

Twitterにおけるユーザが持つフォロー基準の自動推定

長谷川 智咲† 若原 徹†

法政大学 情報科学部†

1. まえがき

ユーザにフォロワー推薦をする従来手法としては、文献[1][2]のようにユーザの興味領域やユーザ間での共通話題を考慮する手法が主流である。[2]では話題の共通性に加えて、話題に対する感情的態度が類似するユーザをフォロワー候補として推薦する手法が提案された。しかし、被推薦ユーザが持つフォロー基準を考慮していないため、どんなユーザに対しても共通話題とそれに対する感情的態度の類似性のみを考慮した推薦となっており、共通話題をフォロー基準としていないユーザにとっては適切なフォロワー推薦とはならなかった。

本研究ではフォロー基準の自動推定手法を提案する。分析対象ユーザとそのフォロワーから情報を取得し、「ユーザが関心のある話題とその感情的態度」「フォロワーのツイート頻度」「公式アカウントのフォロー数」「フォロワーのツイートタイプ」「相互フォロー率」の計5つの指標から分析し、各ユーザが各指標をフォロー基準にしているかどうかを判別する。

2. データの収集

TwitterAPI[3]を用いて分析対象ユーザと分析対象ユーザのフォロワーからデータの収集を行う。

まず、データ収集の対象となる分析対象ユーザのフォロワー数 n は、標本調査において標本サイズを決定する公式[4]に従って決定した。

次いで、分析対象ユーザのフォロワー n 人と分析対象ユーザからタイムラインのツイート（リツイートを含む）をそれぞれ200件ずつ取得した。具体的には、ユーザID、ユーザのスクリーンネーム、ツイートID、ツイート本文、ツイート日時、さらに、それぞれのユーザのプロフィール文とそのユーザが認証済みアカウントかどうかについての情報を取得した。

3. 提案手法

本研究では、収集したデータからフォロー基準に関して、以下に記す5つの指標の情報を取得する。

3.1. 関心のある話題とそれに対する感情的態度

まず、話題に対する感情的態度をネガティブ、ニュートラル、ポジティブの3つで定めるために、日本語評価極性辞書[5]を用いる。さらに、日本語WordNet[6]を用いて当該辞書の拡張を行った結果、ネガティブを表す感情語は8150語から8784語、ニュートラルは4973語から10133語、ポジティブは5470語から7126語となった。

次に、各ユーザのツイートとプロフィールを”はてなキーワード自動リンクAPI[7]”に与え、ツイートとプロフィールが持つキーワードとカテゴリを取得する。その際、

形態素解析ツール”MeCab”の登録単語を拡張するシステム辞書”mecab-ipadic-Neologd[8]”を用いて、一般名詞と固有名詞のみを取得してキーワードとする。そして、各ユーザで出現回数が多い上位10個のキーワードを特徴語の候補となるキーワードとした。

続いて、キーワード数をさらに絞り込むためのキーワード評価値を提案する。文献[1]に従って算出したキーワード k の興味領域度 I_k と、キーワード k の出現回数 n_k を用いて、キーワード評価値 E_k を次式で計算する。

$$E_k = \frac{n_k}{\bar{n} + \sigma_n} + \frac{I_k}{\bar{I} + \sigma_I} \quad (1)$$

ただし、 \bar{I}, σ_I は I_k の平均と標準偏差、 \bar{n}, σ_n は n_k の平均と標準偏差である。当該評価値が高いキーワード上位20個を分析対象ユーザにとって関心のある話題（特徴語）であるとした。

最後に、得られた特徴語に付帯する単語の感情値を決定する。まず、得られた特徴語を含むツイートをすべて取得し、”MeCab”を用いて形態素解析を行う。形態素解析から得られた語が前述の感情語データベースに存在する場合、PN値を取得し、足し合わせる。これを取得したツイート全てに対して行い、PN値の総和を取得したツイート数で割った値が、特徴語に対する感情値となる。算出された感情値が -0.1 未満の場合はネガティブ、 $[-0.1, +0.1]$ の範囲をニュートラル、 $+0.1$ を超える場合はポジティブと定めた。

3.2. ツイート頻度

分析対象の各ユーザのフォロワー n 人についてツイート頻度（1日当たりの平均ツイート数）の標準偏差 σ_t を求める。ツイート頻度をフォロー基準としているかどうかは、当該ユーザのフォロワーのツイート頻度の標準偏差が小さければフォロー基準としている、大きければフォロー基準にはしていないと判定することとする。

3.3. 公式アカウントのフォロー数

TwitterAPIでユーザプロフィールを取得する際に、そのアカウントが認証済みアカウントかどうか取得した。しかし認証済みアカウントではないが企業や芸能人等の公式アカウントもある。そこで、文献[9]に従い、「公式（非公式は除く）」「オフィシャル」「official」という単語をプロフィールに含むユーザは認証済みアカウントでなくとも公式アカウントとみなすことにした。上記手順により、公式アカウントの数を算出することができる。公式アカウントのフォロー数がある閾値以上のとき、分析対象ユーザは公式アカウントを多くフォローしている、閾値未満のとき、分析対象ユーザは公式アカウントを多くフォローしていない、と判定することとする。

3.4. ユーザのツイートタイプ

本研究では、ツイートを次の3タイプに分けて考える。
タイプ1: 情報発信・共有をしているツイート（公式アカウントのツイートとリツイート）

タイプ2: ユーザ同士のコミュニケーションをしているツイート（リプライ）

タイプ3: 純粋なツイート (タイプ1, 2を除いた, ある話題に対しての感想や日常生活のツイート等)

次に, 分析対象ユーザのフォロワーのツイートから, 当てはまる各タイプのツイート数を算出し, そこから各タイプの標準偏差を算出する.

そして, フォロー基準であるかどうかの判定は, 各タイプの標準偏差が小さければそのタイプをフォロー基準としているとし, 標準偏差が大きければそのタイプはフォロー基準としていないとみなすことにより行う.

3.5. 相互フォロー率

相互フォロー率は, 分析対象ユーザの全フォロワー数を分母, 分析対象ユーザをフォローしてくれているフォロワー数を分子とし, 比で定義することとする.

フォロー基準であるかどうかの判定は, 相互フォロー率が高い場合は相互フォローを望んでいる, つまり相互フォローしてもらうことをフォロー基準としているとし, 相互フォロー率が低い場合は相互フォローしてもらうことをフォロー基準としていないとする, ことにより行う.

4. 評価実験

Twitterを利用しているユーザ15名を被験者として, 提案システムの評価実験を行った.

設問Q1~Q5において, 推定された5つのフォロー基準に対して, 自らが持っていたフォロー基準とどの程度一致したかを, 1~5点の5段階評価で回答させた.

さらに設問Q6において, 提案手法により推定されたフォロー基準に基づいたユーザ推薦システムを使ってみたと思うかを, 「はい」, 「いいえ」で回答させた.

表1に, Q1~Q5のアンケート結果の全体評価を示す.

表2に, Q6のアンケート結果を示す.

表1 提案手法の全体評価

フォロー基準の指標	評価
Q1. 関心のある話題とその話題に対する感情	3.47
Q2. ユーザのツイート頻度	3.33
Q3. 公式アカウントのフォロー数	3.8
Q4. ユーザのツイートタイプ	4.13
Q5. 相互フォロー率	4.53

表2 提案手法の今後の有効性の評価

Q6. 提案手法を用いたユーザ推薦システムを利用してみたいか	回答者数 (被験者15名)
はい	9
いいえ	6

5. 考察

表1より, 「関心のある話題とその話題に対する感情」では全体評価が3.47となった. この指標の精度を上げるためには, まず感情語データベースのさらなる拡張が考えられる. 次に, ユーザの特徴語で「しま」や「ちゃん」などが取得されてしまうことを防ぐため, 固有名詞のみ取得して特徴語候補とすることが考えられる. さらに, 式(1)で定義したキーワード評価値の妥当性や適用法を再検討する必要がある.

「ユーザのツイート頻度」では, 全体評価が3.33となった. このような評価になった理由は2点考えられる. 1点は, フォロー基準の判別で用いた閾値が不適切であった可能性である. もう1点は, ユーザ自身がツイート頻度をフォロー基準としていなかったが, フォロワーのツイート頻度に偶然偏りがあり, 推定結果ではフォロー基準と推定されてしまった可能性である.

「公式アカウントのフォロー数」では, 全体評価が3.8となった. この指標の精度を上げるためには, まずフォロー基準の判別で用いた閾値の見直す必要がある. さらに, 公式アカウントと判断する要素を増やすことが考えられる. 公式アカウントはフォロワーとフォロワーの数に極端な差があるという特徴が有効かもしれない.

「ユーザのツイートタイプ」では, 全体評価が4.13, 「相互フォロー率」では, 4.53となった. これらは, フォロー基準として有効であったと判断できる.

表2より, 「この提案手法を用いたユーザ推薦システムを利用してみたいか」という設問に対して「はい」が6割, 「いいえ」が4割であったことより, 本研究で目指したフォロー基準の自動推定能力がユーザ推薦システムの有用性を大きく向上する可能性があると言える.

6. 結び

本研究では, Twitterのユーザが持つフォロー基準の自動推定手法を提案した. 評価実験より, 提案手法を用いたユーザ推薦システムを利用したいと考えるユーザはまだ6割程度であり, 「関心のある話題とその話題に対する感情」, 「ユーザのツイート頻度」, 「公式アカウントのフォロー数」がフォロー基準かどうかの推定方法に改善が必要であることが明らかになった. これらの改善により, 提案手法を用いたフォロワー推薦技術を展開することが今後の課題である.

文 献

- [1] 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”, 情報処理学会研究報告, vol. 2015-ICS-179, no. 1, 2015.
- [2] 熊本忠彦, 灘本明代, “共通話題に対する感情の態度の類似度に基づくフォロワー推薦”, 電子情報通信学会論文誌D, vol. J100-D, no. 4, pp. 500-509, 2017.
- [3] “Twitter Developers”, <https://developer.twitter.com/>
- [4] 福祉サービス分析法 (標本サイズの決定), http://www.rikkyo.ne.jp/~ssakata/class/analysis/2002/07/sample_size.htm
- [5] “日本語評価極性辞書”, <http://www.cl.ecei.tohoku.ac.jp/index.php?Open%20Resources%2FJapanese%20Sentiment%20Polarity%20Dictionary>
- [6] “日本語 WordNet”, <http://compling.hss.ntu.edu.sg/wnja/>
- [7] “はてなキーワード自動リンク API”, <http://developer.hatena.ne.jp/ja/documents/keyword/apis/autolink>
- [8] MeCab用システム辞書“mecab-ipadic-Neologd”, <https://github.com/neologd/mecab-ipadic-neologd/blob/master/README.ja.md>
- [9] 鈴木祥平, 倉田陽平, 石川博, “Twitterのユーザプロフィールを用いた公式アカウントの検出とフォロー形成パターンの分析”, DEIM Forum 2017 XX-Y