

SNS におけるユーザ推薦のための潜在的なフォロー関係予測手法

屈 剣鋒[†] 牛尼 剛聡[‡]

[†]九州大学芸術工学府 〒815-8540 福岡県福岡市南区塩原 4-9-1

[‡]九州大学芸術工学研究院 〒815-8540 福岡県福岡市南区塩原 4-9-1

E-mail: [†]2ds20009n@s.kyushu-u.ac.jp, [‡]ushiana@design.kyushu-u.ac.jp

あらまし 近年、インターネットの情報が増大し、ユーザにとって価値のある情報を取得するためには、SNS が重要な役割を果たしている。SNS では、興味のある友人やアイテムをユーザに推薦する機構が重要である。しかし、ユーザ数とアイテム数が増加するにつれて、データの疎性問題はますます深刻になり、推薦システムの精度も低下する。この問題を解決するために、本研究では、信頼ネットワークに基づくユーザ潜在的なフォロー関係予測手法を提案する。本手法では、フォロー関係行列を利用してユーザ間の信頼度と類似度を計算し、潜在的なフォロー関係を予測し、推薦システムに応用する。

キーワード SNS、推薦システム、データスパースネス、信頼ネットワーク

1. はじめに

SNS (Social Networking Service) の発展により、人々のコミュニケーション方法や除法の取得方法が変化した。インターネットを利用するユーザは、主に SNS 上で興味のある人や事柄をフォローすることによって、興味がある最新のニュースを取得する[1]。ユーザにとって、興味のある情報を提供してくれる可能性が高いユーザを推薦できれば、SNS に対する満足度を高めることができる。したがって、SNS でフォローするユーザを効果的に推薦するための推薦アルゴリズムが重要な研究テーマとなっている。

従来の推薦システムでは、ユーザ-アイテム評価行列データセットが多く使われているが[2]、電子商取引が多くなるにつれて、データセットのユーザとアイテムの数が増え、ユーザ-アイテム評価行列のスパース性が高くなる。スパース性が高いユーザ-アイテム評価行列を使用する場合、推薦の基準となるユーザ分野類似度やアイテム間の類似度がますます不正確になり、推薦精度が低くなるという問題が発生する。

社会的な交流関係は人々の実際の生活において不可欠な貴重な資源であり、SNS においても非常に重要であり、データが少ない推薦システムで有効に活用されることが期待される。単純な SNS に基づく友人推薦システムは、主にユーザの明示的な個人情報と明示的な社交関係[3, 4]を利用している。さらに、ハイブリッド協調フィルタリングで明示的な個人情報と明示的な社交関係を使用すると、推薦の精度も向上することが報告されている[5]。ここで利用される明示的な個人情報とは、ユーザの地理的位置、タグ情報、テキスト情報な

どである。一方明示的な社交関係とは、一般的に SNS ユーザ間の強い関係を指し、直接フォローされた関係である。本研究ではフォロー関係のネットワーク（信頼ネットワーク）に注目する。

信頼ネットワークにはスパース性の問題があることが知られている。例えば、表 1 の通り、研究で使った Epinions データセットの疎さは 99.98%である[6]。

表 1. フォロー関係行列

	user ₁	user ₂	user ₃	...	user _n
user ₁	-	1	?		?
user ₂	?	-	1		?
user ₃	1	?	-		?
...				-	
user _n	?	?	?		-

この問題を解決するために、本論文では、信頼ネットワークの潜在的なフォロー関係を予測することによって、スパース性が高い行列においても、より高精度に推薦を実現可能とするための手法を提案する。具体的には、ソーシャルネットワークにおける信頼は伝播する性質が存在すると考えられるので[7]、その特性を利用して、伝播による潜在的なフォロー関係を予測する。予測された潜在的なフォロー関係を利用して、信頼ネットワークを拡大することで、推薦精度を向上させることを本研究の目的とする。

2. 関連研究

フォロー関係に基づく代表的なアルゴリズムとしては、ソーシャルフィルタリングと PageRank がある[12]。ソーシャルフィルタリングアルゴリズムは主にターゲ

ットユーザとの共通フォロワーが最も多いユーザを推薦する[8]。Weng の手法では、PageRank アルゴリズムを SNS に適用して、ユーザの影響力を計算し、ターゲットユーザに影響力が最も大きいユーザを推薦する[9]。Wang の手法はユーザの社交関係に基づいて、友人の分類と関係の重みの分類を使用して、友人が多すぎて情報が急増する問題を解決した[10]。Chen は、ユーザの影響力がユーザ間のリンク構造に関連していると考え、ユーザの社交関係を使用してユーザ間の関係強度マトリックスを計算し、PageRank アルゴリズムを組み合わせてユーザの影響力を計算し、PeopleRank と呼ばれる手法で友人推薦を行う[11]。PeopleRank のアルゴリズムは以下の通りである

- (1) ターゲットユーザが一人のユーザをフォローし、現在のユーザに関連するユーザを特定の確率 d でフォローし、他のユーザを $(1-d)$ の確率でランダムにフォローすることを前提とする。
- (2) ユーザ間の有向グラフを作成する。ユーザ u_i とユーザ u_j を同時にフォローしているユーザがいる場合、 u_i から u_j までの有向エッジがある。
- (3) 行列 $M = (m_{ij})_{n \times n}$ が $n \times n$ 行列、 n がユーザ数、 m_{ij} が有向エッジ $u_i \rightarrow u_j$ の重みであり、 u_i をフォローしたユーザが u_j をフォローする確率を表す。式(1)に示す。 u_i のサポート度 $Support(u_i)$ として、 u_i 、 u_j を除く残りの $n-2$ ユーザの中で u_i をフォローしたユーザ数を定義する。 $Support(u_i, u_j)$ は u_i, u_j を同時にフォローしているユーザ数である。

$$m_{ij} = \begin{cases} \frac{Support(u_i, u_j)}{Support(u_i)}, & Support(u_i) > 0, \\ 0, & else. \end{cases} \quad (1)$$

- (4) PageRank 法を用いてユーザの PeopleRank 値を計算する。式(2)に示す。 $PR_u(u_i)$ はユーザ u に推薦されるユーザ u_i の PeopleRank 値を表し、 d は減衰係数であり、 U は全部のユーザである。

$$PR_u(u_i) = d \times \sum_{u_j \in U} PR_u(u_j) \times m_{ji} + \frac{1-d}{n} \quad (2)$$

- (5) ターゲットユーザ u に対して、フォローしているユーザを first-user としてマークし、first-user がフォローしているユーザを second-user としてマークする。second-user ユーザセットの中の PeopleRank 値が高いユーザ topN をユーザ u に推薦する。

本研究では、ユーザ間の類似度と影響力も考慮して推薦を行うことを考える。

3. 潜在的なフォロー関係の予測

本論文で提案する手法は図 1 に示すような 3 つの部分から構成される。第 1 の要素はフォロー関係からユーザ間類似度を計算し、類似のユーザ行列を構築する部分である。第 2 の要素はユーザのフォロー関係を正規化し、信頼度に変換する部分である。第 3 の要素は類似度と信頼度による潜在的なフォロー関係のスコアを予測する部分である。最終的に予想されるスコアが高いほど、ユーザがフォローしたい可能性が高いと考える。

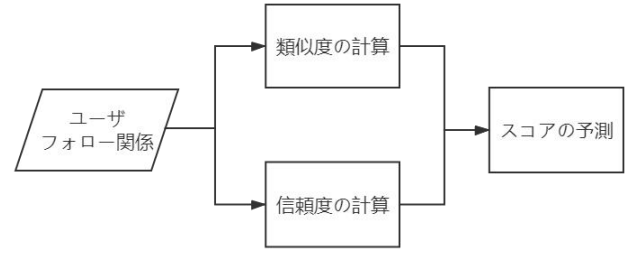


図 1: 潜在的なフォロー関係の予測手法

以下に、潜在的なフォロー関係のスコアを推定するために利用する類似度と信頼度の計算手法について詳しく説明する。

3.1. 類似度の計算

類似ユーザとは、2 つのユーザの間に信頼があり、ある程度の類似性があるユーザを指す。そして類似性は類似度によって測定される必要がある。ここでの類似度はユーザ-アイテム評価行列に基づいて計算された類似度は、異なる関心ターゲットに基づいており、計算方法も異なる。本研究では、2 ユーザがフォローするユーザ集合において、共通のユーザ数が多いほど、2 つのユーザの類似度が高いと考える。この原理に従って、ユーザ u とユーザ v の類似性の計算を式(3)として定義する。

$$S(u, v) = \frac{|S_u \cap S_v|}{|S_u|} \quad (3)$$

ここで、 S_u と S_v はユーザ u と v がフォローしているユーザ集合を表している。計算されたユーザ類似度は $[0, 1]$ の範囲にあり、ユーザーアイテム評価行列に基づいて計算されたユーザ類似度と同じ範囲にあるため、偏差は大きくない。

3.2. 信頼度の計算

一般的に、もし多くのユーザが特定のユーザを信頼しているなら、そのユーザはより多くのユーザから信頼されやすいと考えられる。しかし、2 値の信頼ネッ

トワークでは、これらの状況を区別しておらず、信頼 (1) と不信頼 (0) の 2 つの値しかない。

図 2 に示すように、ユーザ u_a が u_b および u_c とフォロー関係にある場合を考える。フォロー関係行列では、それらの信頼度は 1 である。しかし、 u_b をフォローしているユーザの数は u_c の数より多いため、 u_b の方がより信頼できると考えられ、 u_a が u_b の提案を採用する可能性は u_c よりも高いと考えられる。

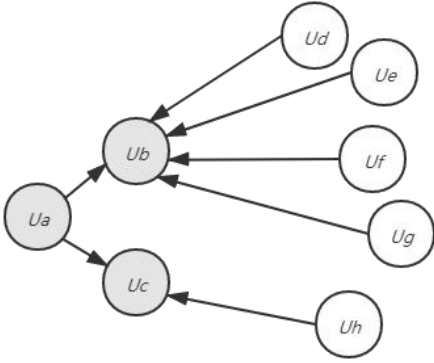


図 2：信頼ネットワーク

以上の考えに基づいて、ユーザ u は v をフォローしているとき、 $In(v)$ を使ってユーザ v の In-degree、すなわち v をフォローしているユーザの数を表すとする。 T_u はユーザ u が直接フォローするユーザのセット

である。 $\sum_{k \in T_u} In(k)$ は、ユーザ u を直接フォローしている

すべてのユーザの In-degree です。本研究における信頼度の定義を式 (4) に示す。

$$T(u, v) = \frac{In(v)}{\sum_{k \in T_u} In(k)} \quad (4)$$

計算された信頼度は $[0, 1]$ の範囲にあり、信頼の非対称性も満足する [13]。

3.3. 潜在的なフォロー関係スコアの予測

フォローネットワークから直接得られるフォロー関係行列は 0 と 1 の値しかなく、スパース性が高い。信頼の伝播性を考えることにより、信頼度を計算してから類似度によって信頼を伝播させ、潜在的なフォロー関係を予測する。ユーザ u との類似度が最も高いユーザ $topN$ のセットを S_u で表す。 $S(u, v)$ は u と v のフォロー関係に基づく類似度であり、 $T(k, v)$ は k と v の変換した信頼度である。 u と v の潜在的なフォロー関係スコア予測を式 (5) に示す。潜在的なフォロー関係スコアの計算を表す例を図 3 に表す。

$$P(u, v) = \frac{\sum_{k \in S(u)} S(u, k) \times T(k, v)}{\sum_{k \in S(u)} S(u, k)} \quad (5)$$

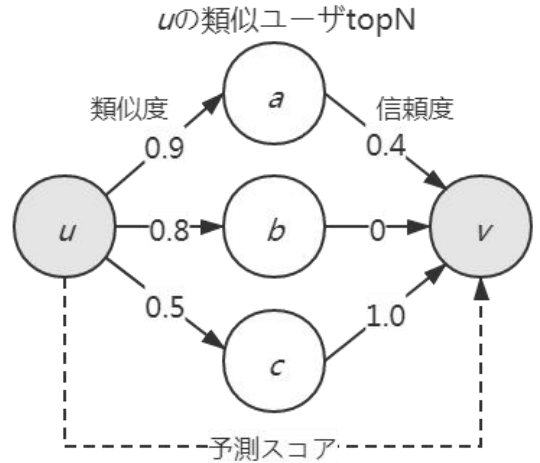


図 3. スコアの予測手法

潜在的なフォロー関係スコアを予測した後、一定のしきい値以上のスコアがあるユーザ間にフォロー関係が存在すると考えることにより、ユーザ間行列の組成が解消され、効果的な推薦が可能となると期待できる。

4. 評価実験

提案手法の有効性を評価するために実験を行った。実験には Epinions データセット (<http://www.trustlet.org/epinions.html>) を使用する。このデータセットは Paolo Massa が 1999 年に設立されたレビューサイト Epinions.com で 5 週間収集および整理したデータセットである。このデータセットには以下の情報が含まれている。

- 少なくとも 1 回評価した 49,290 人のユーザ
- 少なくとも 1 回評価された 139,738 個のアイテム
- 664,824 の評価 (5 レベル)
- 487,181 件のフォロー関係

本研究では、データセットから 300 人のユーザを選択しました。これらのユーザには 5,813 のフォロー関係がある。データスパースネスは約 93.5% である。これに基づいて予測を行う。Python 言語を使ってアルゴリズムを実装する。

潜在的なフォロー関係のスコアを予測することに加えて、元の 5,813 件のフォロー関係の信頼度を 1 つずつ 0 に設定して予測し、前者と比較する。各ユーザに対して、予測スコア topN を友人推薦リストとして、N

は 5, 10, 15, 20, 25, 30 を取り、適合率 (Precision) と再現率 (Recall) を使用して評価する [14]。

- 適合率 = 推薦のユーザはフォロー関係がある数 / 推薦のユーザ数
- 再現率 = 推薦のユーザはフォロー関係がある数 / 全部のフォロー数

そして、PeopleRank 法で得られた結果と比較する。

5. 実験結果

データセットが小さいので、潜在的なフォロー関係をスコア予測する際に、類似ユーザ topN または全部の類似ユーザを使っても、最後の結果に大きな違いはない。以下の結果は全部の類似ユーザを使用して予測したものである。

5,813 件のフォロー関係によると、0 以上の 56,245 件の潜在的なフォロー関係スコアが予測される。元のフォロー関係の予測スコアとの分布は図 4 の通りである。縦軸は確率密度で、両者の面積は正規化して 1 になる。元のフォロー関係の予測スコアの分布が高く、提案手法の有効性を示している。

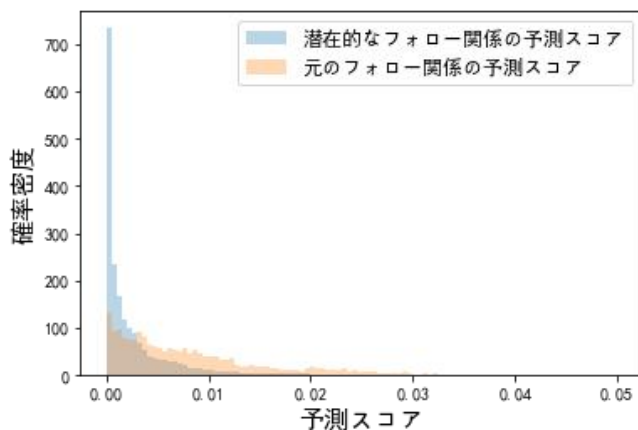


図 4. 予測スコアの分布

予測スコア topN を友人推薦リストとして使用する場合の評価を図 5 に示す。予測スコア top5 のユーザを推薦すると適合率は 50% を超える。また、データセット (300 人) が小さい原因かもしれませんが、PeopleRank 方法を利用した場合と比較を行ったところ、PeopleRank よりも高精度で推薦が可能であった。これは、潜在的なフォロー関係の予測手法が小データセットにおいて、より効果的に表現されていることを示している。

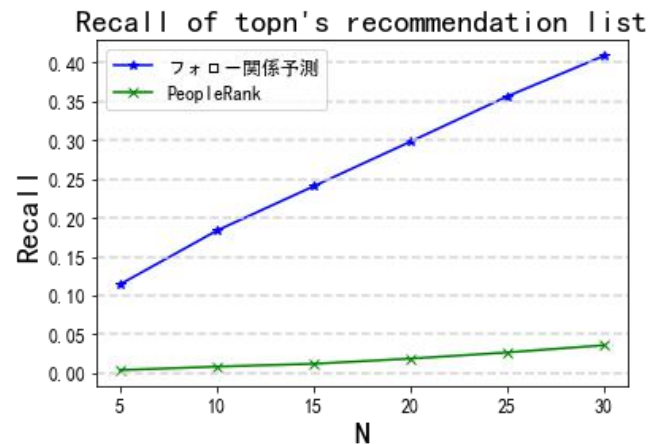
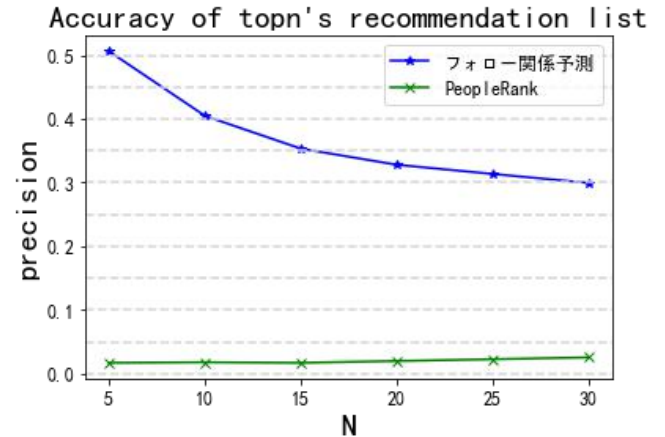


図 5. 推薦リストの評価

6. まとめ

本論文では、潜在的な信頼ネットワークを予測することに基づいた友人推薦手法を提案し、その有効性を評価した。潜在的なフォロー関係を予測する際にユーザ間の類似度と信頼度を利用する妥当性を説明し、フォロー関係行列に基づく潜在的なフォロー関係予測手法を提案した。実験結果を通して手法の有効性を証明し、50%以上の推薦精度を達成した。

今回の実験では 300 人のデータセットしか使用しなかったため、予測で使用した類似ユーザ topN の数が結果に大きな影響を与えていない。今後、サイズの異なるデータセットで実験を行い、手法の有効性を検討する。そして、実際に Twitter や他の SNS でデータを収集し、本研究の手法を適用して検証する必要がある。

また、本実験の推薦リストの精度も高くない。本研究はまた、ユーザーアイテム評価マトリックスに結合するなど、様々な方法を用いて改善する必要がある。

謝辞

本研究は JSPS 科研費 19H04219 の助成を受けたものです。

参 考 文 献

- [1] Chen Yun. Research on Behavior of China Social Networking Application users[J]. China Internet Network Information Center, 2014 (01): 57-63.
- [2] Recommender Systems Handbook. Bracha Shapira. Springer-Verlag New York Inc . 2010
- [3] Shriver, Scott K., Harikesh S. Nair, and Reto Hofstetter. "Social ties and user-generated content: Evidence from an online social network." *Management Science* 59.6 (2013): 1425-1443.
- [4] Huang, W.-H & Meng, X.-W & Wang, L.-C. (2011). A Collaborative Filtering Algorithm Based on users' Social Relationship Mining in Mobile Communication Network. *J. Electron. Inf. Technol.* 33. 3002-3007. 10.3724/SP.J.1146.2011.00364.
- [5] Dou Lisha. Research On Collaborative Recommendation Algorithm Based On Probabilistic Matrix Factorization[D]. Yanshan university. 2015.
- [6] Frank Edward Walter, Stefano Battiston, Frank Schweitzer. A model of a trust-based recommendation system on a social network. 2008, 16(1):57-74.
- [7] Chen Wenjun, Ni Jing. Research on recommendation model based on trust transitivity in social network environment[J]. 2020, 37(05):1364-1367.
- [8] Gao Yongbing, Yang Honglei, Liu Chunxiang, Hu Wenjiang. Friends recommendation algorithm based on the content and social filtering[J]. 2013, 32(14):75-78+82.
- [9] WENG, Jianshu; LIM, Ee Peng; JIANG, Jing; and HE, Qi. Twitterank: Finding topic-sensitive influential Twitterers. (2010). *Proceedings of the Third ACM International Conference on Web Search & Data Mining*: February 3-6, 2010, New York. 261-270. Research Collection School Of Information Systems.
- [10] Wang Ping, Long Yihong, Tang Zhihong, Liu Xu. Study on the Social Relation Based Internet Trust Establishment[J]. 2011, 32(04):12-15.
- [11] C. Chen and H. Feng, "MicroBlog recommendation based on user interaction," *Proceedings of 2012 2nd International Conference on Computer Science and Network Technology*, Changchun, 2012, pp. 2107-2111.
doi: 10.1109/ICCSNT.2012.6526334
- [12] Jing, N., et al. "Friend Recommendation Algorithm based on user Relations in Social Networks." *Chin. J. Manag. Sci* 25 (2017): 164-171.
- [13] Jian Wu, Francisco Chiclana, Enrique Herrera-Viedma. Trust based consensus model for social network in an incomplete linguistic information context. 2015, 35:827-839.
- [14] Bin, H.. "Friend recommendation algorithm based on association rules and tags." *Computer Engineering and Science* (2013): n. pag.