

アクティビティ評価情報とNMFの融合による ソーシャルメディアの信頼リンク予測

Combining Activity-evaluation Information with NMF for Trust-link Prediction in Social Media

松谷 貫司¹ 熊野 雅仁¹ 木村 昌弘^{1*}
Kanji Matsutani¹ Masahito Kumano¹ Masahiro Kimura¹
斉藤 和巳² 大原 剛三³ 元田 浩⁴
Kazumi Saito² Kouzou Ohara³ Hiroshi Motoda⁴

¹ 龍谷大学理工学部電子情報学科

¹ Department of Electronics and Informatics, Ryukoku University

² 静岡県立大学経営情報学部

² School of Administration and Informatics, University of Shizuoka

³ 青山学院大学理工学部情報テクノロジー学科

³ Department of Integrated Information Technology, Aoyama Gakuin University

⁴ 大阪大学産業科学研究所

⁴ Institute of Scientific and Industrial Research, Osaka University

Abstract: Acquiring networks of trust relations among users in social media sites such as item-review sites is important for analyzing users' behaviour and efficiently finding reliable information on the Web. For an item-review site, we address the problem of predicting trust-links among users. Recently, non-negative matrix factorization (NMF) methods have been shown to be useful for trust-link prediction in such an site, where both the link and activity information is employed. Here, a user activity in an item-review site means posting a review and giving a rating for an item. Aiming to improve NMF methods for trust-link prediction, in this paper, we propose such an NMF method that incorporates information of people's evaluations for users' activities as well as information of trust-links and users' activities. Also, we apply it to an analysis of users' behaviour. Using real data of an item-review site, we experimentally demonstrate the effectiveness of the proposed method.

1 はじめに

近年, Digg, eBay, Epinions, Facebook などのソーシャルメディアの普及に伴い, オンライン世界に大規模ネットワークが生成されている. ソーシャルメディアユーザの間の信頼ネットワークはソーシャルネットワークの一種であり, Web 上で信頼できる情報を効良く発見することに役立つ. ソーシャルメディアサイトに投稿されたニュースや意見は, ソーシャルネットワークを介して急速かつ広域的に拡散し, 多数の人々に共有されるようになる. このように, 信頼ネットワークは人々の日々の生活の中で重要な役割を果たす. したがって, 信頼ネッ

トワーク(ソーシャルネットワーク)のマイニングに対して, 情報拡散分析 [1], [2], [3], [4], [5], [6], [7] や, ソーシャルリンク予測 [8], [9], [10], [11], [12], [13], [14] などの様々な研究が行われている.

ソーシャルメディアサイト上のユーザの行動を分析するためには, そのサイトにおけるユーザ間の信頼関係により構築されるネットワークを獲得することが必要不可欠である. しかし, 信頼ネットワークは時間経過と共に継続的に進化し, かつ, プライバシーの問題も存在することから, 分析に必要な特定の時点についての完全な信頼ネットワーク構造を獲得することは一般に困難である. また, オンライン世界において, 人々の行動パターンや嗜好は長期間では大きく変化すると考えられる. そこで, 本論文では, 直近の過去において活動的なユーザを対象に, 直近の未来に生成される信頼リンクを予測す

*連絡先: 〒520-2194 大津市瀬田大江町横谷 1-5
龍谷大学理工学部電子情報学科
E-mail: kimura@rins.ryukoku.ac.jp

る問題を論じ、この問題の効率的な解法を提案する。ここに、我々の提案法は未観測リンクの予測問題にも適用可能であることを注意しておく。

多くのソーシャルメディアサイトでは、信頼リンク生成機能とともに、アクティビティ実行機能が提供されており、ユーザは与えられたアクティビティ群の中から一つを選択して実行することができる。たとえば、アイテムレビューサイトにおいて、ユーザは与えられたアイテム群の各アイテムに対してレビューの投稿および評価付けを行うことができる。信頼リンクとアクティビティを提供するソーシャルメディアサイトにおける信頼リンク予測問題に対して、Tang ら [14] はリンクとアクティビティの情報を利用し、類同性 (*homophily*) と呼ばれる社会学の理論を利用した非負値行列分解 (NMF) 法を提案した。ここに、類同性理論では類似するユーザ (類似するアクティビティを実行したユーザ) ほど信頼関係を確立する可能性が高いことが示唆されている [10], [15]。 *hTrust* と呼ばれる彼らの手法は NMF に基づいており、NMF 法 [16] は協調フィルタリング、文書クラスタリング、リンク予測などの多くの応用に対して有効であることが示されている。彼らは、 *hTrust* が従来法より高性能であることをアイテムレビューサイトの実データを用いた実験で示した [14]。

Tang ら [14] と同様に、我々もアイテムレビューサイトに焦点を当てる。最近の多くのアイテムレビューサイトでは、あるユーザのアイテムへのレビューに対して、別のユーザがそれを有益であると思ったときに感謝メッセージを投稿することができる。信頼リンク予測に対する NMF の性能向上のために、我々はそのような情報と NMF との融合を考える。本論文では、信頼リンク予測問題に対して、信頼リンクとユーザアクティビティの情報とともに、ユーザアクティビティの評価情報を組み込んだ新たな NMF を提案し、それをソーシャルメディアサイトにおけるユーザ行動の分析に応用する。また、アイテムレビューサイトの実データを用いて提案法を評価する。まず、データセットの統計解析を行い、特に、受け取った感謝メッセージ数と信頼リンクの入次数との間に相関があることを示す。次に、信頼リンク予測問題に対して、提案法が *hTrust* とそれに類似した手法に比べて高性能であることを実証する。さらに、信頼リンク生成の観点から、データセットにおけるユーザ行動を分析する。

本論文の構成は以下の通りである。2 節では、信頼リンク予測問題を定式化し、信頼リンク予測に対してリンク情報とアクティビティ情報を用いた NMF 法を説明する。3 節では、提案 NMF 法を示し、4 節において評価実験の結果を報告する。5 節では、主たる結果を要約し、本論文の結論を述べる。

2 NMF による信頼リンク予測

Tang ら [14] はソーシャルメディアにおける信頼リンク予測に対して、リンク情報とアクティビティ情報を用いた非負値行列分解 (NMF) 法である *hTrust* を提案し、その有効性を示した。また、類似する NMF 法が Zhu ら [17] によって提案されている。彼らは Web ページの分類に対してリンクと文書の情報を用いた NMF 法である *Joint Link-Content Matrix Factorization (JLCMF)* を提案した。本節では、ソーシャルメディアにおける我々の信頼リンク問題を定式化したのち、 *hTrust* と *JLCMF* を再訪する。

2.1 問題の定式化

本論文では、信頼リンクとアクティビティを提供するソーシャルメディアサイトに対して、直近の過去において活動的なユーザを対象に、直近の未来に生成される信頼リンクを予測する問題を扱う。ここに、このようなサイトにおけるユーザアクティビティとは、与えられたアイテム群の各アイテムに対してレビュー投稿および評価付けを行うことであるとする。さらに、サイト内のアイテムにユーザが行うレビューに対して、別のユーザがそれを有益であると思ったときに感謝メッセージを投稿することができる。つまり、ユーザアクティビティの評価情報をサイト内にて観測することができる。

我々は、直近の Δt_0 か月内 (観測期間 I_0 と呼ぶ) に少なくとも n_1 本の信頼リンクを結び、かつ、少なくとも n_2 本の信頼リンクが結ばれたような活動的なユーザの集合 \mathcal{V} に焦点を当てる。ここでは、期間 I_0 はさほど長くはない (たとえば $\Delta t_0 = |I_0|$ が 5, 6 程度) と仮定する。 $\mathcal{V} = \{v_1, \dots, v_N\}$ とし、 \mathcal{V} において現在までに生成された信頼リンクを表す $N \times N$ 行列を $G = (G_{i,j})_{i,j=1,\dots,N}$ とおく。ここに、現在までにユーザ v_i からユーザ v_j への信頼リンクが生成されているならば $G_{i,j} = 1$ 、さもなければ $G_{i,j} = 0$ である。本論文では、予測期間 I_1 が短い (たとえば Δt_1 が 2, 3 程度) という仮定の下に、次の Δt_1 か月内に \mathcal{V} において生成される信頼リンクを予測する問題を扱う。ここに、 $\Delta t_0 = |I_0|$ と $\Delta t_1 = |I_1|$ の値は、ユーザの行動パターンは短い期間 (たとえば数か月) においては大きく変化しないという事実が考慮されていることに注意しておく。特に、我々はこの予測問題を行列 G を分解する NMF アプローチにより解くことを考える。

信頼リンク予測問題では、リンク情報の他にも考慮すべき情報がある。 \mathcal{V} に属するユーザが現在までにレビュー投稿や評価付けを行ったアイテムの集合を $\mathcal{A} = \{a_1, \dots, a_M\}$ とし、現在までのユーザアクティビティを表す $N \times M$ 行列を $X = (X_{i,\alpha})$ とする。ここに、 $X_{i,\alpha}$ はユーザ v_i がアイテム a_α に対して付けた評点である。ただし、ユーザ v_i がアイテム a_α に対してレビューを投稿

しなかった場合は $X_{i,\alpha} = 0$ とする。さらに、 \mathcal{V} における現在までのユーザアクティビティに対する評価の情報を表す $N \times M$ 行列を $Y = (Y_{i,\alpha})$ とする。ここに、 $Y_{i,\alpha}$ は、ユーザ v_i がアイテム a_α のレビューに対して得た感謝メッセージの数である。 $hTrust$ と $JLCMF$ は行列 G と X のみを利用したのに対し、本論文では行列 G, X および Y を利用する。

2.2 hTrust

NMF アプローチは潜在空間を用いて有効な特徴量表現を求めるものである。 $hTrust$ は単一の潜在空間を用いる。潜在空間の次元を K とする。 $hTrust$ では非負値 $N \times K$ 行列 $U = (U_{i,k})$ と、非負値 $K \times K$ 行列 $H = (H_{k,\ell})$ を導入する。ここに、 $U_{i,k}$ は潜在因子 k に対するユーザ v_i の強さを、 $H_{k,\ell}$ は信頼リンク生成に対する潜在因子 k から潜在因子 ℓ への関係の強さをそれぞれ表す。 U と H の関数

$$\mathcal{F}_0(U, H) = \|G - UHU^T\|^2 + \lambda_U \|U\|^2 + \lambda_H \|H\|^2 + \lambda_X \text{Tr}(U^T S_X U) \quad (1)$$

の最小化を考える。ここに、 λ_U, λ_H および λ_X は正定数(超パラメータ)である。ただし、任意の行列 B に対して、 $B^T, \|B\|$ および $\text{Tr}(B)$ はそれぞれ B の転置行列、 B のフロベニウスノルム、および B のトレースを表す。 $S_X = ((S_X)_{i,j})$ は、

$$(S_X)_{i,j} = \bar{\xi}_i \delta_{i,j} - \xi_{i,j} \quad (2)$$

で定義される $N \times N$ 対称行列である。ここに、 $\delta_{i,j}$ はクロネッカーのデルタであり、

$$\xi_{i,j} = \frac{\sum_{\alpha=1}^M X_{i,\alpha} X_{j,\alpha}}{\sqrt{\sum_{\alpha=1}^M X_{i,\alpha}^2} \sqrt{\sum_{\alpha=1}^M X_{j,\alpha}^2}} \quad (3)$$

かつ、

$$\bar{\xi}_i = \sum_{j=1}^N \xi_{i,j} \quad (4)$$

である。ただし、 $\text{Tr}(U^T S_X U)$ は類同性正則化 (*homophily regularization*) [14] と呼ばれる項であり、

$$\text{Tr}(U^T S_X U) = \frac{1}{2} \sum_{i,j=1}^N \xi_{i,j} \sum_{k=1}^K (U_{i,k} - U_{j,k})^2 \quad (5)$$

であることに注意しておく。この項は類似度の高いユーザ同士はより信頼関係を確立しやすいという性質を組み込むために導入されている。ここに、 $\xi_{i,j}$ はアクティビティに関するユーザ v_i と v_j の間の類似度である。 $\xi_{i,j}$ については別の尺度の選択が可能であり、Jaccard 係数や Pearson 相関係数を含む様々な尺度を使うことができる

ことに注意しておく。非負値拘束 $U \geq 0, H \geq 0$ の下での関数 $\mathcal{F}_0(U, H)$ の最小化問題に対して、 $hTrust$ は U と H の反復更新アルゴリズムを提案した(詳細については [14] を参照されたい)。そのアルゴリズムにより得られた最適な U と H をそれぞれ $U^* = (U_{i,k}^*), H^* = (H_{k,\ell}^*)$ とする。我々は $G_{i,j} = 0$ であるペア (i, j) をリンク候補と呼ぶ。 $hTrust$ は $G_{i,j}^* = \sum_{k,\ell=1}^K U_{i,k}^* H_{k,\ell}^* U_{j,\ell}^*$ の値に従ってリンク候補 (i, j) をランキングすることにより、期間 I_1 に生成される信頼リンクを予測する。

2.3 Joint Link-Content Matrix Factorization

$JLCMF$ も単一の潜在空間を利用する。潜在空間の次元を K とする。 $JLCMF$ では非負値 $N \times K$ 行列 $U = (U_{i,k})$ 、非負値 $K \times K$ 行列 $H = (H_{k,i})$ 、および非負値 $M \times K$ 行列 $\Phi = (\Phi_{\alpha,k})$ を導入する。ここに、 $\Phi_{\alpha,k}$ はアイテムへのレビュー投稿の観点における、アイテム a_α と潜在因子 k の間の関係の強さを表す。 U, H および Φ の関数

$$\mathcal{F}_1(U, H, \Phi) = \|G - UHU^T\|^2 + \lambda_H \|H\|^2 + \lambda_\Phi \|\Phi\|^2 + \lambda_X \|X - U\Phi^T\|^2 \quad (6)$$

の最小化を考える。ここに、 λ_H, λ_Φ および λ_X は正定数(超パラメータ)である。非負値拘束 $U \geq 0, H \geq 0$ の下での $\mathcal{F}_1(U, H, \Phi)$ の最小化問題に対して、 $hTrust$ [14] と通常の NMF [16] を応用することにより、 U, H および Φ の反復更新アルゴリズムを容易に得ることができる。 $JLCMF$ は Web ページの分類を目的としているが、信頼リンク予測に対しても適用可能である。したがって、本論文では $JLCMF$ を信頼リンク予測に対してリンクとアクティビティの両方を利用した比較の NMF 法とする。

3 提案法

我々は信頼リンク情報、ユーザアクティビティ情報、およびユーザアクティビティに対する評価情報を組み込んだ新たな NMF 法、すなわち、行列 G, X 、および Y を用いた NMF 法を提案する。まず、信頼リンク予測に対する新たな NMF 法を提案する。次に、提案 NMF 法に対する最適化アルゴリズムを導出する。最後に、ユーザの行動分析への提案 NMF 法の応用法を示す。

3.1 信頼リンク予測のための NMF 法

我々はユーザが好む分野の概念と、ユーザが信頼を獲得する分野の概念を区別することを考える。前者の分野を P 分野、後者の分野を T 分野と呼ぶ。 $hTrust$ や $JLCMF$ とは異なり、提案 NMF モデルは 2 つの潜在空間を用いる。一つは P 分野の空間 (PF 空間と呼ぶ) に対応し、も

う一つは T 分野の空間 (TF 空間と呼ぶ) に対応する。さらに, P 分野と T 分野の各因子をそれぞれ潜在 P 因子と潜在 T 因子と呼ぶ。

PF 空間の次元を K とし, TF 空間の次元を L とする。提案 NMF モデルでは非負値 $N \times K$ 行列 $U = (U_{i,k})$, 非負値 $N \times L$ 行列 $W = (W_{i,\ell})$, および非負値 $K \times L$ 行列 $H = (H_{k,\ell})$ を導入する。ここに, $U_{i,k}$ は潜在 P 因子 k に対するユーザ v_i の強度を表し, $W_{i,k}$ は潜在 T 因子 k に対するユーザ v_i の強度を表す。さらに, $H_{k,\ell}$ は信頼リンク生成に対する潜在 P 因子 k から潜在 T 因子 ℓ への関係の強さを表す。 U, W および H の関数

$$\mathcal{F}(U, W, H) = \|G - UHW^T\|^2 + \lambda_U \|U\|^2 + \lambda_W \|W\|^2 + \lambda_H \|H\|^2 + \lambda_X \text{Tr}(U^T S_X U) + \lambda_Y \text{Tr}(W^T S_Y W) \quad (7)$$

の最小化を考える。ここに, $\lambda_U, \lambda_W, \lambda_H, \lambda_X$, および λ_Y は正定数 (超パラメータ) である。 $S_X = ((S_X)_{i,j})$ は式 (2), (3) および (4) によって定義される $N \times N$ 対称行列である。 $S_Y = ((S_Y)_{i,j})$ は, $(S_Y)_{i,j} = \bar{\eta}_i \delta_{i,j} - \eta_{i,j}$ で定義される $N \times N$ 対称行列である。ここに,

$$\eta_{i,j} = \frac{\sum_{\alpha=1}^M Y_{i,\alpha} Y_{j,\alpha}}{\sqrt{\sum_{\alpha=1}^M Y_{i,\alpha}^2} \sqrt{\sum_{\alpha=1}^M Y_{j,\alpha}^2}}$$

かつ,

$$\bar{\eta}_i = \sum_{j=1}^N \eta_{i,j}$$

である。ただし,

$$\text{Tr}(W^T S_Y W) = \frac{1}{2} \sum_{i,j=1}^N \eta_{i,j} \sum_{\ell=1}^L (W_{i,\ell} - W_{j,\ell})^2 \quad (8)$$

である。提案法は正則化項 $\text{Tr}(U^T S_X U)$ と $\text{Tr}(W^T S_Y W)$ により, ユーザアクティビティとユーザアクティビティに対する評価情報を組み込む。ここに, $\xi_{i,j}$ はアクティビティ評価に関するユーザ v_i, v_j 間の類似度であり, $\eta_{i,j}$ はアクティビティ評価に関するユーザ v_i, v_j 間の類似度である。 $\xi_{i,j}$ と $\eta_{i,j}$ については他の尺度の選択が可能であることに再度注意しておく。式 (5) と (8) は, アクティビティの類似するユーザは潜在 P 因子の点で類似する表現を持ち, また, アクティビティ評価の類似するユーザは潜在 T 因子の点で類似する表現を持つということの意味する。さらに, 項 $\|U\|^2, \|W\|^2$ および $\|H\|^2$ は過学習を防ぐための平滑正則化として加えられている。

拘束条件 $U \geq 0, W \geq 0$ および $H \geq 0$ の下での $\mathcal{F}(U, W, H)$ の最小化のために, U, W および H の反復更新アルゴリズム (3.2 節参照) を示す。 $U^* = (U_{i,k}^*)$, $W^* = (W_{i,\ell}^*)$ および $H^* = (H_{k,\ell}^*)$ をそれぞれそのアルゴリズムにより得られた U, W および H の最適値とする。提

案法はリンク候補 (i, j) を $G_{i,j}^* = \sum_{k=1}^K \sum_{\ell=1}^L U_{i,k}^* H_{k,\ell}^* W_{j,\ell}^*$ の値に従ってランキングすることにより, 期間 I_1 に生成される信頼リンクを予測する。

3.2 最適化アルゴリズム

非負値 $N \times K$ 行列 $U = (U_{i,k})$, 非負値 $N \times L$ 行列 $W = (W_{i,\ell})$ および非負値 $K \times L$ 行列 $H = (H_{k,\ell})$ に対して, 式 (7) で定義される関数 $\mathcal{F}(U, W, H)$ の最小化問題を考える。本節ではこの最適化問題を解くための U, W および H の反復更新アルゴリズムを提案する。以降, 任意の行列 B に対して, $B_{p,q}$ は B の (p, q) 成分を示すものとする。

\hat{U}, \hat{W} および \hat{H} をそれぞれ U, W および H の現在の推定値とする。以下において U, W および H の更新式を導出する。まず, 3つの補助関数

$$\begin{aligned} \mathcal{E}_U(U, \hat{U}; W, H) = & \|G\|^2 + \sum_{i=1}^N \sum_{k=1}^K \frac{(\hat{U}HW^TWH^T)_{i,k}}{\hat{U}_{i,k}} U_{i,k}^2 \\ & - 2 \sum_{i=1}^N \sum_{k=1}^K (GWH^T)_{i,k} \hat{U}_{i,k} \left(1 + \log \frac{U_{i,k}}{\hat{U}_{i,k}}\right) + \lambda_U \|U\|^2 \\ & + \lambda_X \sum_{i,j=1}^N \xi_{i,j} \sum_{k=1}^K \left\{ U_{i,k}^2 - \hat{U}_{i,k} \hat{U}_{j,k} \left(1 + \log \frac{U_{i,k} U_{j,k}}{\hat{U}_{i,k} \hat{U}_{j,k}}\right) \right\} \\ & + \lambda_W \|W\|^2 + \lambda_H \|H\|^2 + \lambda_Y \text{Tr}(W^T S_Y W), \quad (9) \end{aligned}$$

$$\begin{aligned} \mathcal{E}_W(W, \hat{W}; U, H) = & \|G\|^2 + \sum_{i=1}^N \sum_{\ell=1}^L \frac{(\hat{W}H^T U^T UH)_{i,\ell}}{\hat{W}_{i,\ell}} W_{i,\ell}^2 \\ & - 2 \sum_{i=1}^N \sum_{\ell=1}^L (G^T UH)_{i,\ell} \hat{W}_{i,\ell} \left(1 + \log \frac{W_{i,\ell}}{\hat{W}_{i,\ell}}\right) + \lambda_W \|W\|^2 \\ & + \lambda_Y \sum_{i,j=1}^N \eta_{i,j} \sum_{\ell=1}^L \left\{ W_{i,\ell}^2 - \hat{W}_{i,\ell} \hat{W}_{j,\ell} \left(1 + \log \frac{W_{i,\ell} W_{j,\ell}}{\hat{W}_{i,\ell} \hat{W}_{j,\ell}}\right) \right\} \\ & + \lambda_U \|U\|^2 + \lambda_H \|H\|^2 + \lambda_X \text{Tr}(U^T S_X U), \quad (10) \end{aligned}$$

$$\begin{aligned} \mathcal{E}_H(H, \hat{H}; U, W) = & \|G\|^2 + \sum_{k=1}^K \sum_{\ell=1}^L \frac{(U^T U \hat{H} W^T W)_{k,\ell}}{\hat{H}_{k,\ell}} H_{k,\ell}^2 + \lambda_H \|H\|^2 \\ & - 2 \sum_{k=1}^K \sum_{\ell=1}^L (U^T G W)_{k,\ell} \hat{H}_{k,\ell} \left(1 + \log \frac{H_{k,\ell}}{\hat{H}_{k,\ell}}\right) + \lambda_U \|U\|^2 \\ & + \lambda_W \|W\|^2 + \lambda_X \text{Tr}(U^T S_X U) + \lambda_Y \text{Tr}(W^T S_Y W). \quad (11) \end{aligned}$$

を定義する。このとき, 以下の不等式と等式が成り立つ

ことは容易に証明される.

$$\begin{aligned}
\mathcal{E}_U(U, \hat{U}; W, H) &\geq \mathcal{F}(U, W, H), \\
\mathcal{E}_W(W, \hat{W}; U, H) &\geq \mathcal{F}(U, W, H), \\
\mathcal{E}_H(H, \hat{H}; U, W) &\geq \mathcal{F}(U, W, H), \\
\mathcal{F}(U, W, H) &= \mathcal{E}_U(U, U; W, H) = \mathcal{E}_W(W, W; U, H) \\
&= \mathcal{E}_H(H, H; U, W) \quad (12)
\end{aligned}$$

したがって, 式 (9) と (12) から, $\mathcal{E}_U(U, \hat{U}; W, H)$ を U について最小化することで, $U = (U_{i,k})$ に対する以下の更新式が導出される.

$$U_{i,k} = \hat{U}_{i,k} \sqrt{\frac{(GWH^T)_{i,k} + \lambda_X \sum_{j=1}^N \xi_{i,j} \hat{U}_{j,k}}{(\hat{U}HW^TWH^T)_{i,k} + \lambda_U \hat{U}_{i,k} + \lambda_X \bar{\xi}_i \hat{U}_{i,k}}} \quad (13)$$

また, 式 (10) と (12) から, $\mathcal{E}_W(W, \hat{W}; U, H)$ を W について最小化することで, $W = (W_{i,\ell})$ に対する以下の更新式が導出される.

$$W_{i,\ell} = \hat{W}_{i,\ell} \sqrt{\frac{(G^T UH)_{i,\ell} + \lambda_Y \sum_{j=1}^N \eta_{i,j} \hat{W}_{j,\ell}}{(\hat{W}H^T U^T UH)_{i,\ell} + \lambda_W \hat{W}_{i,\ell} + \lambda_Y \bar{\eta}_i \hat{W}_{i,\ell}}} \quad (14)$$

同様に, 式 (11) と (12) から, $\mathcal{E}_H(H, \hat{H}; U, W)$ を H について最小化することで, $H = (H_{k,\ell})$ に対する以下の更新式が導出される.

$$H_{k,\ell} = \hat{H}_{k,\ell} \sqrt{\frac{(U^T GW)_{k,\ell}}{(U^T U \hat{H} W^T W)_{k,\ell} + \lambda_H \hat{H}_{k,\ell}}} \quad (15)$$

次に, 更新式 (13), (14), (15) を繰り返し用いて導かれる, U, W および H の反復更新アルゴリズムを説明する. 任意の非負整数 t に対して, U, W および H の t 回目の更新値をそれぞれ U_t, W_t および H_t とする. U_{t+1} を

$$U_{t+1} = \arg \min_U \mathcal{E}_U(U, U_t; W_t, H_t)$$

により定義すると, これは式 (13) から得られる. また, W_{t+1} を

$$W_{t+1} = \arg \min_W \mathcal{E}_W(W, W_t; U_{t+1}, H_t)$$

と定義すると, これは式 (14) から得られる. 同様に, H_{t+1} を

$$H_{t+1} = \arg \min_H \mathcal{E}_H(H, H_t; U_{t+1}, W_{t+1})$$

と定義すると, これは式 (15) から得られる. このとき, 式 (12) より,

$$\begin{aligned}
\mathcal{F}(U_t, W_t, H_t) &= \mathcal{E}_U(U_t, U_t; W_t, H_t) \geq \mathcal{E}_U(U_{t+1}, U_t; W_t, H_t) \\
&\geq \mathcal{F}(U_{t+1}, W_t, H_t) = \mathcal{E}_W(W_t, W_t; U_{t+1}, H_t) \\
&\geq \mathcal{E}_W(W_{t+1}, W_t; U_{t+1}, H_t) \geq \mathcal{F}(U_{t+1}, W_{t+1}, H_t) \\
&= \mathcal{E}_H(H_t, H_t; U_{t+1}, W_{t+1}) \geq \mathcal{E}_H(H_{t+1}, H_t; U_{t+1}, W_{t+1}) \\
&\geq \mathcal{F}(U_{t+1}, W_{t+1}, H_{t+1})
\end{aligned}$$

を得る. したがって, 目的関数 $\mathcal{F}(U, W, H)$ はこの反復更新アルゴリズムにより単調に減少して行く. ゆえにこのアルゴリズムは収束する.

3.3 行動分析への応用

提案アルゴリズムにより得られた U, W および H , すなわち $U^* = (U_{i,k}^*)$, $W^* = (W_{i,\ell}^*)$ および $H^* = (H_{k,\ell}^*)$ を用いて, 我々は対象とするソーシャルメディアサイトにおけるユーザの行動分析を行う. 潜在 P 因子 (すなわち, P 分野) および潜在 T 因子 (すなわち, T 分野) を調べるために, 非負値 $M \times K$ 行列 $\Phi = (\Phi_{\alpha,k})$ および非負値 $M \times L$ 行列 $\Psi = (\Psi_{\alpha,\ell})$ を導入する. ただし, $\Psi_{\alpha,k}$ はアイテムに対するレビュー投稿の観点からのアイテム a_α と潜在 P 因子 k の間の関係の強さを表し, また, $\Psi_{\alpha,\ell}$ はアイテムのレビューに対して感謝メッセージを受け取るという観点からのアイテム a_α と潜在 T 因子 ℓ の間の関係の強さを表す. ここで, 以下の関数

$$\mathcal{G}_X(\Phi) = \|X - U^* \Phi^T\|^2 + \lambda_\Phi \|\Phi\|^2$$

および,

$$\mathcal{G}_Y(\Psi) = \|Y - W^* \Psi^T\|^2 + \lambda_\Psi \|\Psi\|^2$$

の最小化を考える. ただし, λ_Φ と λ_Ψ は正定数 (超パラメータ) である. これらの最小化問題に対して, 3.2 節と同様の方法で Φ と Ψ の反復更新アルゴリズムを導出することができる. これらのアルゴリズムにより得られた Φ と Ψ の最適値をそれぞれ $\Phi = (\Phi_{\alpha,k})$ と $\Psi = (\Psi_{\alpha,\ell})$ とする. Φ^* と Ψ^* を用いて, 我々はアイテムに関する潜在 P 因子と潜在 T 因子の解釈を試みる. そして, それらの解釈に基づき, 信頼リンク生成の観点からサイトにおけるユーザの行動を分析する.

4 評価実験

アイテムレビューサイト Epinions の実データを用いて提案法を評価する. まず, データセットの統計解析から始め, その後, 2.1 節に示した信頼リンク予測問題の解法としての提案法の評価を行う. 次に, 提案法を応用することで, 信頼リンク生成の観点からユーザ行動の性質を分析する.

4.1 ソーシャルメディアデータ

我々はアイテムレビューサイト Epinions¹ の実データを収集した. ここに, Epinions は製品レビューと消費者

¹<http://www.epinions.com/>

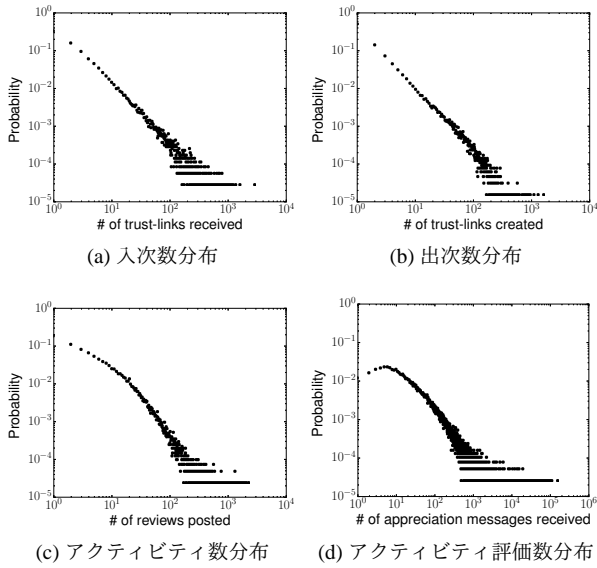


図 1: Epinions データの基本統計解析.

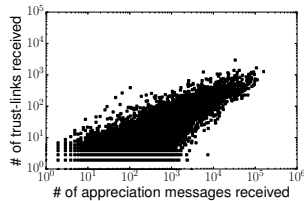


図 2: アクティビティ評価数と入次数の相関.

報告のソーシャルメディアサイトである. Epinions において, ユーザは他のユーザへの信頼リンクの生成のみならず, 与えられたアイテム群のアイテムに対してレビュー投稿や評点付けを行うことができる. さらに, あるユーザのアイテムへのレビューに対して, 別のユーザはそれを有益であると思ったときに感謝メッセージを投稿することができる. Epinions に関して, 我々は 2012 年 10 月に最人気ユーザとして紹介されたユーザを起点として, 新たなユーザが出現しなくなるまで信頼リンクを幅優先探索し, 信頼リンク, レビューと評点, および感謝メッセージのデータを収集した. 収集したデータは, ユーザ数 64,268 人, 信頼リンク数 509,293 本, アイテム数 268,891 個, レビュー数 809,517 件および感謝メッセージ数 18,960,792 件であった.

Epinions データに対して, まず, 基本的な統計的性質を分析するために, 入次数分布 (ユーザが受け取った信頼リンク数の割合), 出次数分布 (ユーザが生成した信頼リンクの割合), アクティビティ分布 (ユーザが投稿したレビュー数の割合) およびアクティビティ評価分布 (ユーザが受け取った感謝メッセージ数の割合) を解析した. 図 1 にその結果を示す. いずれの分布も後部がべき乗則に従うことが観測される. これらの結果は, Epinions デー

タがオンライン世界におけるソーシャルデータの典型的な性質を満足することを示している. 次に, アクティビティ評価数と入次数の間の相関を調査した. 図 2 にその結果を示す. ユーザが受け取った感謝メッセージ数とユーザが受け取った信頼リンク数との間には正の相関があることが分かる. これは, 信頼リンク予測の NMF 法がアクティビティ評価情報を組み込むことにより改善される可能性を示唆している.

4.2 実験設定

Epinions データを用いて, 信頼リンク予測問題 (2.1 節参照) の解法である提案法の性能を評価した. Epinions データに対して, 4 つのデータセット D_1, D_2, D_3 および D_4 を以下のように構築した. まず, $n_1 = n_2 = 1, \Delta t_0 = 6$, および $\Delta t_1 = 3$ (2.1 節参照) と設定した. 評価期間 I_1 は, D_1 に対しては 1 月から 3 月, D_2 に対しては 4 月から 6 月, D_3 に対しては 7 月から 9 月, D_4 に対しては 10 月から 12 月とした. 2006 年には一定数以上の信頼リンクが常に生成されていたことからその年に生成された信頼リンクの予測を試みた. ここに, たとえば, データセット D_1 に対する I_1 は 2006 年の 1 月から 3 月であり, 観測期間 I_0 は 2005 年の 7 月から 12 月である. 表 1 は各データセットの基本統計である.

表 1: データセットの基本統計.

	D_1	D_2	D_3	D_4
ユーザ数, N	771	722	734	727
アイテム数, M	56,642	57,886	59,522	61,327
観測信頼リンク数	27,154	25,382	26,096	26,581
アクティビティ数	83,786	83,933	86,396	88,980
アクティビティ評価数	1,074,042	1,039,505	1,127,326	1,170,950
I_1 に生成された信頼リンク数	1,670	1,308	1,441	1,684

信頼リンク予測問題に対して, 我々は提案法と $hTrust$ (式 (1) 参照) および $JLCMF$ (式 (6) 参照) を比較する. また, $JLCMF$ を拡張することで提案法に類似する手法を定義し, 提案法との比較を行う. ここに, 非負値 $N \times K$ 行列 U , 非負値 $N \times L$ 行列 W , 非負値 $K \times L$ 行列 H , 非負値 $M \times K$ 行列 Φ および非負値 $M \times L$ 行列 Ψ に対して, 以下の関数

$$\mathcal{F}_2(U, W, H, \Phi, \Psi) = \|G - UHW^T\|^2 + \lambda_U \|U\|^2 + \lambda_W \|W\|^2 + \lambda_H \|H\|^2 + \lambda_X \|X - U\Phi^T\|^2 + \lambda_Y \|Y - W\Psi^T\|^2 + \lambda_\Phi \|\Phi\|^2 + \lambda_\Psi \|\Psi\|^2$$

の最小化を考える. ただし, $\lambda_U, \lambda_W, \lambda_H, \lambda_X, \lambda_Y, \lambda_\Phi$ および λ_Ψ は正定数 (超パラメータ) である. この最小化問題に対して, 提案法 (3.2 節参照) と同様の導出により U, W, H, Φ および Ψ の反復更新アルゴリズムが容易に得られる. 我々はこの手法を $JLCMF2$ と呼ぶ.

各手法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が、本論文では [14] に従って単純に $\lambda_U = \lambda_W = \lambda_H = \lambda_\Phi = \lambda_\Psi = 0.01$, $K = L = 10$, $\lambda_X = \lambda_Y = 10$ と設定した².

4.3 信頼リンク予測

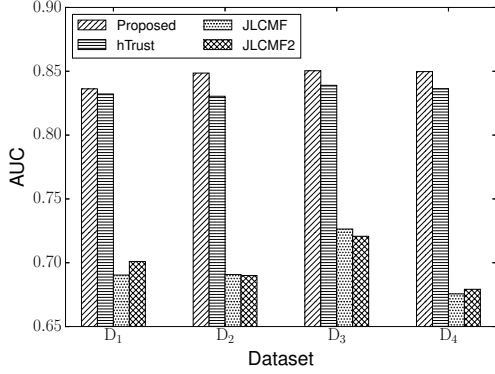


図 3: 信頼リンク予測性能の比較.

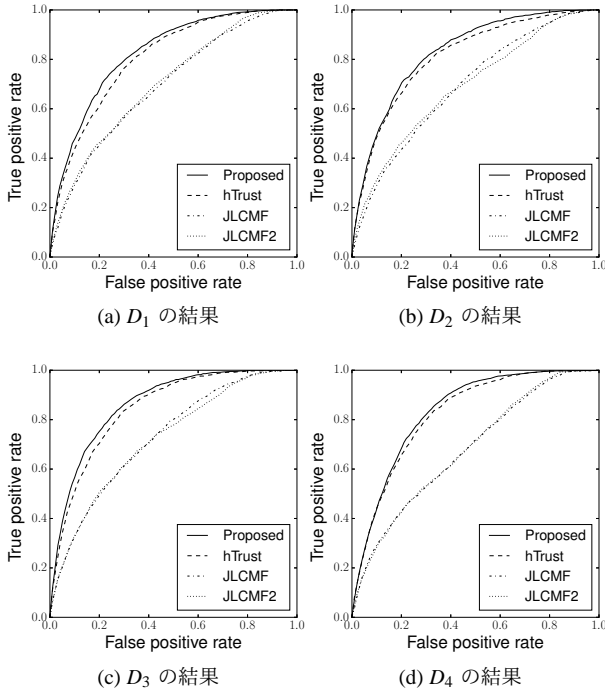


図 4: ROC 曲線.

信頼リンク予測問題 (2.1 節参照) に対して、提案法と *hTrust*, *JLCMF* および *JLCMF2* を比較する。実験では、ROC 曲線下の面積 (AUC) で予測性能を測定した。図 3 にその結果を示す。図 4 は ROC 曲線の結果である。提

²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。

案法が最も高性能であり、*hTrust* がそれに続いていることが観察される。性能差はデータ依存であるが、*JLCMF* と *JLCMF2* はこれら 2 つの手法を常に下回っていた。これらの結果はアクティビティ評価情報を組み込むことの重要性を示しており、2 種類の潜在因子 (すなわち、2 つの潜在空間) である潜在 P 因子および潜在 T 因子を適切に融合した提案法の有効性を実証している。

4.4 行動分析

提案法を応用し、*Epinions* での信頼リンク生成に対するユーザ行動の分析を行う。ここでは、データセット D_3 に対する分析結果のみを報告する。

図 5 に $H^* = (H_{k,\ell}^*)$ の可視化結果を示す。 k 行 ℓ 列の明るさは、潜在 P 因子 k から潜在 T 因子 ℓ への関係の強さ $H_{k,\ell}^*$ の値を表している。 $H_{k,\ell}^*$ は $(k, \ell) = (6, 6)$ において最大であった。また、 $\Phi^* = (\Phi_{\alpha,k}^*)$ と $\Psi^* = (\Psi_{\alpha,\ell}^*)$ を用いて、潜在 P 因子 $\{k\}$ と潜在 T 因子 $\{\ell\}$ がどのようなものを調べた。その結果、潜在 P 因子と潜在 T 因子の間には明らかな違いが見られたことに注意しておく。

表 2 と表 3 にそれぞれ $\Phi_{\alpha,6}^*$ と $\Psi_{\alpha,6}^*$ に関する上位 4 アイテム $\{\alpha\}$ を示す。これらは潜在 P 因子 $k = 6$ と潜在 T 因子 $\ell = 6$ を例示していることに注意しておく。したがって、*Epinions* のデータセット D_3 では、テレビゲームと SF 映画を好むユーザから SF・ファンタジーで信頼を獲得するユーザへと信頼リンクが結ばれる傾向があることが分かる。

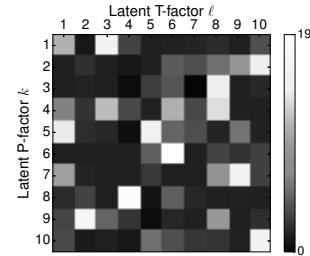


図 5: D_3 における $H^* = (H_{kl}^*)$ の可視化.

5 結論

本論文では、アイテムレビューサイト (ソーシャルメディアサイト) に対して、直近の過去において活動的な

表 2: D_3 における潜在 P 因子の例 ($k = 6$).

Item	Category
Nintendo Game Cube White Console	Video Game Consoles
Sony PlayStation 2 Slimline Console	Video Game Consoles
Star Wars Episode III: Revenge of the Sith	Movies
Sega Dreamcast Grey Console	Video Game Consoles

表 3: D_3 における潜在 T 因子の例 ($\ell = 6$).

Item	Category
Lord of the Rings: The Return of the King	Movies
Star Wars Episode III: Revenge of the Sith	Movies
The Incredibles by Pixar & Disney	Movies
Lord of the Rings: The Fellowship of the Ring	Movies

ユーザを対象に、直近の未来に生成される信頼リンクを予測する問題を扱った。近年 NMF 法を用いた $hTrust$ の有効性が示されたことから、我々はこの問題に対して NMF アプローチを採用した。 $hTrust$ は信頼リンク情報とユーザアクティビティ情報を用いる。ここに、ユーザアクティビティはアイテムに対してレビュー投稿や評価付けを行うことを意味する。多くのアイテムレビューサイトでは、有益であったレビューに感謝メッセージを投稿することができる。つまり、ユーザアクティビティ評価情報をサイト内にて観測することができる。信頼リンク予測に対する NMF 法の性能の向上を目指して、我々は信頼リンク情報とアクティビティ情報に加えてアクティビティ評価情報を組み込んだ新たな NMF 法を提案した。そして、アイテムレビューサイト Epinions の実データを用いて提案法の評価を行った。

まず、我々は受け取った感謝メッセージ数と受け取った信頼リンク数の間に相関があることを示し、アクティビティ評価情報を組み込むことが有効である可能性を示した。次に、信頼リンク予測問題において提案法が $hTrust$ やそれと類似した手法 $JLCMF$ および $JLCMF2$ と比べて高性能であることを示した。これらの結果はアクティビティ評価情報を組み込むことの重要性を示しており、2 種類の潜在因子 (すなわち、2 つの潜在空間) である潜在 P 因子および潜在 T 因子を適切に融合した提案法の有効性を実証している。さらに、提案法をアイテムレビューサイトにおけるユーザの行動分析に応用し、信頼リンク生成の観点から Epinions のいくつかの特徴的な性質を明らかにした。

謝辞

本研究は JSPS 科研費 26330352 の助成を受けたものである。

参考文献

- [1] D. Kempe, J. Kleinberg, and E. Tardos, “Maximizing the spread of influence through a social network,” in *ACM SIGKDD*, 2003, pp. 137–146.
- [2] M. Kimura, K. Saito, and H. Motoda, “Blocking links to minimize contamination spread in a social network,” *ACM TKDD*, vol. 3, pp. 9:1–9:23, 2009.
- [3] M. Kimura, K. Saito, R. Nakano, and H. Motoda, “Extracting influential nodes on a social network for information diffusion,” *Data Mining and Knowledge Discovery*, vol. 20, pp. 70–97, 2010.
- [4] K. Saito, M. Kimura, K. Ohara, and H. Motoda, “Selecting information diffusion models over social networks for behavioral analysis,” in *ECML-PKDD*, ser. LNAI, vol. 6323. Springer, 2010, pp. 180–195.
- [5] W. Chen, L. Lakshmanan, and C. Castillo, “Information and influence propagation in social networks,” *Synthesis Lectures on Data Management*, vol. 5, pp. 1–177, 2013.
- [6] M. Kimura, K. Saito, K. Ohara, and H. Motoda, “Learning to predict opinion share and detect anti-majority opinionists in social networks,” *Journal of Intelligent Information Systems*, vol. 41, pp. 5–37, 2013.
- [7] K. Saito, M. Kimura, K. Ohara, and H. Motoda, “Detecting changes in information diffusion patterns over social networks,” *ACM TIST*, vol. 4, pp. 55:1–55:23, 2013.
- [8] R. Guha, R. Kumar, P. Raghavan, and A. Tomkins, “Propagation of trust and distrust,” in *ACM WWW*, 2004, pp. 403–412.
- [9] D. Liben-Nowell and J. Kleinberg, “The link-prediction problem for social networks,” *Journal of the American Society for Information Science and Technology*, vol. 58, pp. 1019–1031, 2007.
- [10] H. Liu, E. Lim, H. Lauw, M. Le, A. Sun, J. Srivastava, and Y. Kim, “Predicting trusts among users of online communities: an opinion case study,” in *ACM EC*, 2008, pp. 310–319.
- [11] H. Mannila and E. Terzi, “Finding links and initiators: a graph-reconstruction problem,” in *SIAM SDM*, 2009, pp. 1207–1217.
- [12] V. Nguyen, E. Lim, J. Jiang, and A. Sun, “To trust or not to trust? predicting online trusts using trust antecedent framework,” in *IEEE ICDM*, 2009, pp. 896–901.
- [13] M. Gomez-Rodriguez, J. Leskovec, and A. Krause, “Inferring networks of diffusion and influence,” in *ACM SIGKDD*, 2010, pp. 1019–1028.
- [14] J. Tang, H. Gao, X. Hu, and H. Liu, “Exploiting homophily effect for trust prediction,” in *ACM WSDM*, 2013, pp. 53–62.
- [15] D. Crandall, D. Cosley, D. Huttenlocher, J. Kleinberg, and S. Suri, “Feedback effects between similarity and social influence in online communities,” in *ACM SIGKDD*, 2008, pp. 160–168.
- [16] D. Lee and H. Seung, “Algorithms for non-negative matrix factorization,” in *NIPS*, 2001, pp. 556–562.
- [17] S. Zhu, K. Yu, Y. Chi, and Y. Gong, “Combining content and link for classification using matrix factorization,” in *ACM SIGIR*, 2007, pp. 487–494.