

社会的紐帯と類似性に基づいた ソーシャルメディアにおけるユーザ行動の分析

津田 洋平^{†1} 熊野 雅仁^{†2}
小野 景子^{†2} 木村 昌弘^{†2}

ソーシャルメディアの実データを用いて、ソーシャルネットワークにおける社会的紐帯と類似性間の相互作用を同定する。さらに、ソーシャルネットワークにおいて、個人が将来行うリンク行動とアクティビティ行動を、社会的影響に関する特徴量と類似性に関する特徴量の両方に基づいて予測する手法を提案する。社会的紐帯と類似性の両方の特徴量を結合した提案法は、社会的影響に関する特徴量のみ従来法や類似性に関する特徴量のみ従来法よりも行動予測において有効であることを、ソーシャルメディアデータを用いた実験で実証する。

Behavioral Analyses of Social Media by Social Ties and Similarity

YOHEI TSUDA,^{†1} MASAHITO KUMANO,^{†2} KEIKO ONO^{†2}
and MASAHIRO KIMURA^{†2}

We identify the interplay between social ties and similarity in a social network using real social media data. Moreover, we propose a method for predicting the future links and activities of an individual in a social network on the basis of both social influence and similarity features. We experimentally demonstrate using the social media data that combining features based on social ties and similarity is more predictive of future behavior than either social influence or similarity features alone.

^{†1} 龍谷大学大学院 理工学研究科 電子情報学専攻
Division of Electronics and Informatics, Ryukoku University

^{†2} 龍谷大学 理工学部 電子情報学科
Department of Electronics and Informatics, Ryukoku University

1. はじめに

ソーシャルメディアにおけるユーザの将来行動を予測する手法は、次世代高度推薦システムのコア技術の一つとして重要と考えられる。ソーシャルメディアにおけるユーザの代表的行動として、お気に入りリンク作成のような社会的紐帯（ソーシャルリンク）の構築や、商品の購買やレビューのようなユーザアクティビティの実行が挙げられるので、ソーシャルメディアデータを対象として、ソーシャルネットワークの成長を予測する手法や、ユーザの将来アクティビティを予測する手法が盛んに研究されている¹⁾⁻⁷⁾。例えば、Liben-Nowell と Kleinberg は、成長ネットワークにおいてノードの“proximity”に基づいたリンク予測法を提案しており¹⁾、また、Resnick らは協調フィルタリングにおける最近傍アプローチに基づいたユーザの将来アクティビティ予測法を提案している²⁾。

従来、ソーシャルネットワークの成長予測問題とユーザの将来アクティビティの予測問題は、別々に研究されることが多かった。しかし最近 Crandall らは、Wikipedia 編集コミュニティに所属するユーザらの Wikipedia ページ編集アクティビティと、LiveJournal においてフレンドリストで結ばれたユーザらのコミュニティ参加アクティビティを調べ、ソーシャルメディアにおける、ソーシャルネットワークの成長とユーザアクティビティの類似性変化の間に、相互作用が存在することを明らかにした⁸⁾。本研究では、Crandall らの研究に基づき、ソーシャルネットワークの成長とユーザアクティビティの将来を、ソーシャルネットワークとユーザアクティビティの類似性を用いて予測するというアプローチを提案する。そして、実ソーシャルメディアデータとして化粧品クチコミサイトの@cosme^{*1} のデータを用いた実験により、提案アプローチの有効性を示す。

本報告の構成は以下のとおりである。2 節では、@cosme データにおけるソーシャルネットワークおよびユーザアクティビティを定義し、それらを統計的に分析する。そして、ソーシャルネットワークの成長とユーザアクティビティの類似性変化の間に、たしかに相互作用が存在することを示す。3 節では提案する予測アプローチを述べ、4 節では@cosme データを用いた評価実験の結果を述べる。5 節はまとめである。

2. 分析データ

実ソーシャルメディアデータとして、@cosme データを用いる。

*1 みんなのクチコミサイト@cosme <http://www.cosme.net/>

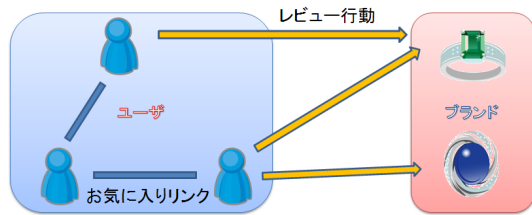


図 1 @cosme データにおける社会的紐帯とユーザアクティビティの定義

Fig. 1 The definition of the social tie and the user activity in the @cosme data.

2.1 @cosme データ

@cosme サービスでは、「ユーザが他のユーザへのお気に入りリンクを作成する行動」や、「ユーザがあるブランドをレビューする行動」が存在する(図 1 参照)。本研究では、前者をユーザ間の社会的紐帯と定義し、後者をユーザアクティビティと定義する。すなわち、@cosme データでは、ユーザをノードとし、お気に入りリンクを無向リンクと考えると、ユーザ間のソーシャルネットワークを定義する。また、ユーザ i がブランド m をレビューしたという行動を、ユーザ i がアクティビティ m を実行したと定義する。我々は、ユーザアクティビティ全体の集合(注: ブランド全体の集合と同一視される)を \mathcal{M} と記述する。

@cosme データを収集する方法として、1 個のノードを起点としたネットワークを幅優先探索で辿り、分析データを自動構築した。ただし、登録ノードからランダムに選んだ 3 個のノードを起点として、3 つのネットワークを構築し、相互のノードの一致度を検証した結果、99.5% のノードが一致した。このことから、この手法により構築されたネットワークのノード群はリンクで結ばれた一般的なノード群を反映していると思われる。そこでこの手法により選択した 1 個のノードに基づくネットワークを対象として分析を行った。1999 年 9 月～2010 年 6 月までの期間を取得し、ブランド数 9113 種類、ノード数 46256 人となった。

2.2 統計分析

まず、収集した @cosme ソーシャルネットワークの特性を、複雑ネットワーク科学分野で用いられているネットワーク統計量に基づいて調べ、それは一般的なソーシャルネットワークがもつ統計的性質を満足していることを確認する。また、ソーシャルネットワークの成長とユーザアクティビティの類似性変化の関係を分析する。

2.2.1 ソーシャルネットワーク成長分析

まず、本研究では、2010 年 6 月時点のソーシャルネットワークの統計量分析を行った。図 2 は 2010 年 6 月時点のソーシャルネットワークに関する度数分布であり、スケールフリー

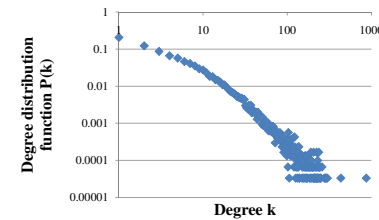


図 2 @cosme ソーシャルネットワークの度数分布

Fig. 2 The degree distribution of the @cosme social network.

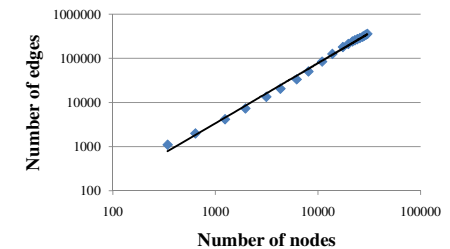


図 3 ノード数 $n(t)$ に対するリンク数 $e(t)$ (log-log scales)

Fig. 3 Number of edges $e(t)$ versus number of nodes $n(t)$ in log-log scales.

性が認められる。2010 年 6 月時点のユーザ数 46256 人のソーシャルネットワークにおいて、平均パス長は 4.18 であり、クラスター係数は 0.022 であった。松尾らによれば、mixi で行ったネットワーク統計量のうち、50000 人規模のネットワークにおいて平均パス長は 4.63 であり、クラスター係数は 0.35 であった⁹⁾。これにより、mixi に比べ、平均パス長は大きな差がないものの、クラスター係数が 1 桁小さく、クラスター性が低いネットワークであると考えられる。

次に、1999 年 9 月～2010 年 6 月までのソーシャルネットワークの成長分析を行った。図 3 に、ノード数 $n(t)$ に対するリンク数 $e(t)$ の増加率の変遷を示す。Leskovec らは¹⁰⁾、時刻 t でのノード数を $n(t)$ 、リンク数を $e(t)$ とすると、

$$n(t) = b \times e(t)^\beta \quad (1)$$

の関係が成り立つと指摘している。これは、平均リンク数が時間と共に増大し、ネットワークが稠密 (dense) になることを意味しており、densification power law と呼ばれる。図 3 から、ネットワーク成長過程のノード数 $n(t)$ とリンク数 $e(t)$ の関係において、グラフの傾きが直線的になり、べき乗則がほぼ成り立っていることがわかる。なお、図中の直線は最小二乗法で求めた近似直線である。ソーシャルネットワークでは、時刻 t でのリンク数 $e(t)$ とノード数 $n(t)$ との間に式 (1) が成り立つと言える。べき指数と係数を調べた結果、 $\beta = 0.84, b = 0.59$ となった。べき指数が 1 より小さいことから、時間経過と共に、ネットワーク全体の平均リンク数が増加することがわかる。

図 4 で、ネットワーク成長における有効直径について述べる。有効直径とはネットワーク

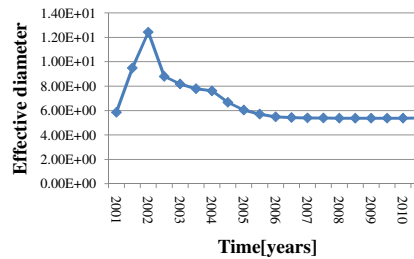


図 4 有効直径の変動

Fig. 4 Fluctuation in effective diameter.

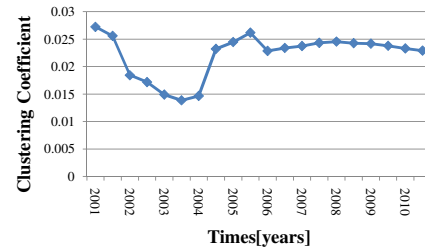


図 5 クラスタ係数の変動

Fig. 5 Fluctuation in clustering coefficient.

の経路長のことである．今，任意のノード i, j に対して， $d_{i,j}$ をそのグラフのノード間距離とする．なおノード i と j が非連結の場合は便宜上 $d_{i,j}(t) = \infty$ と定める．このときの有効直径 $ED(t)$ は，Leskvec らの研究¹⁰⁾ にしたがって，

$$ED(t) = \min \left\{ d; \frac{|\{(i, j); i < j, d_{i,j}(t) \leq d\}|}{|\{(i, j); i < j, d_{i,j}(t) < \infty\}|} \geq \beta \right\} \quad (2)$$

によって定義できる．ここで， $|\{\cdot\}|$ は集合の要素数を表す．また， $\beta = 1.0$ とすると，通常のネットワーク直径 (graph diameter) となる．本研究では，Leskvec らの論文に従い $\beta = 0.9$ を採用した．このように有効直径を採用する理由は，ネットワークの成長過程において，外れ値と見なせるほど距離の大きい少数のノードペア群の影響を除き，安定して評価値を求めるためである．図 4 より，2002 年をピークとして有効直径が 6 程度から 12 程度へと急激に増加するが 2006 年頃から 6 程度に収束する様子が伺える．

図 5 で，ネットワーク成長におけるクラスタ係数について述べる．クラスタ係数とは，ノード同士がどの程度密集しているのかを調べるためのものである．クラスタ係数は各ノード毎にクラスタ係数を求めて平均する Watts & Strogatz の定義¹¹⁾ がよく知られている．しかし本研究では，ネットワーク全体に対するクラスタとして三角形 $(a_{i,j}(t)a_{i,k}(t)a_{j,k}(t))$ が時刻 t までに増加したときの個数に着目した定義¹²⁾ を採用する．今，時刻 t におけるノード i と j の隣接関係を $a_{i,j}(t) \in \{0, 1\}$ とし，ノード i の次数 (リンク数) を $d_i(t)$ としたとき，クラスタ係数 $CC(t)$ を，Newman らの研究にしたがって，

$$CC(t) = \frac{6 \sum_{i < j < k} a_{i,j}(t)a_{i,k}(t)a_{j,k}(t)}{\sum_i d_i(t)(d_i(t) - 1)} \quad (3)$$

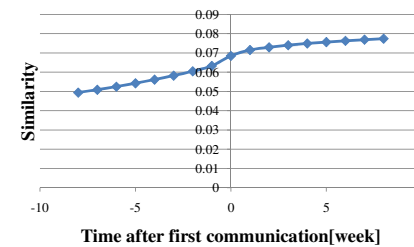


図 6 ユーザペアが初めて結ばれてから，時間が経つにつれての類似性変化を示した結果

Fig. 6 Results for the average similarity of user pairs as a function of the time after first communication.

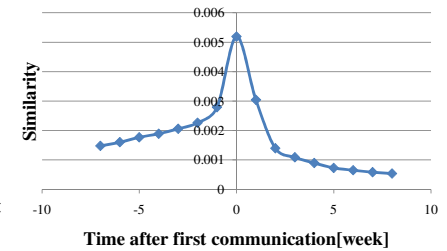


図 7 ユーザペアが初めて結ばれてから，時間が経つにつれての類似性変化量の結果

Fig. 7 Results for the difference of average similarity of user pairs as a function of the time after first communication.

によって定義する．

図 5 より，図 4 の有効直径変動期と思われる 2001 ~ 2006 年頃の期間に連動するが如く，クラスタ係数も 2001 ~ 2006 年頃まで大きな変動が見られる．しかし，2006 年以降比較的安定する傾向があることがわかる．図 3 から，ノード数とリンク数はこの 10 年間で常に増加傾向を示しているにも関わらず，2006 年以降の有効直径とクラスタ係数は安定する傾向があることになる．

2.2.2 ソーシャルネットワークの成長とユーザアクティビティの類似性変化の関係

ここでは，ソーシャルネットワーク成長とユーザアクティビティの類似度変化の間に，相互作用が存在するかを検証する．ノード i とノード j のユーザアクティビティの類似度 $B_{i,j}$ を，Crandall らの研究⁸⁾ にしたがって，

$$B_{i,j} = \frac{|\Gamma(i) \cap \Gamma(j)|}{|\Gamma(i) \cup \Gamma(j)|} \quad (4)$$

によって定義する．ここに， $\Gamma(i)$ はノード i が実行したアクティビティ全体の集合， $\Gamma(i) = \{m \in \mathcal{M}; u_{i,m} > 0\}$ である．ただし， $u_{i,m}$ はノード i がアクティビティ $m \in \mathcal{M}$ を実行した数である．

図 6 は，あるノードペアが初めてリンクを結んだから，1 週間毎にそのノードペアのアクティビティの類似性変化を示す．また，図 7 で，図 6 での変化量を示す．ここで，図 6，

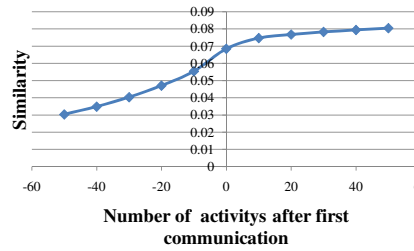


図 8 ユーザペアが初めて結ばれてから、アクティビティを実行した回数毎での類似性変化を示した結果
Fig.8 Results for the average similarity of user pairs as a function of the number of activities after first communication.

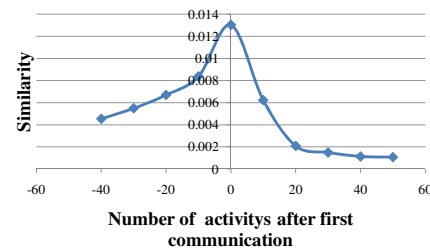


図 9 ユーザペアが初めて結ばれてから、アクティビティを実行した回数毎での類似性変化量の結果
Fig.9 Results for the difference of the average similarity of user pairs as a function of the number of activities after first communication.

図 7 は、ノードペア 359,817 組の平均である。図 7 より、この横軸 0 の時点をピークとしてその近傍で類似度が高まっていることがわかり、図 6 の横軸 0 以降も上昇傾向を示すことから、リンクを結んだ以後も類似度が増加していくことがわかる。

次に、図 8 より、あるノードペアが初めてリンクを結んでから、アクティビティを実行した回数毎に、そのノードペアのアクティビティの類似性変化を示す。また、図 9 で、図 8 での変化量を示す。その結果、図 8 と図 9 は、図 6 と図 7 と同様の傾向を示した。このことから、リンクが結ばれる前後で類似度が急激に変化するため、ソーシャルネットワークの成長とユーザアクティビティの類似性変化の間には、たしかに相互作用が存在した。

3. 提案法

本節では、ソーシャルネットワークの成長とユーザアクティビティの将来を、ソーシャルネットワークとユーザアクティビティの類似性を用いて予測する手法を提案する。

観測されたソーシャルネットワークデータ $G = (V, E)$ に対して、任意のノード i の近傍ノード全体の集合を $F(i)$ とする。ここに、 V はノード全体の集合、 E はリンク全体の集合を表す。 \mathcal{M} をユーザアクティビティ全体の集合とする。観測期間においてノード $i \in V$ がアクティビティ $m \in \mathcal{M}$ を実行した数を $u_{i,m}$ とし、 $\mathcal{U} = \{u_{i,m}; i \in V, m \in \mathcal{M}\}$ を観測期間におけるユーザアクティビティデータと呼ぶ。我々は、観測されたソーシャルネットワーク

におけるノード i とノード j の proximity $A_{i,j}$ を、Adamic と Adar の研究³⁾ にしたがって、

$$A_{i,j} = \sum_{k \in F(i) \cap F(j)} \frac{1}{\log |F(k)|} \quad (5)$$

によって定義する。また、観測期間におけるノード i とノード j のユーザアクティビティの類似度 $B_{i,j}$ を、Crandall らの研究⁸⁾ にしたがって、

$$B_{i,j} = \frac{|\Gamma(i) \cap \Gamma(j)|}{|\Gamma(i) \cup \Gamma(j)|} \quad (6)$$

によって定義する。ここに、 $\Gamma(i)$ はノード i が観測期間に実行したアクティビティ全体の集合、

$$\Gamma(i) = |\{m \in \mathcal{M}; u_{i,m} > 0\}| \quad (7)$$

である。

3.1 ソーシャルネットワークの成長予測法

まず、ソーシャルネットワークの成長を、ソーシャルネットワークとユーザアクティビティの類似性を用いて予測する手法を提案する。我々は、観測されたソーシャルネットワークデータ $G = (V, E)$ と、観測期間におけるユーザアクティビティデータ \mathcal{U} が与えられたとき、任意のノード $i \in V$ が将来リンクするノード $j \in V$ を、スコア $Q_{i,j}$ の値に基づいてランキングすることにより予測する。ここに、スコア $Q_{i,j}$ は、ソーシャルネットワークとユーザアクティビティの類似性を用いて、

$$Q_{i,j} = A_{i,j} B_{i,j} \quad (8)$$

と定義する。

3.2 ユーザアクティビティの予測法

次に、ユーザアクティビティの将来を、ソーシャルネットワークとユーザアクティビティの類似性を用いて予測する手法を提案する。我々は、観測されたソーシャルネットワークデータ $G = (V, E)$ と、観測期間におけるユーザアクティビティデータ \mathcal{U} が与えられたとき、任意のノード $i \in V$ が将来実行するアクティビティ $m \in \mathcal{M}$ を、スコア $R_{i,m}$ の値に基づいてランキングすることにより予測する。ここに、スコア $R_{i,m}$ は、ソーシャルネットワークとユーザアクティビティの類似性を用い、協調フィルタリング法に基づいて、

$$R_{i,m} = \sum_{j \in V} A_{i,j} B_{i,j} \frac{u_{j,m}}{\sum_{\ell \in \mathcal{M}} u_{j,\ell}} \quad (9)$$

と定義する。

4. 実験

本節では、ソーシャルネットワークの成長予測とユーザアクティビティの予測における提案法と比較法の性能を比較する。

4.1 実験データ

本研究では、観測期間 2009 年 12 月～2010 年 5 月までの 6ヶ月間と評価期間 2010 年 6 月の 1ヶ月間でのソーシャルネットワークデータとユーザアクティビティデータのデータセットを D_1 とし、観測期間 2000 年 5 月～2010 年 5 月までの 10 年間と評価期間 2010 年 6 月の 1ヶ月間でのデータセットを D_2 とする。また、観測期間 2000 年 5 月～2010 年 5 月までの 10 年間のソーシャルネットワークデータ、2009 年 12 月～2010 年 5 月までの 6ヶ月間でのユーザアクティビティデータと評価期間 2010 年 6 月の 1ヶ月間でのソーシャルネットワークデータとユーザアクティビティデータのデータセットを D_3 とする。

4.2 ソーシャルネットワークの成長予測

ソーシャルネットワークの成長予測の比較検証を行う。

4.2.1 比較法

提案法と以下の手法を比較した。比較法は、観測されたソーシャルネットワークデータ $G = (V, E)$ と、観測期間におけるユーザアクティビティデータ U が与えられたとき、任意のノード $i \in V$ が将来リンクするノード $j \in V$ を、スコア $Q_{i,j}$ の値に基づいてランキングすることにより予測する。

- A_J 法

スコア $Q_{i,j}$ を、観測期間におけるノード i とノード j ユーザアクティビティの類似度を用いて

$$Q_{i,j} = \frac{|\Gamma(i) \cap \Gamma(j)|}{|\Gamma(i) \cup \Gamma(j)|} \quad (10)$$

によって定義する。

- S_CN 法

スコア $Q_{i,j}$ を、観測されたソーシャルネットワークにおけるノード i とノード j の proximity を用い、Newman の研究⁴⁾ にしたがって、

$$Q_{i,j} = |F(i)| \cap |F(j)| \quad (11)$$

によって定義する。

- S_J 法

スコア $Q_{i,j}$ を、観測されたソーシャルネットワークにおけるノード i とノード j の proximity を用い、Liben-Nowell と Kleinberg の研究¹⁾ にしたがって、

$$Q_{i,j} = \frac{|F(i) \cap F(j)|}{|F(i) \cup F(j)|} \quad (12)$$

によって定義する。

- S_A/A 法

スコア $Q_{i,j}$ を、観測されたソーシャルネットワークにおけるノード i とノード j の proximity を用い、Adamic と Adar の研究³⁾ にしたがって、

$$Q_{i,j} = \sum_{k \in F(i) \cap F(j)} \frac{1}{\log |F(k)|} \quad (13)$$

によって定義する。

- S_K 法

スコア $Q_{i,j}$ を、観測されたソーシャルネットワークにおけるノード i とノード j の proximity を用い、Katz の研究⁶⁾ にしたがって、

$$Q_{i,j} = \sum_{l=1}^{\infty} \beta^l |\text{paths}_{i,j}^{(l)}| \quad (14)$$

によって定義する。paths_{i,j}^(l) は、ノード i からノード j への長さ l のパスの集合である。本実験では $\beta = 0.05$ とした。

- S_PA 法

スコア $Q_{i,j}$ を、観測されたソーシャルネットワークにおけるノード i とノード j の proximity を用い、Barabasi の研究⁷⁾ にしたがって、

$$Q_{i,j} = |F(i)| \cdot |F(j)| \quad (15)$$

によって定義する。

4.2.2 結果

各予測法の性能をそのランキング法が抽出する上位 r 個のノードの精度によって評価する。図 10, 図 11, 図 12 は、各手法の精度 a をランク r の関数としてプロットしている。

図 10 は、観測期間 2009 年 12 月～2010 年 5 月までの 6ヶ月間、評価期間 2010 年 6 月の 1ヶ月間としたデータセット D_1 を用いて、ソーシャルネットワークの成長予測を行った結果であり、図 11 は、観測期間 2000 年 5 月～2010 年 5 月までの 10 年間、評価期間 2010

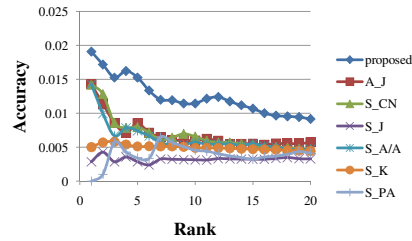


図 10 データセット D_1 での結果
Fig. 10 Results for dataset D_1 .

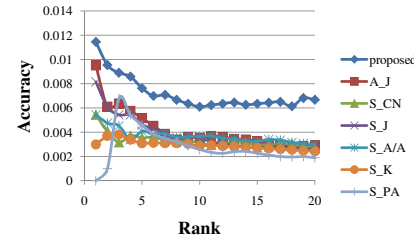


図 11 データセット D_2 での結果
Fig. 11 Results for dataset D_2 .

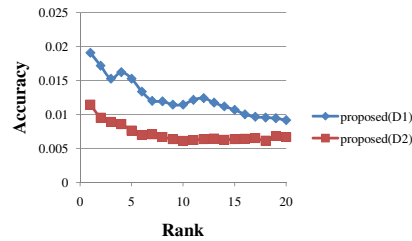


図 12 観測期間の長さに関する性能比較
Fig. 12 Performance comparison for the length of an observation period.

年 6 月の 1ヶ月間としたデータセット D_2 での結果である。どちらの場合においても、提案法が比較法よりも高い精度を示した。

図 12 は、データセット D_1 の提案法の精度と、データセット D_2 の提案法の精度を比較している。 D_2 よりも D_1 での提案法が高い精度を示した。また、提案法には、6ヶ月間のデータを使用した方が精度が高くなることが示唆された。

4.3 ユーザアクティビティの予測

ユーザアクティビティの予測について比較検証を行う。

4.3.1 比較法

提案法を以下の手法と比較した。比較法は、観測されたソーシャルネットワークデータ $G = (V, E)$ と、観測期間におけるユーザアクティビティデータ U が与えられたとき、任意

のノード $i \in V$ が将来実行するアクティビティ $m \in M$ を、スコア $R_{i,m}$ の値に基づいてランキングすることにより予測する。

• N1 法

スコア $R_{i,m}$ を、ノード i のソーシャルネットワーク G の第 1 近傍までのノードが、アクティビティ m を実行した数と定義する。

• N2 法

スコア $R_{i,m}$ を、ノード i のソーシャルネットワーク G の第 2 近傍までのノードが、アクティビティ m を実行した数と定義する。

• N3 法

スコア $R_{i,m}$ を、ノード i のソーシャルネットワーク G の第 3 近傍までのノードが、アクティビティ m を実行した数と定義する。

• N4 法

スコア $R_{i,m}$ を、ノード i のソーシャルネットワーク G の第 4 近傍までのノードが、アクティビティ m を実行した数と定義する。

• All 法

スコア $R_{i,m}$ を、ノード i のソーシャルネットワーク G の全てのノードが、アクティビティ m を実行した数と定義する。

• CF 法

スコア $R_{i,m}$ は、ユーザアクティビティの類似性を用い、協調フィルタリング法に基づいて、

$$R_{i,m} = \sum_{j \in V} B_{i,j} \frac{u_{j,m}}{\sum_{\ell \in M} u_{j,\ell}} \quad (16)$$

と定義する。

4.3.2 結果

各予測法の性能をそのランキング法が抽出する上位 r 個のブランドの精度によって評価する。図 13、図 14、図 15、図 16 は、各手法の精度 a をランク r の関数としてプロットしている。

図 13 は、観測期間 2009 年 12 月～2010 年 5 月までの 6ヶ月間、評価期間 2010 年 6 月の 1ヶ月間としたデータセット D_1 を用いて、ユーザアクティビティの予測を行った結果であり、CF 法が提案法より高い精度を示した。図 14 は、観測期間 2000 年 5 月～2010 年 5 月までの 10 年間、評価期間 2010 年 6 月の 1ヶ月間としたデータセット D_2 での結果であ

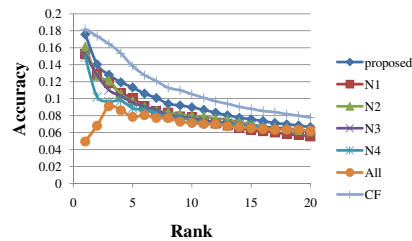


図 13 データセット D_1 での結果
Fig. 13 Results for dataset D_1 .

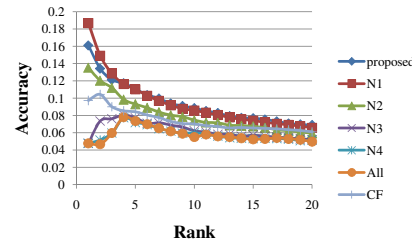


図 14 データセット D_2 での結果
Fig. 14 Results for dataset D_2 .

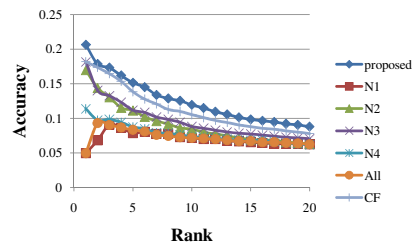


図 15 データセット D_3 での結果
Fig. 15 Results for dataset D_3 .

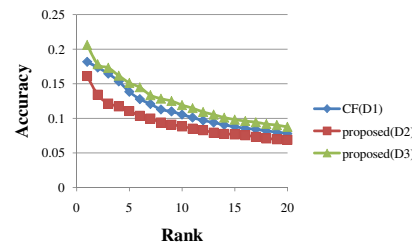


図 16 観測期間の長さに関する性能比較
Fig. 16 Performance comparison for the length of an observation period.

り、提案法と N1 法が他の比較法よりも高い精度を示した。図 15 は、観測期間 2000 年 5 月～2010 年 5 月までの 10 年間のソーシャルネットワークデータ、2009 年 12 月～2010 年 5 月までの 6 ヶ月間でのユーザアクティビティデータと評価期間 2010 年 6 月の 1 ヶ月間でのソーシャルネットワークデータとユーザアクティビティデータのデータセット D_3 での結果であり、提案法が比較法より高い精度を示した。

図 16 は、データセット D_1 の CF 法の精度、データセット D_2 の提案法の精度、データセット D_3 の提案法の精度を比較している。 D_3 での提案法が、 D_1 の CF 法、 D_3 の提案法より高い精度を示した。また、提案法には、ソーシャルネットワークデータには 10 年間のデータ、ユーザアクティビティデータには、6 ヶ月間のデータを使用すると精度が高くなることが示唆された。

5. ま と め

本研究では、実ソーシャルメディアデータとして化粧品クチコミサイト@cosme のデータを用いて、ソーシャルネットワークにおける社会的紐帯と類似性との相互作用を同定した。そしてその結果に基づき、ソーシャルネットワークにおいて個人が将来行うリンク行動とアクティビティ行動を、社会的紐帯に関する特徴量と類似性に関する特徴量の両方を用いて予測する手法を提案した。@cosme データを用いた実験により、社会的紐帯と類似性の両方の特徴量を結合した提案法は、社会的影響に関する特徴量のみからの従来法や類似性に関する特徴量のみからの従来法よりも、リンク行動予測およびアクティビティ行動予測において精度が向上することを確認した。

参 考 文 献

- 1) Liben-Nowell, D. and Kleinberg, J.: The Link Prediction Problem for Social Networks, in Proceedings of the Twelfth International Conference on Information and Knowledge Management (CIKM), pp.556–559 (2004).
- 2) Resick, P., Iacovou, N., Suchak, M., Bergstrom, P. and Riedl, J.: GroupLens; an open architecture for collaborative filtering of networks, Proc. 9th ACM Conf. on Computer-Supported Cooperative Work, pp.175–186 (1994).
- 3) Adamic, L. A. and Adar, E.: Friends and neighbors on the Web, Social Networks, Vol.25, No.2, pp.211–230 (2003).
- 4) Newman, M. E. J.: Clustering and preferential attachment in growing networks, Physical Review Letters E, Vol.64(025102) (2001).
- 5) Baeza-Yates, R. A. and Ribeiro-Neto, B, A.: Modern Information Retrieval, ACM Press / Addison-Wesley (1999).
- 6) Katz, L.: A new status index derived from sociometric analysis, Psychometrika, Vol.18, No.1, pp.39–43 (1953).
- 7) Barabasi, A. L., Jeong, J., Neda., Z., Ravasz, E., Shubert, A., and Vicsek, T.: Evolution of the social network of scientific collaborations, Physica A, Vol.311, No.3-4, pp.590–614 (2002).
- 8) Crandall, D., Cosley, D., Huttenlocher, D., Kleinberg, J. and Suri, S.: Feedback Effects between Similarity and Social Influence, Proc. KDD-2008, pp.160–168 (2008).
- 9) 松尾 豊, 安田 雪: SNS における関係形成原理 – mixi のデータ分析 –, 人工知能学会論文誌, Vol.22, No.5, pp.531–541 (2007) .
- 10) Leskovec, J. Kleinberg, J. and Faloutsos, C.: Graphs over time: densification laws, shrinking diameters and possible explanations, Proc. KDD-2005, pp.177–187

(2005).

- 11) Watts, D. J. and Strogatz, S. H.: Collective dynamics of 'small-world' networks, Nature, Vol.393, pp.440-442 (1998).
- 12) Newman, M. E. J.: The structure and function of complex networks, SIAM Review, Vol.45, No.2, pp.167-256 (2003).