

# 属性情報を融合したNMFによる ソーシャルメディアの共有イベント予測

藤原稜 熊野雅仁 木村昌弘

龍谷大学理工学部電子情報学科

t140169@mail.ryukoku.ac.jp kumano@rins.ryukoku.ac.jp kimura@rins.ryukoku.ac.jp

**概要** NMFは、ソーシャルメディアにおけるユーザの行動分析に有効な手法であることが示されている。我々は、アイテムレビューサイトを対象として、2種類の潜在空間に基づき評価情報と感謝ポイントという属性情報をNMFに融合することにより、ユーザ間の信頼リンクを予測する手法を提案した。そして、そのような属性情報を融合したNMFは、通常のNMFよりもユーザ間の信頼リンクの予測精度が高いことを実証した。本研究では、その研究を拡張して、2種類の潜在空間に基づき属性情報を融合したNMFにより、オンラインアイテムのユーザによる共有イベントを予測する手法を提案し、コスメ・美容の総合サイトと料理レシピ共有サイトの実データを用いた実験でその有効性を評価する。

**キーワード** ソーシャルメディアマイニング, 非負値行列因子分解, 共有イベント予測

## 1 はじめに

近年、ソーシャルメディアサイト上のユーザ行動を分析する研究が注目されている。NMF(Nonnegative Matrix Factorization)法 [1] はソーシャルメディアにおけるユーザの行動分析に有効であることが知られているが、Tangら [2] は、リンクとアクティビティの両方の情報を利用できるhTrust法を与え、信頼リンク予測問題に対してNMFを改良した。我々は、評価と感謝ポイントの情報に基づいた2種類の潜在空間を用いてNMFを拡張することにより、アイテムレビューサイトにおけるユーザ間の信頼リンクを予測する手法を提案し、通常のNMFやhTrustよりも高い予測性能を得た [3]。本研究では、その研究を拡張して、2種類の潜在空間に基づき属性情報を融合したNMFにより、オンラインアイテムのユーザによる共有イベントを予測する手法を提案し、コスメ・美容の総合サイトと料理レシピ共有サイトの実データを用いた実験でその有効性を評価する。

## 2 提案法

本章では、我々が以前に与えたリンク予測のNMF法 [3] を拡張して、2種類の潜在空間に基づき属性情報を融合したNMFにより共有イベントを予測する手法を提案する。

ユーザ集合を  $\mathcal{V} = \{v_1, \dots, v_N\}$ 、アイテム集合を  $\mathcal{R} = \{r_1, \dots, r_M\}$  とし、ユーザがアイテムを共有したことを表す  $N \times M$  行列を  $G = (G_{i,\alpha})$  とする。ただし、ユーザ  $v_i$  がアイテム  $r_\alpha$  を共有したなら  $G_{i,\alpha} = 1$  であり、そうでないなら  $G_{i,\alpha} = 0$  である。我々は、ユーザの指向を表す  $K$  次元潜在空間とアイテムの特徴を表す  $L$  次

元潜在空間を用いて、非負値  $N \times K$  行列  $U = (U_{i,k})$ 、非負値  $K \times L$  行列  $H = (H_{k,\ell})$ 、非負値  $L \times M$  行列  $W = (W_{\ell,\alpha})$  を導入し、

$$\begin{aligned} \mathcal{F}(U, W, H) = & \|G - UHW^T\|^2 + \lambda_U \|U\|^2 + \lambda_W \|W\|^2 + \lambda_H \|H\|^2 \\ & + \frac{\lambda_X}{2} \sum_{i,j=1}^N \xi_{i,j} \sum_{k=1}^K (U_{i,k} - U_{j,k})^2 \\ & + \frac{\lambda_Y}{2} \sum_{\alpha,\beta=1}^M \eta_{\alpha,\beta} \sum_{\ell=1}^L (W_{\alpha,\ell} - W_{\beta,\ell})^2 \end{aligned}$$

を最小化することにより、 $U, W, H$  の最適値  $U^* = (U_{i,k}^*)$ 、 $W^* = (W_{\ell,\alpha}^*)$ 、 $H^* = (H_{k,\ell}^*)$  を探索する。ただし、 $\|\cdot\|$  は行列のFrobeniusノルムであり、 $\lambda_U, \lambda_W, \lambda_H, \lambda_X, \lambda_Y > 0$  は超パラメータである。また、 $\xi_{i,j}$  はユーザ  $v_i$  とユーザ  $v_j$  の類似度を、 $\eta_{\alpha,\beta}$  はアイテム  $r_\alpha$  とアイテム  $r_\beta$  の類似度をそれぞれ表すものであり、対象とするサイトで提供された属性情報を用いて定義される。我々は、 $G_{i,\alpha} = 0$  であるユーザとアイテムの候補  $(v_i, r_\alpha)$  に対して、

$$G_{i,\alpha}^* = \sum_{k=1}^K \sum_{\ell=1}^L U_{i,k}^* H_{k,\ell}^* W_{\ell,\alpha}^*$$

の値に基づきそれら候補をランキングすることにより、共有イベント予測を行う。

コスメ・美容の総合サイトの実験では、 $v_i$  と  $v_j$  はフォローしているユーザがどのくらい同じかという観点から  $\xi_{i,j}$  を定義し、 $r_\alpha$  と  $r_\beta$  は同じユーザによる評価がどのくらい類似していたかという観点から  $\eta_{\alpha,\beta}$  を定義した。また、料理レシピ共有サイトの実験では、 $v_i$  と  $v_j$  は投稿したレシピで使われた食材がどのくらい類似していたかという観点から  $\xi_{i,j}$  を定義し、 $r_\alpha$  と  $r_\beta$  は料理の作り

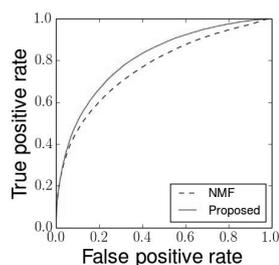


図 1: @cosme データセットにおけるレビュー予測結果

方を記載した文書情報から抽出した単語がどのくらい類似しているかという観点から  $\eta_{\alpha,\beta}$  を定義した。

### 3 実験

#### 3.1 実験設定

本研究では二つのソーシャルメディアを対象に実データによる実験を行った。まず、コスメ・美容の総合ポータルサイト「@cosme」の1999年9月から2010年6月までの実データを用いた。アイテム数122,927個、ユーザ数46,256人、レビュー数1,177,589回であった。なお、予測性能を検証する上で、レビューが活発に行われているアイテムとレビューを活発に行うユーザに焦点を当てるため、レビュー数10回以上のアイテムと、レビューを10回以上しているユーザを抽出した。1年間を学習して次の1カ月を予測する、6つのデータセット  $D_1, \dots, D_6$  を構築した。紙数の関係上、 $D_1$  のみの結果を示す。 $D_1$  は、アイテム数8,139 ユーザ数8,121 レビュー数300,050 となった。また、料理レシピ共有サイト「cookpad」の2000年1月から2014年8月までの実データを用いた。レシピ数90,209個、ユーザ数272,552人、つくれば数3,196,445回であった。なお、つくればが行われているレシピとつくればを行うユーザに焦点を当てるため、つくれば数10回以上のレシピと、つくればを1回以上しているユーザを抽出した。6カ月を学習して次の6カ月を予測する、6つのデータセット  $E_1, \dots, E_6$  を構築した。紙数の関係上、 $E_1$  のみの結果を示す。 $E_1$  は、レシピ数505、ユーザ数2,096、つくれば数13,772 となった。各実験における超パラメータは交差検証により決定することが可能である。本研究では単純に [2] に従って設定した。特に、@cosme データでは  $K=L=100$ ,  $\lambda_U=\lambda_W=\lambda_H=0.01$  とし、cookpad データでは  $K=L=100$ ,  $\lambda_U=\lambda_W=\lambda_H=0.01$  と設定した。

#### 3.2 実験結果

##### 3.2.1 @cosme データの結果

データセット  $D_1$  における NMF と提案法のレビュー予測性能を図 1 に ROC 曲線で示す。なお、この結果は、 $D_1$  において高い性能を示した超パラメータ  $\lambda_X=4$ ,  $\lambda_Y=1$  のときの結果である。NMF に対して提案法は予

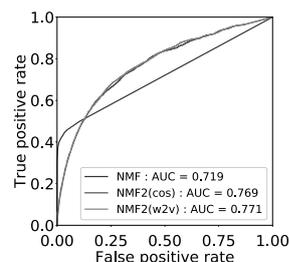


図 2: Cookpad データセットにおけるつくれば予測結果

測性能が高いことが分かる。また、残りのデータセットにおいても提案法の予測性能が NMF を上回った。

##### 3.2.2 Cookpad データの結果

データセット  $E_1$  における NMF と、提案法においてユーザ間の類似性とレシピ間の類似性の計算について  $\cos$  類似度を用いた方法および  $\cos$  類似度に word2vec による類似度を組み込んだ方法に関するつくれば予測性能を図 2 に ROC 曲線で示す。なお、この結果は、 $E_1$  において高い性能を示した超パラメータ  $\lambda_X=6$ ,  $\lambda_Y=1$  のときの結果である。AUC で評価した結果 (図 2 の凡例参照), NMF 法に対して提案法の予測性能が高い値を示すことがわかる。また、AUC が最も高い値を示したのはユーザ間およびレシピ間の類似度において  $\cos$  類似度に word2vec による類似度を組み込んだ方法の場合であったが、 $\cos$  類似度を単純に用いた方法との違いはほとんどないことがわかった。

### 4 まとめ

本研究では、2種類の潜在空間に基づく属性情報を融合した NMF により、オンラインアイテムのユーザによる共有イベントを予測する手法を提案し、アイテムレビューサイト「@cosme」と料理レシピ共有サイト「cookpad」の実データを用いた実験により、NMF 法と提案 NMF 法とを比較し、その有効性を示した。潜在空間の最適な次元数や超パラメータの影響を詳細に調査することは今後の課題である。

### 謝辞

本研究では、クックパッド株式会社と国立情報学研究所が提供するクックパッドデータを利用した。

### 参考文献

- [1] Lee, D. D. and Seung, H. S., “Algorithms for non-negative matrix factorization,” in *Proceedings of NIPS’01*, 2001, pp. 556–562.
- [2] Tang, J., Gao, H., Hu, X. *et al.*, “Exploiting homophily effect for trust prediction,” in *Proceedings of WSDM’13*, 2013, pp. 53–62.
- [3] Matsutani, K., Kumano, M., Kimura, M. *et al.*, “Combining activity-evaluation information with NMF for trust-link prediction in social media,” in *Proceedings of BigData’15*, 2015, pp. 2263–2272.