

ツイートの感情値とアカウントの不快要素を用いて 交流の減少を防ぐフォロー推奨・非推奨ユーザ提示手法

舛田悠[†] 塚田晃司[†]

和歌山大学システム工学部[†]

1. はじめに

Twitter では嗜好の近いユーザ同士がフォロー・フォロワーの関係となり、ツイートと呼ばれる140字以内の短文を投稿してコミュニケーションを取る。

新たにフォローするユーザを見つけるためには、タイムラインや検索、Twitter 公式が提供するおすすめユーザ等から自らユーザの選定を行う必要がある。しかし、アカウントホームの少ない情報からユーザの選定を行うことは、Twitter の利用期間の短いユーザほど困難である。フォロー後に相性が良くないと判明すると、フォローの解除等の関係悪化や疎遠に繋がる恐れがある。

上記を踏まえて、フォロー前後のトラブルを避けて長期的に交流可能なユーザを推薦するために、相性の良いフォローを推奨するユーザだけでなく推奨しないユーザも同時に提示する手法を提案する。

2. 関連研究

[1]では、Twitter におけるユーザのツイート内容や指定した単語に関してどれ程、どの時間帯にツイートしているかのデータを用いることで、ユーザ同士の相性を判断する手法を提案していた。

また[2]では、ツイートからユーザごとに全ての単語に対して感情情報を付加することで、単語に対して同じ感情を抱くユーザを推薦する手法を提案していた。

既存の研究では、推薦精度を向上させるために感情値と更に他のデータと組み合わせることが課題として挙げられていた。

3. 提案手法

今回実装したシステムでは、推薦を受ける対象ユーザ X と被推薦者である判定ユーザ Y のツイ

A new method for follower recommendation that Twitter beginner may or should not follow on the basis of emotional value of tweets and discomfort expression causing communication decrease.

Yu Masuda[†] Koji Tsukada[†]

Faculty of Systems Engineering Wakayama University[†]

ートやリツイートをもとめたツイートデータを用いて対象ユーザとツイートの傾向が近い判定ユーザをピックアップし、その中からフォロー推奨ユーザと非推奨ユーザを分類する。(図1)

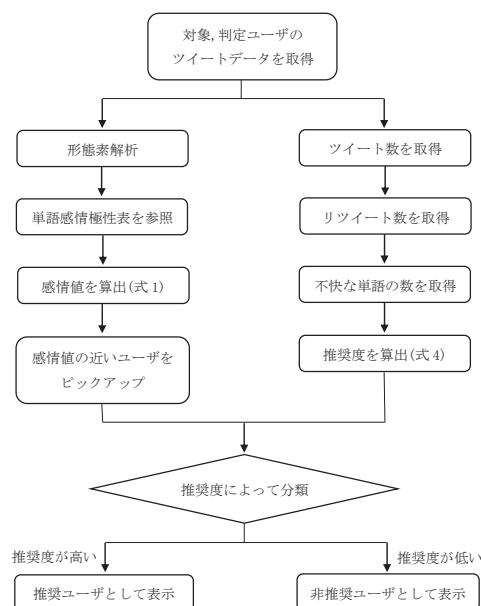


図1: システム概要

まず、ツイート内の単語ごとにどれほどポジティブ・ネガティブであるかを示した数値である感情値を算出する。これより、ユーザ同士のツイートやリツイートの内容の類似度を求める。

感情値は単語感情極性表[3]を用いて、全ての単語の感情値の平均から求める。更に式(1)よりユーザ間の感情値の距離 XY を求め、類似している度合いを判定するためにユーザ X, Y の感情類似度 F を式(2)で求める。

$$XY = \begin{cases} |x - y| (x > 0, y > 0 \text{ または } x < 0, y < 0) \\ |x| + |y| (x < 0, y > 0 \text{ または } x > 0, y < 0) \end{cases} \quad (1)$$

$$F = \frac{1}{(XY + 1)} \quad (2)$$

次にユーザの推奨度を算出する。推奨度とはユーザの1日の平均ツイート数等の要素を用いて、

ユーザ同士の相性を示した数値である。これより、ユーザ同士の Twitter の使い方の類似度を求める。

推奨度を求めるための要素として対象ユーザ X と判定ユーザ Y それぞれの 1 日の平均ツイート数 TX, TY, 1 日の平均リツイート数 RX, RY, 1 日の不快な単語の平均発言回数 GX, GY (以下平均 NG ワード数と言う) を用いる。この要素を決定するために [4]~[7] からフォローを解除する原因を調査し (表 1), 直接評価法により用いる要素と重み付けに利用する割合を選定した。重み付けの値はツイート数, リツイート数, NG ワード数のそれぞれを 0.109, 0.126, 0.291 とした。

表 1: フォローを解除する理由

理由	割合
誹謗中傷やネガティブなツイートが多い	29.1%
大量のリツイート	12.6%
大量のツイート	10.9%
話題が合わない	10.5%
しつこいリプライ	7.7%
自撮りツイート	6.9%
大量のハッシュタグの利用	6.5%
その他	15.7%

平均ツイート数はツイートデータの行数と、平均リツイート数はツイートデータの rt と表示されている行数と取得日数の平均から求めた。

NG ワード数は、単語感情極性対応表内で感情値が -0.99 以下の単語と [8] を参考に目に入ると不快と感じる可能性のある単語で構成した NG ワード表を参照して、取得日数との平均から求めた。

これを用いて重み付きユークリッド距離 E を式 (3) より、推奨度 R を式 (4) より求める。

$$E = \sqrt{0.109 * (TX - TY)^2 + 0.126 * (RX - RY)^2 + 0.109 * (GX - GY)^2} \quad (3)$$

$$R = \frac{1}{E + 1} \quad (4)$$

以上より、感情類似度 F が 0.91 以上、推奨度 R が 0.53 以上の判定ユーザをフォロー推奨ユーザ、それ未満の判定ユーザをフォロー被推奨ユーザとして表示する。評価に用いた値は、事前に長期間相互フォローの関係にある、自分と Twitter 上の友人のツイートデータ 100 組からシステムと同様の方法で算出した結果の平均から求めた。

4. 評価

今回提案した手法を用いてフォロー推奨・被推奨ユーザをピックアップし、アンケート調査により表示されたユーザが適切かどうかを評価する。

ランダムに募集した Twitter ユーザと知人を対象としてそれぞれツイートデータを収集し、当手法を用いて推奨・非推奨ユーザどちらも 3 人ずつ提示した。そして提示した推奨・非推奨ユーザをフォローしたい・したくないと感じるか、ランダムにピックアップしたユーザとどちらをフォローしたいか等を問うアンケートを行った。

5. まとめ

今回の提案システムでは、ユーザの感情値と不快感に繋がる要素を用いることで、Twitter の利用期間の短いユーザがフォロー前後のトラブルを避けて長期的に交流可能なフォローユーザを選定できる手法を提案した。

今後の課題としては、Twitter API を導入した実用的なシステムの開発や、当手法を利用してフォローした対象ユーザの長期的な観測による評価等が挙げられる。

参考文献

- [1] 川口辰弥, 塚田晃司: Twitter における利用者に即したタイムラインを構成するためのユーザー評価推薦手法の提案, pp. 1-42 (2018)
- [2] 加藤慎一郎・濱川礼: Twitter 感情分析を用いた感情値可視化とユーザー推薦システム, 情報処理学会研究報告, pp. 1-8 (2013)
- [3] 高村大也: 単語感情極性対応表, 入手先<http://www.lr.pi.titech.ac.jp/%7Etakamura/pndic_ja.html> (参照 2019-12-26)
- [4] Twitter のフォローをやめる理由アンケート, 入手先<<https://news.mynavi.jp/article/20120616-twitter/>> (参照 2019-12-26)
- [5] 植田智明: Twitter における不愉快情報への接触を防ぐための応答メッセージ不可視化システムの提案, 情報処理学会研究報告, pp. 1-46 (2015)
- [6] 友達をリムーブした理由アンケート, 入手先<<https://cancam.jp/archives/270223>> (参照 2019-12-26)
- [7] 都筑学: 大学生における LINE や Twitter の利用目的とその心理についての研究, 中央大学保険体育研究所紀, pp. 23-27 (2017)
- [8] 悪口一覧, 入手先<<https://nomucom.com/提案/【悪口一覧】日本の悪口をア行からワ行まで200個>> (参照 2019-12-26)