of items with many evaluation oriented and (ii) the matrix based on the approach of items with small number of evaluation obtained. This approach is expected to improve the prediction accuracy for the evaluation.

## 1. はじめに

近年、EC サイト上では大量の商品が扱われるようになり、ユーザの嗜好も多様化している。これに伴い、自動で各ユーザの嗜好に合致した商品を提示する推薦システムの重要性が高まっている。推薦システムの一つに、各ユーザの評価履歴に基づき、類似した他のユーザとの関係性から評価値予測を行うことでアイテムを推薦する協調フィルタリングがある。協調フィルタリングには様々な手法があるが、本研究では、行列分解に基づく手法である EM Non-negative Matrix Factorization[1](以下、EM-NMF) に着目する。行列分解の手法としては、Non-negative Matrix Factorization(NMF)[2] が基本的なモデルであるが、協調フィルタリングでは、欠損している要素の予測評価値を推定するため、欠損値を推定する手法である EM アルゴリズム [3] を適用した EM-NMF が用いられる。EM-NMF は、アイテムとユーザがそれぞれ共通のクラスタに所属すると考え、評価値行列をそれらへの所属度合いを要素に持つ二つの低次元非負値行列に分解し、分解後の行列の積により評価値を予測する手法である。

しかし一般に評価値は、被評価数の多い少数の人気アイテム(MAJOR アイテム)と、大多数の被評価数が少ないアイテム(MINOR アイテム)から構成されている。EM-NMF の学習における評価関数は、評価値と予測値との二乗誤差の総和であることから、この評価関数に含まれる評価データ数が多い人気アイテムの情報が重視されるように行列分解が行われてしまう。そのため、人気アイテムを嗜好するユーザに対しては適切な推薦を行うことができる一方、被評価数の少ないアイテムを嗜好するユーザに対する予測評価値は、相対的に予測精度が低くなってしまおうという問題がある。

これに対して、斎藤 [4] は、MAJOR アイテムを一般的なユーザ(MAJOR ユーザ)がよく嗜好するアイテム、MINOR アイテムをニッチなユーザ(MINOR ユーザ)がよく嗜好するアイテムと定義した。そして、MAJOR ユーザと MINOR ユーザでは嗜好が異なると考え、あらかじめ学習データを MAJOR ユーザと MINOR ユーザに分割する。それぞれのデータにお

いて学習した部分評価値行列から推定した予測評価値と評価値行列全体から推定した予測値をそれぞれ算出し、その予測値をアンサンブル [5] することで、予測精度を向上させた。

しかし、斎藤の手法は予め学習データを分割する必要があるが、その分割基準は経験的に決める必要があり、最適である保証はない。その上、学習データを分割しているため、各評価値データはどちらか一方の行列分解にしか使われず、部分評価値行列の学習が十分行えない可能性がある。

そこで本研究では、学習データを十分に活用するために、データを分割するのではなく評価関数に対して重み付き最小二乗誤差の考え方を導入する。この重みを調整することで、MAJOR アイテムと MINOR アイテムのそれぞれを重視した推定を可能にする。MAJOR アイテムと MINOR アイテムのそれぞれを重視したモデルを独立して学習し、その学習されたモデルから得られた二つの予測評価値の行列と、通常の EM-NMF から得られた予測評価値の行列をアンサンブルすることによって、予測精度の向上を目指す。

## 2. 従来手法

### 2.1 EM-NMF

EM-NMF とは評価値行列  $A$  を二つの低次元かつ非負の特徴行列の積で近似し、未評価要素の予測を行う手法である。いま、 $N$  個からなるアイテム集合を  $\mathcal{I} = \{I_i : 1 \leq i \leq N\}$ 、 $M$  人からなるユーザ集合を  $\mathcal{J} = \{J_j : 1 \leq j \leq M\}$  と定義する。アイテム  $I_i$  に対し、ユーザ  $J_j$  が  $Y$  段階評価で  $y$  点の評価をした場合は  $a_{ij} = y$  を要素、未評価の場合の  $a_{ij}$  は欠損とする評価値行列を  $A = [a_{ij}] \in R^{N \times M}$  と定義する。二つの分解した行列をそれぞれアイテム  $I_i$  がクラスタ  $k$  ( $1 \leq k \leq K$ ) に所属する度合いを要素に持つ行列を  $U = [u_{ik}] \in R^{N \times K}$ 、ユーザ  $J_j$  がクラスタ  $k$  に所属する度合いを要素に持つ行列を  $V = [v_{kj}] \in R^{K \times M}$  とし、 $U, V$  の積によって算出される評価値行列の近似行列を  $X = [X_{ij}] \in R^{N \times M}$  と定義する。EM-NMF の学習では、 $A'$  が更新される都度、式 (1) を最小化するように行列  $U, V$  を更新する。なお、 $\|\cdot\|_F$  はフロベニウスノルムを表す。

$$\min_{U,V} \sum_{i,j} |a'_{ij} - u_{ik}v_{kj}|^2 \quad (1)$$

$$s.t. \forall u_{ik} \geq 0, \forall v_{kj} \geq 0$$

いま、評価値行列  $A$  の評価要素の集合を  $A^\circ$ 、未評価要素の集合を  $A^u$  と定義する。式 (1) の解を得るために、 $a_{ij} \in A^\circ$  に該当する要素は評価値で固定し、 $a_{ij} \in A^u$  に該当する要素を繰り返し演算により更新して求める。具体的には、 $t$  回目の更新により算出された行列  $X$  の値を  $x_{ij}^{(t)}$ 、行列  $A$  の未評価要素に  $x_{ij}^{(t-1)}$  を代入した行列を  $A^{(t)} = [a_{ij}^{(t)}] \in R^{N \times M}$  とすると、次の式 (2)~(4) を用いて行列  $A'$  と  $U, V$  を繰り返し更新することで式 (1) の問題を解くことができる。行列  $A', U, V, X$  について、 $t$  回目の更新後の各行列を  $A^{(t)}, U^{(t)}, V^{(t)}, X^{(t)}$  とし、および各行列の  $(i, j)$  要素を  $a'_{ij}, u_{ij}, v_{ij}, x_{ij}^{(t)}$  とする。また  $t = 0$  は各行列の初期値を表す。

$$u_{ik}^{(t)} \leftarrow u_{ik}^{(t-1)} \frac{\sum_j a_{ij}^{(t-1)} v_{kj}^{(t-1)}}{\sum_j v_{kj}^{(t-1)} \sum_{k'} u_{ik'}^{(t-1)} v_{k'j}^{(t-1)}} \quad (2)$$

$$v_{kj}^{(t)} \leftarrow v_{kj}^{(t-1)} \frac{\sum_i a_{ij}^{(t-1)} u_{ik}^{(t-1)}}{\sum_i u_{ik}^{(t-1)} \sum_{k'} u_{ik'}^{(t-1)} v_{k'j}^{(t-1)}} \quad (3)$$

$$a'_{ij} = \begin{cases} a_{ij} & (a_{ij} \in A^\circ) \\ x_{ij}^{(t-1)} & (a_{ij} \in A^u) \end{cases} \quad (4)$$

以上の式より求めた収束後の行列  $A^{(T)}$  の値を用いて、対象ユーザーに推薦を行う。

## 2.2 ユーザの嗜好を考慮した非負値行列因子分解 [4]

斎藤の手法は被評価数の多いアイテムと被評価数の少ないアイテムを被評価数の観点から分割し、それぞれのアイテムごとにモデルを構築する。そして、複数のモデル (EM-NMF と被評価数を考慮した EM-NMF) を組み合わせて学習機を構成することで単体のモデルよりも予測精度が向上させるアンサンブル [5] を用いてモデルを構築する。具体的には、まず被評価数に応じてアイテムを MAJOR アイテムと MINOR アイテムに分類し、さらに MAJOR アイテムをよく嗜好するユーザーを MAJOR ユーザー、そうでないユーザーを MINOR ユーザーとして分類する。そして、MAJOR ユーザーと MINOR ユーザーに分割したそれぞれの学習データから学習した部分評価値行列を全体のデータから学習した評価値行列と適切にアンサンブルすることで被評価数の少ないアイテムを重視した評価値行列を構築することができる、予測精度が向上する。

実行手順として、MAJOR ユーザーの評価値行列を  $\bar{A} (\in N^{N \times \bar{M}})$ 、MINOR ユーザーの評価値行列を  $\underline{A} (\in N^{N \times \underline{M}})$  とする。ただし、 $\bar{M}$  は MAJOR ユーザー数、 $\underline{M}$  は MINOR ユーザー数とする。すなわち、 $M = \bar{M} - \underline{M}$  である。また、分割した行列の  $(i, j)$  要素を、それぞれ  $\bar{a}_{ij}, \underline{a}_{ij}$  とする。そして、 $A, \bar{A}, \underline{A}$  それぞれに対し EM-NMF を行い、予測値をアンサンブルすることによって、予測評価値を算出する。分割した各予測評価値行列を  $\bar{A}', \underline{A}'$  とし、それらの行列の  $(i, j)$  要素をそれぞれ  $\bar{a}'_{ij}, \underline{a}'_{ij}$  とする。その際のパラメータを MAJOR ユーザーの混合比率  $\alpha (0 \leq \alpha \leq 1)$ 、MINOR ユーザーの混合比率  $\beta (0 \leq \beta \leq 1)$  とする。以下の式 (5) より、予測評価値行列  $\tilde{A} = [\tilde{a}_{ij}] \in R^{N \times M}$  を算出する。

$$\tilde{a}_{ij} = \begin{cases} \alpha \bar{a}'_{ij} + (1 - \alpha) a'_{ij}, & \text{if } a'_{ij} \in \bar{A} \\ \beta \underline{a}'_{ij} + (1 - \beta) a'_{ij}, & \text{if } a'_{ij} \in \underline{A} \end{cases} \quad (5)$$

## 3. 提案手法

### 3.1 着想

従来手法は、EM-NMF における被評価数に応じて偏った学習をしてしまう問題を MAJOR ユーザーと MINOR ユーザーの部分評価値行列を用いることで解決した。しかし、斎藤の手法は MAJOR アイテムと MINOR アイテムに事前に分ける必要がある。分けるための最適な閾値は数式的には求められないため、誤って分類されてしまい、予測精度が下がってしまう可能性がある。また、MAJOR ユーザーと MINOR ユーザーの評価値行列の学習の際には、分割によって学習データも少なくなるため、全体のデータを用いて学習する評価値行列と比較して、部分評価値行列は十分に学習されない可能性がある。

そこで、アイテムの被評価数の偏りを考慮するためにユーザーを分割するのではなく、全体のデータを用いて、評価関数に対して重み付き二乗誤差の考え方を導入し、MINOR アイテムと MAJOR アイテムのそれぞれを重視した学習を独立に行うことを考える。そして、得られた二つの評価値行列と通常の EM-NMF から得られた評価値行列をアンサンブルすることで予測精度を向上させる。重みには事前に設定するパラメータを導入し、そのパラメータを調節することで MAJOR アイテムと MINOR アイテムのそれぞれを重視した学習を可能にする。

### 3.2 提案モデル

$l_j$  を  $j$  番目のアイテムの被評価数、 $\bar{l}$  を全アイテムの平均被評価数、 $\gamma$  を事前に決定するパラメータとした時、MINOR アイテムと MAJOR アイテムのそれぞれを重視して学習を行うために、MINOR アイテムは式 (6)、MINOR アイテムは式 (7) のように被評価数に基づく重みを学習に導入する。

$$w_j^{major} = \left( \frac{1}{1 + e^{\frac{l_j - \bar{l}}{\gamma}}} \right) \quad (6)$$

$$w_j^{minor} = 1 - \left( \frac{1}{1 + e^{\frac{l_j - \bar{l}}{\gamma}}} \right) \quad (7)$$

式 (6) は被評価数が多いほど 1 に近い値を取り、少ないほど 0 に近い値を取る。反対に、式 (7) は被評価数が多いほど 0 に近い値を取り、少ないほど 1 に近い値を取る。これらの重みを目的関数に導入することで、MAJOR アイテムと MINOR アイテムのそれぞれを重視した学習が可能となる。

式 (6), (7) を導入した MAJOR アイテム、MINOR アイテムのそれぞれの目的関数は式 (8), (9) のようになる。

$$\min_{U,V} \sum_{i,j} |w_j^{major} (a'_{ij} - u_{ik}v_{kj})|^2 \quad (8)$$

$$\min_{U,V} \sum_{i,j} |w_j^{minor} (a'_{ij} - u_{ik}v_{kj})|^2 \quad (9)$$

目的関数である式 (8), (9) を最小化する  $u_{ik}, v_{kj}$  の更新式は以下のようになる。

#### MAJOR アイテム :

$$u_{ik}^{(t)} \leftarrow u_{ik}^{(t-1)} \frac{\sum_j w_j^{major} a_{ij}^{(t-1)} v_{kj}^{(t-1)}}{\sum_j w_j^{major} v_{kj}^{(t-1)} \sum_{k'} u_{ik'}^{(t-1)} v_{k'j}^{(t-1)}} \quad (10)$$

#### MINOR アイテム :

$$u_{ik}^{(t)} \leftarrow u_{ik}^{(t-1)} \frac{\sum_j w_j^{minor} a_{ij}^{(t-1)} v_{kj}^{(t-1)}}{\sum_j w_j^{minor} v_{kj}^{(t-1)} \sum_{k'} u_{ik'}^{(t-1)} v_{k'j}^{(t-1)}} \quad (11)$$

共通：

$$v_{kj}^{(t)} \leftarrow v_{kj}^{(t-1)} \frac{\sum_j a_{ij}^{(t-1)} u_{ik}^{(t-1)}}{\sum_j u_{ik}^{(t-1)} \sum_{k'} u_{ik'}^{(t-1)} v_{k'j}^{(t-1)}} \quad (12)$$

$$a_{ij}^{(t)} = \begin{cases} a_{ij} & (a_{ij} \in A^o) \\ x_{ij}^{(t-1)} & (a_{ij} \in A^u) \end{cases} \quad (13)$$

目的関数が収束するまで、これらの式 (10) ~ (13) を用いて更新を繰り返す。そして、それぞれ式 (14), (15) のように得られた評価値行列をアンサンブルし、予測評価値を算出する。

$$\tilde{a}_{ij}^{major} = w_j^{major} a_{ij}^{major} + (1 - w_j^{major}) a_{ij}^{normal} \quad (14)$$

$$\tilde{a}_{ij}^{minor} = w_j^{minor} a_{ij}^{minor} + (1 - w_j^{minor}) a_{ij}^{normal} \quad (15)$$

$a_{ij}^{major}$ ,  $a_{ij}^{minor}$  は MAJOR アイテム, MINOR アイテムをそれぞれ重視して学習した EM-NMF から算出した予測評価値であり,  $a_{ij}^{normal}$  は通常の EM-NMF から得られた予測評価値である。

式 (14), (15) から得られた予測評価値は最終的に式 (16) のように統合する。

$$\tilde{a}_{ij} = \begin{cases} \tilde{a}_{ij}^{major}, & \text{if } w_j^{major} \geq w_j^{minor} \\ \tilde{a}_{ij}^{minor}, & \text{if } w_j^{major} < w_j^{minor} \end{cases} \quad (16)$$

このように統合することで、それぞれ重視して学習されていない予測評価値を取り除くことができ、予測精度が向上する。

## 4. 実験

提案手法の有効性を示すため、推薦システムのベンチマークデータを用いた評価実験を行い、従来手法である EM-NMF と斎藤の手法との比較を行った。

### 4.1 実験条件

実験では、公開データセット MovieLens[6] の映画評価データ 100 万件を用いた。ユーザ数  $M = 6,040$ (人), アイテム数  $N = 3,952$ (本), 評価段階数は (1~5) である。また、従来手法において、被評価数上位二割を MAJOR アイテム、それ以外を MINOR アイテムと定義した。ユーザに関しては、評価している映画の八割が MAJOR アイテムであれば MAJOR ユーザ、それ以外のユーザを MINOR ユーザと定義した。それぞれのユーザ数は、 $\overline{M} = 2,018, \underline{M} = 4,022$  である。アンサンブルの割合は  $\alpha = 0.2, \beta = 0.5$  とした。提案手法において、パラメータ  $\gamma = 100$  とした。また、すべての手法において  $K = 10$  とした。五分割交差検定実験を行い、その平均値により予測精度の比較を行う。評価指標には MAE (平均絶対誤差) を用いる。

### 4.2 実験結果

MAJOR アイテムと MINOR アイテムのそれぞれに対する MAE を以下の表 1 に示す。

表 1. MAJOR アイテム, MINOR アイテムにおける各手法の MAE

|       | EM-NMF | 斎藤の手法 | 提案手法  |
|-------|--------|-------|-------|
| MAJOR | 0.719  | 0.709 | 0.702 |
| MINOR | 0.739  | 0.733 | 0.722 |

表 1 より、提案手法は MAJOR アイテムや MINOR アイテムの両方において従来手法よりも優れていることがわかる。この結果より、EM-NMF に対して、斎藤の手法や提案手法が優れていることから、被評価数を考慮することで予測精度が向上していることがわかる。また、斎藤の手法よりも提案手法の方が MAJOR アイテムや MINOR アイテムに対する予測精度が向上していることから、全体のデータを用いた学習の効果が出ていることが分かる。

次に、表 2 に全体のテストデータに対する MAE を示す。

表 2. 各手法における MAE

|     | EM-NMF | 斎藤の手法 | 提案手法  |
|-----|--------|-------|-------|
| MAE | 0.736  | 0.730 | 0.719 |

表 2 より、提案手法は全体のテストデータにおいても従来手法よりも優れていることがわかる。この結果より、提案手法は独立に学習を行った MAJOR アイテムと MINOR アイテムの評価値行列と通常の NMF から得られた評価値行列を適切にアンサンブルできていることがわかる。

## 5. まとめと今後の課題

評価関数に対して被評価数に応じた重みを導入することで、被評価数の多いアイテムと少ないアイテムのそれぞれを重視して学習を行い、その学習されたモデルから得られた二つの予測評価値の行列と、通常の EN-MMF から得られた予測評価値の行列をアンサンブルする手法を提案した。また、ベンチマークデータを用いた実験により提案手法の有効性を示すことができた。

今後の課題として、本実験では学習において用いる重みに必要なパラメータを経験的に求めたが、そのパラメータの適切な選択方法の検討などが挙げられる。また、提案手法においてはアイテムにおける被評価数を重みに活用したが、ユーザごとにおける被評価数も用いたモデルに拡張することで、予測精度が向上が期待できる。

## 参考文献

- [1] Zhang, S, Wang, W, Ford, J, Makedon, F, "Learning from Incomplete Ratings Using Non-negative Matrix Factorization," *6th SIAM Conference on Data Mining (SDM)*, pp.549-553, 2006.
- [2] L. Daniel, and H. Seung, "Algorithms for nonnegative matrix factorization," *NIPS' 00 Proceeding of the 13th International Conference on Neural Information Processing Systems*, pp.535-541, 2000.
- [3] Dempster, A. P., Laird, N. M., and Rubin, D. B. "Maximum Likelihood from Incomplete Data via The EM Algorithm", *Journal of The Royal Statistical Society (B)*, Vol. 39, No. 1, pp.1-38, 1977.
- [4] 斎藤央樹, "ユーザの嗜好を考慮した非負値行列因子分解の提案", 早稲田大学創造理工学部経営システム工学科 2016 年度卒業論文.
- [5] Robert, S, "The Strength of Weak Learnability," *Machine Learning*, Vol.5(2), pp.197-227, 1990.
- [6] MovieLens. "http://www.movielens.org/", 2019/2/15 アクセス.