

ユーザの関連度を利用したリツイートの推薦手法の提案

情報・通信工学専攻 学籍番号：1531055 寺田研究室 鈴木佑樹 指導教員 寺田 実, 沼尾 雅之

1 背景

Twitterのタイムラインから有益な情報を得るための手段としてリツイートという機能が存在する。これを使うと、自分がフォローしているユーザ(以下フォロワー)以外が発言した内容でも、自分のフォロワーがその発言をリツイートすれば、自分のタイムラインにその発言が表示される。大多数のユーザに拡散される情報は、それだけ情報としての価値が高いことが殆どであるが、興味の不一致などにより、その情報が必ずしも拡散者のフォロワー全員にとって有益な情報であるとは限らない。

2 目的

本研究の目的は、タイムラインに存在するリツイートのうち、自分が有益であると感じる内容を拡散者の情報を基準に推薦を行うことである。

有益な情報は自分と興味内容の近いユーザほど、リツイートの内容も興味があるものを流してくれるのではないだろうかと考えた。そこで本研究ではユーザの関連度を独自の尺度で捉え、その関連度を利用して各ユーザに対して有益なリツイートを推薦する。

3 関連研究

Chunliang[1]らによる研究では、ユーザがそれまで行ったツイートからユーザのプロファイルを構築し、ユーザのタイムライン上から特に関心の高いツイートを推薦するWikiProfileと呼ばれるモデルの提案を行った。Wikipediaの記事間のリンクデータを使って単語ごとの関連度を算出し、ユーザのプロファイル構築を行っている。本研究とは、ユーザの興味によってツイートを推薦するという点では共通するが、Chunliangの研究ではツイートの内容によるプロファイルによって判断するため、本研究と興味の基準が異なる。

近藤[2]らは、ユーザとそのユーザのフォロワーのツイートをを用いて、LDAに基づく興味推定手法の提案を行った。この研究では興味推薦の部分に協調フィルタリングやtf-idfを用いており、また研究内の実験の中でtf-idfよりも協調フィルタリングの方が高い精度の推薦ができたという結果を得られている。

これらのツイートのテキストを解析する推薦では、ツイート内の文字数が特に少ない場合解析が困難であるため、本研究ではツイートのテキストを解析せずに推薦ができないかを検証する。

4 提案手法

本研究では、ユーザ関連度を利用して、フォロワーの行ったリツイートの中からユーザが有益であると思えるリ

ツイートを推薦するシステムを提案する。

システムに推薦されるユーザ(以下、被推薦ユーザと呼ぶ)とそのフォロワーの間でフォロワー一致率を求める。この値をもとに、そのフォロワーからのリツイートからのリツイートに対して推薦スコアを与える。なお、同一のツイートを複数のフォロワーがリツイートした場合は、そのフォロワー全員のフォロワー一致率を足し合わせたものをそのリツイートの推薦スコアとした。

4.1 フォロワー一致率

フォロワー一致率は以下の式で求められる。ここで f_A, f_B は、それぞれユーザA, Bのフォロワー集合である。

$$F_{AB} = \frac{|f_A \cap f_B|}{|f_A \cup f_B|} \quad (1)$$

筆者の卒業研究[3]内では、このパラメータがユーザとの関連度に比例することを示した。

4.2 スコア

特定のツイート T をリツイートしたフォロワーの集合を $U = \{u_1, u_2, \dots, u_n\}$ 、被推薦者とそのとあるフォロワー i とのフォロワー一致率を $F(u_i)$ と表すとすると、そのツイート T に対する推薦スコアは以下の式で表される。

$$Score(T) = \sum_{i=0}^n F(u_i) \quad (2)$$

また、被推薦ユーザと各リツイートの発言者の関係も考慮し、二者間のフォロワー一致率をサブスコアとし、考慮する場合は本来のスコアにかけ合わせたものを推薦スコアとした。サブスコアを考慮した場合のスコアの計算式は、以下の式で表される。

$$Score'(T) = Score(T) * F(u_T) \quad (3)$$

4.3 ノイズ情報の除去

フォロワーのリツイートから有益な情報のみを推薦するために、リツイート数が50に達していないリツイートは、推薦の候補に含めないようにした。このノイズ情報については、リツイート数毎の有益度の分布を調べるための予備実験を行い、有益なリツイートの割合が少ないリツイート数帯を検証した予備実験の結果に基づくものである。

5 評価実験

本システムの推薦アルゴリズムで推薦したリツイートが、被推薦ユーザにとって有益なリツイートであるか、を検証するために実験を行った。

1つ目の実験では、推薦アルゴリズムの妥当性を評価するために、ネット上から無作為に選出したTwitterユーザのフォロワーデータを利用して実験を行った。

この実験では被推薦ユーザのフォロワーが行ったリツイートのうち、被推薦ユーザに対して推薦スコアの高いリツイートを実際に被推薦ユーザがリツイートしているかによって推薦アルゴリズムの有用性を評価する。この実験の結果を表 1, 2 に示した。

表 1 実験 1 においてサブスコア考慮なし、推薦閾値一定場合の結果

ユーザ	ユーザ RT 数	推薦 RT 数	一致	適合率	再現率	F 値
A	22	24	8	0.33	0.36	0.35
B	8	42	5	0.12	0.63	0.2
C	18	27	4	0.15	0.22	0.18
D	27	17	6	0.35	0.22	0.27
E	15	30	9	0.3	0.6	0.4
F	18	33	4	0.12	0.22	0.16
G	17	13	5	0.38	0.29	0.33
H	26	43	10	0.23	0.38	0.29
I	20	31	6	0.19	0.3	0.24
J	24	35	7	0.2	0.29	0.24

表 2 実験 1 においてサブスコア考慮あり、推薦閾値一定場合の結果

ユーザ	ユーザ RT 数	推薦 RT 数	一致	適合率	再現率	F 値
A	22	18	3	0.17	0.14	0.15
B	8	17	2	0.12	0.25	0.16
C	18	29	4	0.14	0.22	0.17
D	27	20	4	0.2	0.15	0.17
E	15	24	7	0.29	0.47	0.36
F	18	22	4	0.18	0.22	0.2
G	17	24	7	0.29	0.41	0.34
H	26	40	7	0.18	0.27	0.21
I	20	22	3	0.14	0.15	0.14
J	24	37	4	0.11	0.17	0.13

2 つ目の実験では、被験者に推薦されるリツイートに対して主観評価を行わせて推薦スコアとの相関を評価し、推薦システムの推薦するリツイートが有益であるかを検証した。

6 人の被験者のアカウントに対し、この推薦アルゴリズムを試してもらった。ランダムに表示されたリツイートを横に設置したラジオボタンで有益度を評価してもらう。被験者に有益であると判断されたリツイートが推薦スコアの高いものであれば、システムの推薦アルゴリズムが有用であると考えられる。

今回の実験の結果について表 3 に示す。またこのうち、文字数が少ないツイート群についての結果を表 4 に示した。

6 考察

サブスコアを考慮した場合、実験 1, 実験 2 の両方において、そうでない場合よりも値をはじめとしたパラメータの

表 3 実験 2 の有益度と推薦スコアの相関係数

ユーザ	有益度相関	サブスコア考慮 有益度相関
a	0.505	0.397
b	0.325	0.259
c	0.604	0.565
d	0.449	0.225
e	0.434	0.337
f	0.472	0.383

表 4 実験 2 における URL を除いた文字数が少ないツイートについての有益度とスコアの相関

文字数	個数	有益度相関	サブスコア考慮 有益度相関
0	14	0.603	0.409
1 ~ 10	38	0.516	0.339

値が低くなっている。通常の推薦スコアは、被推薦ユーザとそのフォロワーの一致率によって求められるものであるのに対し、サブスコアは被推薦ユーザとリツイートの発言者のフォロワーの一致率によって決まるパラメータである。今回の結果を見る限り、それを発言したユーザよりリツイートしたユーザとのユーザの類似度を重視した方がより有益である結果が得られると考えられる。

7 結論

提案した推薦アルゴリズムを 2 つの実験によってその有用性を評価し、提案手法で有益なリツイートの推薦が可能であることを示した。文字数の少ないリツイートに関しても、これらの提案手法が有効であり、テキストの解析を行わずとも推薦を行えることを示した。しかし推薦結果には、拾えきれていない正解データである False Negative も多く見受けられ、これを改善する為の方法を模索する必要がある。

参考文献

- [1] Chunliang Lu, “Twitter User Modeling and Tweets Recommendation Based on Wikipedia Concept Graph”. Intelligent Techniques for Web Personalization and Recommender Systems AAAI Technical Report, WS-12-09, 2012.
- [2] 近藤 直人, “Twitter を用いた LDA に基づくユーザの興味推定手法”. 言語処理学会 第 21 回年次大会発表論文集, 2015.
- [3] 鈴木 佑樹, “フォロワー一致率に基づくタイムラインの 3D 表示の提案”. 電気通信大学 情報理工学部 情報・通信工学科 卒業論文, 2015.

対外発表

本研究は、DEIM2017 にて、登壇発表を行う予定。