

興味領域とキーワードを考慮した Twitter ユーザ推薦に関する考察

牧岡 竜人[†] 平川 豊[‡] 大関 和夫[‡]

芝浦工業大学大学院理工学研究科[†] 芝浦工業大学工学部[‡]

1. はじめに

近年、Social Networking Service(SNS)の利用者が増加しており、中でも Twitter[1]では全世界で約 2 億 8600 万人、日本で約 2590 万人が利用しているとされている。

Twitter では、ユーザをフォローすることによって、ユーザ同士のコミュニケーションや情報収集が行える。しかし、嗜好に合うユーザを探す場合、興味関心のあるキーワードを呟いているユーザを検索しても膨大な検索結果の中からどのユーザがフォローするのに有益なユーザか判断し難い。

本研究ではユーザが関心あるキーワードを用いた Twitter ユーザ推薦システムを提案する。

2. 関連研究

久米ら[2]は、興味領域を考慮した Twitter ユーザ推薦手法を提案しており、被推薦ユーザのツイートから興味のあるカテゴリ（スポーツや音楽等）を抽出し、そのカテゴリとツイートからユーザ間の類似度を算出する事で趣味嗜好が合うユーザの推薦を行っている。

しかし、あるキーワードに対して適切なユーザを推薦する事は考慮されていないため、この手法をそのまま用いることは難しい。

3. 提案手法

ここで、あるキーワードを呟いているユーザ群から共通してフォローされているユーザが存在すれば、そのキーワードの境界で着目されているユーザであり、推薦するにあたって適切なユーザではないか、という考えに基づいて手法を提案する。

本研究で提案するシステムの概要を図 1 に示す。

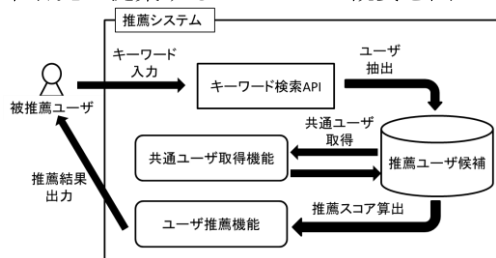


図 1 推薦システムの概要図

3.1. 推薦ユーザ候補の抽出

被推薦ユーザに、興味のあるキーワードを入力してもらい、Twitter のキーワード検索 API を用い

て検索を行って投稿ツイートを最大 2000 件取得する。取得した投稿ツイートからユーザを抽出し、そのキーワードにおける推薦ユーザ候補として保存する。

なお本研究においては、この推薦ユーザ候補内からユーザの推薦を行う。

3.2. 共通ユーザ取得機能

推薦ユーザ候補内のユーザにフォローされた人数を共通ユーザ数と定義し、推薦ユーザ候補内の全てのユーザに対する共通ユーザ数を取得する。例えば、推薦ユーザ候補に含まれるユーザ A が、他の推薦ユーザ候補内のユーザ 4 人にフォローされている場合ユーザ A の共通ユーザ数は 4 となる

3.3. ユーザ推薦機能

推薦ユーザ候補内のユーザからスコアを算出し、その値によって推薦ユーザの提示を行う。

ここで、単に共通ユーザ数の順位によって推薦してしまうと、上位ユーザにはそのキーワードに対して適切ではない著名人などが含まれると考えられる。そのため、著名人などフォロワー数の多いユーザが含まれていても均等な推薦が行えるよう、独自に共通スコアを算出して推薦を行った。ユーザ i の共通スコアは以下の計算式で算出する。

$$score_i = \frac{m_i}{f'_i + \min(f') + 1} \quad (1)$$

m_i は共通ユーザ数 $mutual_i$ を平均 \overline{mutual} と標準偏差 σ_{mutual} で正規化した値であり、以下の計算式で算出を行った。

$$m_i = \frac{mutual_i - \overline{mutual}}{\sigma_{mutual}} \quad (2)$$

f'_i はフォロワー数 $follower_i$ に共通ユーザ数の割合を乗算した値を取り、その値の平均 \overline{f} と標準偏差 σ_f で正規化を行った後、最大共通ユーザ数で割った値であり、以下の計算式で算出を行った。なお $\max(mutual)$ は最大共通ユーザ数を、 $\min(f'_i)$ は f'_i の最小値を指す。

$$f_i = follower_i \times \frac{mutual_i}{\max(mutual) + 1} \quad (3)$$

$$f'_i = \frac{f_i - \overline{f}}{\sigma_f} \quad (4)$$

3.4. 共通スコアと類似度を組み合わせた推薦

3.3 までで提案した共通ユーザによる推薦に、関連研究 2. にて紹介した興味領域を考慮した推薦手法を加えることによって、キーワードに対して適切なユーザかつ被推薦ユーザの趣味嗜好に合うユーザ

A Study on Twitter User Recommendation Considering category of interest and Keywords

[†]Ryuto Makioka, [‡]Yutaka Hirakawa, [‡]Kazuo Ohzeki

[†]Electrical Engineering and Computer Science, Shibaura Institute of Technology, Tokyo, Japan

[‡]Information Science and Engineering, Shibaura Institute of Technology, Tokyo, Japan

が推薦されると考えた。この場合キーワード以外の趣味嗜好も被推薦ユーザに合う、すなわち新たな趣味の発見や意外性のあるユーザを提示する可能性があると考えた。そのため提案手法のシステムに類似度を考慮した推薦機能を加える事で意外性を併せ持つユーザの推薦を試みた。

3.3で算出した共通スコアと関連手法2.によって算出した類似度を用いて、以下の計算式によるスコアを基にした推薦手法を新たに提案する。

$$score' = \alpha \frac{score - \min(score)}{\max(score) - \min(score)} + \beta r_i \quad (5)$$

$\min(score)$ 、 $\max(score)$ はそれぞれ共通スコアの最小値、最大値を指す。 r_i ($0 \leq r \leq 1$)は類似度を指す。 α ($0 \leq \alpha \leq 1$)は共通スコアに対する重み係数であり、 $\beta = 1 - \alpha$ は類似度に対する重み係数とする。事前実験として、 α の値を0.0~1.0まで0.1刻みで変化させ、著者のアカウントで推薦ユーザの評価を行ったところ、 $\alpha = 0.5$ において最善のユーザ推薦がなされた値となった。そのため、重み係数の値を $\alpha = 0.5$ 、 $\beta = 0.5$ とした。

4. 評価実験

キーワード検索結果において、以下の5つの手法による評価を行う。

- **提案手法1:** 共通スコアによる推薦
- **提案手法2:** 共通スコアと類似度の組み合わせによる推薦
- **既存手法1:** フォロワー数による推薦
- **既存手法2:** 類似度のみによる推薦
- **既存手法3:** フォロワー数と類似度の組み合わせによる推薦

既存手法1は、ユーザがキーワードに対して適切なユーザを探す場合に、一般的に考えられる指標としてフォロワー数を基にしたものである。既存手法2は類似度のみで推薦を行う事によって、関連研究との比較を考慮したものである。既存手法3は既存手法1,2の指標であるフォロワー数と類似度を組み合わせた場合での手法であり以下の計算式によって算出したスコアを基に推薦を行う。

$$score_f = \alpha \frac{follower_i - \min(follower)}{\max(follower) - \min(follower)} + \beta r_i \quad (6)$$

なおここで用いる重み係数 α と β は3.4にて述べた重み係数と同じ $\alpha = 0.5$ 、 $\beta = 0.5$ とした。

4.1. 評価項目

評価項目として以下の4つのアンケートを用意し、5段階評価で答えてもらう。

- A) キーワードに対して適切な推薦ユーザか
- B) 自分の趣味嗜好と合っているユーザか
- C) フォローしたいと思うか
- D) キーワード以外の趣味嗜好に興味持てたか

項目Aは適切な範囲のユーザを、項目B,Cは利用者が求めているユーザが推薦されているかを、項目Dは、3.4で述べた新たな趣味嗜好を持つユーザを

発見できるかを確かめるため設定した。

4.2. 実験結果

Twitterを利用しているユーザ4人を被験者として実験を行った。被験者のツイートと興味のあるキーワードを数個用意してもらい手法1~5においてそれぞれ上位10名の推薦ユーザを表示する。被験者は表示されたユーザ、10名×5手法の計50名に対しアンケートに答えてもらった。

以下に実験結果における手法1~5でのアンケートA~Dの結果の平均を取った表1を示す。

表1. 提案手法と既存手法の評価結果

アンケート 手法	A	B	C	D
提案手法1	3.825	3.313	3.050	2.763
提案手法2	3.850	3.313	3.075	2.788
既存手法1	2.450	2.363	2.188	2.338
既存手法2	2.550	2.750	2.688	2.975
既存手法3	2.475	2.488	2.325	2.475

アンケートA,B,Cの結果から、提案手法1,2どちらも全ての既存手法よりも高い評価を得ており、フォロワー数といった一般的な指標や、関連研究で紹介した類似度のみを用いた場合よりも適切なユーザが推薦される事がわかった。

アンケートDの結果は、提案手法1,2でほぼ同等の値が得られ、既存手法2の類似度を用いた手法において最高値となった。これは、3.4で提唱した仮説に反する結果となった。また、既存手法1と3を比較すると手法3のほうが高い値が得られた。これらの事から、類似度を考慮した場合、利用者にとって新たな趣味嗜好や意外性のあるユーザが推薦されやすくなると考えられる。

ここで提案手法2と既存手法2を比べると、アンケートAの結果に大きく差があるがDの結果にはそれほど差はない。そのため、3.4における類似度の重み係数の比重を適切に設定する事によって、新たな趣味嗜好を持ち、かつキーワードに適切なユーザをより推薦しやすくなると考えられる。

5. まとめと今後の課題

本研究では、Twitterにおいて興味のあるキーワードに関する適切なユーザを推薦する手段として、共通ユーザを利用したシステムを提案し、実装及び実験評価を行った。その結果、フォロワー数や類似度を指標とした推薦よりも評価の高いユーザを推薦する事が出来た。

今後の課題として、キーワードに対して適切かつ新たな興味領域を合わせ持つユーザを推薦できるように、共通ユーザスコア及び類似度に付与する重み係数の検討を行う。

参考文献

- [1] Twitter, <https://twitter.com/>
- [2] 久米雄介, 打矢隆弘, 内匠逸, “興味領域を考慮したTwitterユーザ推薦手法の提案と評価”, 研究報告知能システム(ICS), vol.2015, No.1, pp.1-8, 2015-03-13.