

ユーザの嗜好と話題の新鮮度を考慮したツイートの推薦

横山 慎[†] 馬 強[†]

[†] 京都大学大学院 情報学研究科 〒 606-8501 京都府京都市左京区吉田本町

E-mail: [†]yokoyama@db.soc.i.kyoto-u.ac.jp, ^{††}qiang@i.kyoto-u.ac.jp

あらまし マイクロブログを利用して情報収集を行う場合、ユーザが膨大な情報を取捨選択し興味のある情報を過不足なく入手することは通常困難である。ツイート推薦に関する研究は数多くあるが、これらは多くの場合ユーザの嗜好に基づく推薦である。このような手法により推薦されるツイートはユーザにとって興味ある内容を含む重要な情報である可能性が高い反面、それぞれ似通ったものを推薦してしまい、結果的にすでに持っている情報とほとんど同じものしか得られない可能性がある。本研究ではユーザの嗜好と、情報の真新しさを表す新鮮度によるツイートの推薦手法を提案し、ユーザの多様な情報の収集を支援する。本研究における新鮮度とはユーザにとって見慣れない珍しい情報を高価値とみなすような尺度で、ツイート内容が普段ユーザがタイムラインで眺める内容とどのくらい異なっているかをトピックを考慮し推定することで与える。評価実験では、新鮮度を計算する際にトピックを考慮するべきであることをベースラインを用いて示した上で、更にユーザ嗜好度を用いた推薦を提案し、提案手法と比較した上でそれぞれ想定通りに真新しい情報や興味のある情報をよく推薦する傾向にあることを確認した。また、新鮮度・ユーザ嗜好度を同時に考慮した混合推薦手法を提案し、新鮮度のみによる推薦、ユーザ嗜好度のみによる推薦との性能比較を行ったが、推薦能力に大きな変化はなかった。

キーワード twitter, 情報推薦, 新鮮度

1. はじめに

Twitter を始めとしたソーシャルメディアがインターネットの発展とともに急速に普及しつつある [1]。その重要な利用目的の一つに情報収集があり、特に若者の間では既存メディアに劣らない重要な情報収集手段の一つとなりつつある [2]。Twitter の情報源としての側面に注目すると、Twitter では毎日 1 億人以上のユーザから 3 億以上のツイートが投稿されている [3]。個人が情報収集を行うためには、この莫大な情報量に対処するため何らかの取捨選択が必要である。

Twitter にはフォローという仕組みがあり、ユーザはユーザの投稿したツイートのみを受け取り、閲覧することができる。しかしユーザ数が膨大であること、自らにとって必要なツイートしか投稿しないようなユーザは僅かであることから、必要な情報を過不足なく得られるようユーザを選び取りフォローすることは通常困難である。このような背景から、Twitter において、ユーザにとってフォローすべきユーザや見るべきツイートを推薦する手法は数多く提案されている [4], [5], [6], [7]。

既存のツイート推薦手法では、文章間の類似度やトピックモデルを用いるものや、価値ある投稿の多いユーザを特定し、新しいツイートを検知し推薦するといった手法が主流であるが、これらは多くの場合ユーザの嗜好に基づく推薦である。このような手法により推薦されるツイートはユーザにとって興味ある内容を含む重要な情報である可能性が高い一方で、それぞれ似通ったものを推薦してしまい、結果的にすでに持っている情報とほとんど同じものしか得られない可能性がある。

また、ユーザの情報のニーズは人によって多種多様である。

図 1 は Twitter におけるユーザのツイート例であるが、A のようにユーザにとって未知、つまり新鮮度の高い情報が欲しいという場合があれば、B,C のように既存手法により推薦されるようなユーザの嗜好に合った話題に関する情報や、一般の人々の中で話題となっているトレンド情報などについてもニーズがあると考えられる。

そのため本研究ではユーザの嗜好だけでなく情報の真新しさを示す尺度である新鮮度を考慮した推薦手法を提案し、これを

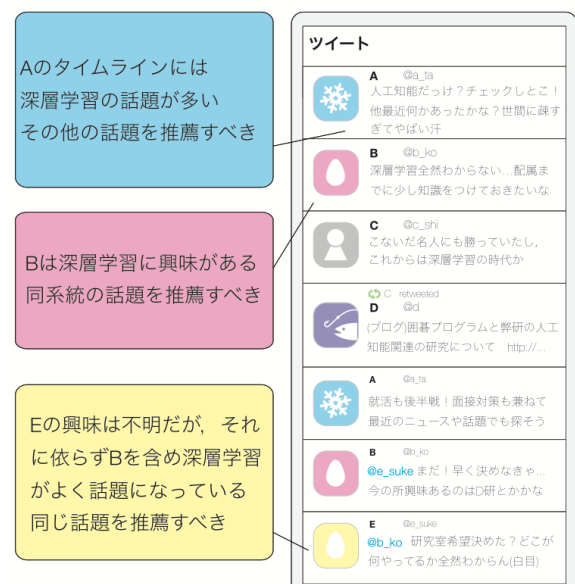


図 1 タイムライン例から見る複数種の推薦のニーズの例

用いたツイートの推薦を試みる。提案手法は、ユーザの嗜好に合った興味深いツイートのうち、なるべく今まで見ていないような内容の情報を推薦することで、全体として得られる情報の網羅性の改善を支援することができると期待できる。

本研究における新鮮度とはユーザにとっての新鮮度である。一般的な新鮮度とは対象の時間経過に伴う価値損失の少なさを示す尺度であるが、本研究では特にユーザにとっての情報の内容の真新しさを推定する。つまり本研究で提案する新鮮度は、新しい情報に加え、新しくなくともユーザにとって見慣れない珍しい情報であればこれを高価値とみなすような尺度である。

ユーザの嗜好を考慮せず、ユーザにとっての新鮮度のみに基づきツイートを推薦する手法も考えられるが、これは得られる情報の多様化に貢献できると期待される反面、ユーザが必要としない情報を推薦してしまう可能性がある。この問題を解決しつつ、ユーザの複数の情報のニーズを含意した手法として、ユーザの嗜好、ユーザにとっての新鮮度を同時に考慮した混合推薦手法を提案することが本研究の目的である。

本論文ではまず第2.節でTwitterにおける情報推薦、ユーザの嗜好に基づく推薦、情報新鮮度にかかわる関連研究について述べた後、第3.節でユーザ嗜好度、情報新鮮度の計算手法を提案する。第4.節で実験にてそれぞれによる推薦手法の妥当性を検証した後、これらを組み合わせた混合推薦手法についても言及、推薦能力を比較する。

2. 関連研究

2.1 Twitterにおける情報推薦

Liら[4]はフォロワーするユーザやコミュニティを推薦するフレームワーク『FRRec』を提案している。ユーザのフォロー関係を示す隣接行列を非負値行列因子分解(Nonnegative Matrix Factorization, NMF)してコミュニティを発見し、ユーザのツイート群とコミュニティ帰属を用いてトピックモデルを提案している[8],[9]。

Wengら[5]は、もっともフォロワー、ツイートの多いユーザ群とそのフォロワー、ツイート等の情報を入力に、ユーザの影響度を推定する手法を提案している。ユーザの類似度を求め、ユーザのフォロー関係をリンクとみなし、リンクの重みをこの類似度を用いて表すことでTwitterにおいてPagerank[10]を適用し、ユーザの影響度を推定し推薦している。

さらに、Uysalら[6]は、ツイート著者の特徴、閲覧者の特徴、ツイートの内容や投稿時間等を分析してツイートの推薦を行う手法を提案している。

2.2 ユーザの嗜好に基づく推薦

Yangら[7]はトピックモデルを用いたツイートのユーザ嗜好度を決定手法を提案している。トピックモデルを用いてユーザとツイートの相関の強さを推定する際、トピック分布の時間的、空間的エントロピーを考慮したペナルティを考慮することで、トピック推定に失敗している部分の影響を軽減している。

2.3 情報の新鮮度

Kotkovら[11]は既存の情報推薦手法を調査し、真新しさや意外性に富んだ情報を推薦する能力の高い情報推薦手法の必要

性を訴えている。

Parkら[12]は、学術論文を入力に新規性の高い部分を推定する手法を提案している。新規性とは今までに言及されていない要素のことであり、情報新鮮度と近い概念である。情報検索の分野を始めとして、時間的・内容的に新しいことが重要な分野において、新規性に関する研究は盛んに行われている[13],[14]。

またMaら[15][16]は、サーバー側でWebサイトを監視し、発生した変更に対して4つの尺度を定義し解析を行い、ユーザーに変更の旨を通知するシステムを提案している。Maらは新鮮度、流行度、更新頻度、アクセス頻度を提案しているが、本研究に関連する新鮮度について以下に議論する。

Maらは変更前後のコンテンツの類似度に着目して、類似度が低い場合、ユーザにとっては従来から大きく様相の異なる見慣れない情報を新たに含んでおり、価値が高いとみなされる可能性が高いとしている。逆に類似度が高い場合、変更前のコンテンツに追記した(同じ文脈の情報を追加した)ものである可能性が高く、元々そのコンテンツ並びに文脈に関心の高いユーザにとっては価値が高いとみなされる可能性が高いと推定している。Maらはこれらを新鮮度・流行度と定義し、それぞれにおける価値計算手法を提案しているが、彼らの手法はキーワードベースであり、トピックを考慮していない。

以上のように主に情報検索の研究において活用されている概念である新鮮度を、本研究ではtwitterにおいてトピックモデルを用いて推定し、適用することを試みる。

3. 提案手法

3.1 概要

既存のツイート推薦手法は単語やトピックをもとにユーザの嗜好を同定しこれに合致するツイートを推薦するものが多く、これらの手法が有益な場合もあるものの、ユーザにとって新規性の高い有益な情報を得ることは困難である。そのため本研究ではユーザの嗜好だけでなく情報の真新しさを示す尺度である新鮮度を考慮した推薦手法を提案し、それを用いたツイートの推薦を試みる。

ユーザにとっての新鮮度は、ユーザとツイート群を入力に、ツイート内容が普段ユーザがタイムラインで眺める内容とどのくらい異なっているかを推定することで与える。ユーザ嗜好度は、ユーザとツイート群を入力に、ユーザが各トピックに対しどのくらい興味を持つか、ツイートは各トピックとどのくらい関係が強いかを推定することで与える。本研究ではユーザとトピックの相関の強さ、トピックと単語の相関の強さに基づいて、新鮮度や嗜好度を計算する手法を提案する。これらの相関は、本研究ではLiら[4]の提案するトピックモデルにより推定されるトピック分布などを用いて推定する。

3.2 単語・トピック分布の推定

Liら[4]の提案するトピックモデルに基づき、ユーザがどのような話題を投稿しているか推定する。本研究では図2に示すUser-Topic Model(以下、UTモデルと呼ぶ)を用いて、ユーザのフォロワーと投稿ツイート群を入力に、ユーザ-トピック分布 $p(z|u)$ 、トピック-単語分布 $p(w|z)$ を数値的に求める。ギブ

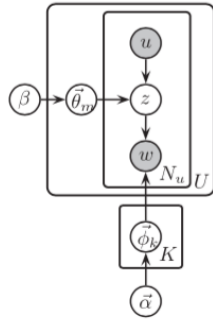


図2 User-Topic Model [4]

ズサンプリング法によりこれを与える場合、以下の更新式を用いる。

$$p(w | z) = \frac{n_z^w + \beta}{n_z^w + W\beta} \quad (1)$$

$$p(z | u) = \frac{n_z^u + \alpha}{n_z^u + K\alpha} \quad (2)$$

十分な反復回数をとって値の更新とサンプリングを繰り返し、これらの値を得る。ただし、 α , β はハイパーパラメータ、 W は単語種類数、 K はトピック数、 n_z^w はトピック z に単語 w が割り当てられた回数、 $n_z^w + \beta$ はトピック z に割り当てられた総単語数、 n_z^u はユーザ文書 (ユーザのツイート群) u においてトピック z が割り当てられた単語数、 $n_z^u + \alpha$ はユーザ文書 u の単語数である。

UT モデルの詳細は文献 [4] を参照されたい。

3.3 新鮮度

本節では Twitter におけるトピックを考慮したツイートの新鮮度計算手法を提案する。

ある情報がユーザにとって見慣れているかどうか判定することを考える際、もっとも直感的な実現方法として単語の出現回数に基づく推薦を行うことが考えられる。しかし単語の出現回数のみでは新鮮度が高いと判断できない場合がある。

図3の例に示されているように、単語「京大」について馴染みのあるユーザにとって、図のツイート全て、ひいては「京大」を含むツイート全てについて見慣れた情報であるとは限らない。

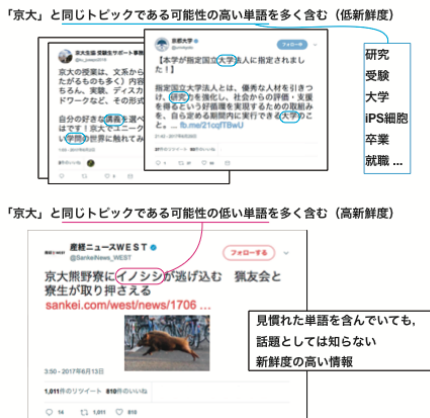


図3 トピックレベルで新鮮度の低いツイート、高いツイートの例

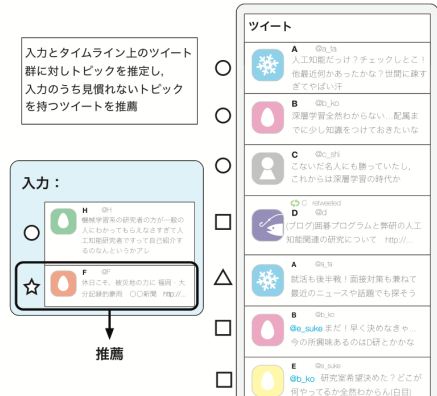


図4 提案手法によるツイート推薦の例

見慣れた単語を含むツイートであってもその背後にある話題として「ありがちな話題」「ありがちでない話題」の2パターンが考えられ、この2つを単語頻度ベースの計算で分離することは困難である。これらを分離し、後者のような場合をも新鮮度が高く推薦すべきと判断できるような計算方法を、本研究ではトピックモデルから得られる推定量を用いて定義する。

アイデアを図4に示す。ツイートのトピックを推定できれば、ユーザのタイムライン上のツイート群と推薦候補ツイート群に対しトピックを推定することで、見慣れないトピックを持つツイートをユーザにとって新鮮度の高いツイートとして推薦することができる。各単語から一様に生起確率の高いトピックを選び、それをツイートのトピックとすることで、単語の出現頻度によらない新鮮度推定を実現する。

以上の考えに基づき、ユーザにとっての新鮮度を表すスコアの計算手法を以下に順に示す。

まずツイートのトピックを推定する。2018年2月現在、Twitterには1ツイート140文字までという文字数制限があり、ツイートの大半が短い文である。このような短文について、その背後に複数のトピックが存在するとは考え難い。このため本研究では1ツイートにただ1つのトピックが存在すると仮定する。この時あるツイート t についてそのトピック z_i は、3.2で示された $p(w | z)$ を用いて次のように推定される。

$$\arg \max_{z_i} \prod_{w \in t} p(w | z_i) \quad (3)$$

ただし、 w はツイート t を構成する各単語を示す。

次に、ユーザ u のあるトピック z_i についての新鮮度は、3.2節で示された $p(z | u)$ を用いて次のように計算される。

$$1 - p(z_i | u) \quad (4)$$

式 (3) と (4) より、トピックを考慮したユーザ u とツイート t の新鮮度 $f_{fresh}(t, u)$ の導出式を以下のように定義する。

$$f_{fresh}(t, u) = 1 - p(\arg \max_{z_i} \prod_{w \in t} p(w | z_i) | u) \quad (5)$$

つまりこの計算手法はツイートのトピックを推定し、ユーザと関連の少ないトピックを持つツイートに対して高い価値を与える。

3.4 ユーザ嗜好度

Liら[4]は論文にて複数の推薦手法を包括したユーザ・コミュニティ推薦システムを提案しており、推薦ユーザを、キーワード群を入力に $\tilde{p}(u|W)$ を計算することで与えている。ユーザ u にとってのツイート t の価値はユーザ毎のツイートの条件付生起確率 $p(t|u)$ であると捉えられるので、ベイズの定理により、ユーザ u のツイート t に対するユーザ嗜好度 $f_{fav}(t,u)$ を以下のようにして求める。

$$f_{fav}(t,u) = p(t|u) = \frac{p(W)}{p(u)} \tilde{p}(u|W) \quad (t=W) \quad (6)$$

なお、 $\tilde{p}(u|W)$ はLiらの論文において以下のようにして求められている。

$$\tilde{p}(u|W) \propto p(u) \sum_z p(z|u) \prod_{w_i} p(w_i|z) \quad (7)$$

式(6),(7)より、直感的には $f_{fav}(t,u)$ は、ユーザのツイートに対する興味の度合いを単語毎のユーザの相関をトピックを介し集計することで与えている。

4. 評価実験

4.1 目的

本論文で提案しているツイートの推薦尺度を評価するために以下の実験を行った。

- 新鮮度の評価実験：単語レベルの新鮮度計算手法(4.2節)をベースラインとし、本研究で提案するトピックを考慮した新鮮度計算手法の有効性を確認する。

- まず、クラウドソーシングを用いた実験により、提案手法であるトピックレベルでの新鮮度計算手法とベースラインである単語レベルでの新鮮度計算手法について、それぞれの推薦能力を比較する。

- 次に、ユーザアンケートを用いた実験により、異なるユーザの異なるタイムラインを用いて新鮮度計算を行い、4.3節で得られた傾向が異なる入力においても一般的に認められる内容であるか確認するとともに、新鮮度だけでなく有用度等を考慮した考察を行い、推薦されるツイートに対して更なる具体的な分析を行う。

- 新鮮度とユーザ嗜好度を用いたツイート推薦の評価実験：新鮮度、ユーザ嗜好度それぞれを考慮した情報推薦とこれらを併用した混合推薦手法について比較検討する。

4.2 ベースライン：単語レベルでの新鮮度

提案手法であるトピックを考慮した新鮮度計算手法と比較するため、ツイートの類似度を用いて単語レベルでツイートの新鮮度を計算する方法をベースラインとする。

ツイート t について単語頻度 (term frequency,tf) 値を成分とするベクトル $v(t)$ を考えると、ツイート t_1 と t_2 の類似度 $sim(t_1,t_2)$ は以下のように計算される。

$$sim(t_1,t_2) = \frac{v(t_1) \cdot v(t_2)}{\|v(t_1)\| \|v(t_2)\|} \quad (8)$$

あるツイート t について、ユーザ u のタイムライン上に存在

するツイート群中の各ツイート t_i ($i=1,2,\dots,N$)との類似度を求めた際に類似度が低いものが多い時、このツイートはユーザにとって見慣れない単語の多い新鮮度の高いツイートであると言える。出現単語レベルでの新鮮度 $f_{rf}(t,u)$ は以下のように計算される。

$$f_{rf}(t,u) = 1 - \frac{1}{N} \sum_{i=1}^N sim(t,t_i) \quad (9)$$

本研究ではこれによるツイートの推薦と提案手法であるトピックを考慮した新鮮度計算によるツイート推薦の性能を評価実験により比較し、提案手法によるツイート推薦の有用性を確かめる。

4.3 クラウドソーシングを用いた新鮮度評価実験

4.3.1 概要

ランサーズ^(注1)を利用し、被験者100名に対しアンケート実験を行った。被験者には、著者が実際に運用するtwitterアカウントによりフォローされている353人の公開アカウントが平成29年11月1日から7日にかけて投稿した全ツイートのうち、後述する手法によりサンプリングされた5000ツイートを閲覧してもらった後、当該アカウントのおすすめユーザ^(注2)100アカウントが同時期に投稿したツイート100件について、既出の内容を含むか回答してもらうことで、新鮮度に関するラベルを収集した。

同時に、モデル生成用には当該アカウント及びフォローしている公開アカウントが平成29年11月7日までに投稿した日本語ツイート532765件をTwitter REST API^(注3)を用いて収集した。これを用いてUser-Topic model (UTモデル)を生成した。これに際し、ノイズとなる情報を取り除くため以下の下処理を行った。パラメータ設定はLiらに従い $\alpha=0.01, \beta=0.01$ とした。ギブズサンプリング試行回数は300回とした。

- juman++[17]を用いて形態素解析を行い、ツイートを単語群に分割し、助詞・助動詞などのそれ自体が意味を持たない品詞は除去した。

- 日本語でない単語や記号、URLなども正規表現を用いて除去した。

- ツイートを構成する単語が3語未満の場合はこれを考慮しないこととした。これは単語群が特徴的な意味ある文章として成立するのに主語・目的語・動詞が一つずつは必要と考えられるからである。

UTモデル生成の後、新鮮度に関するラベルを付与された100ツイートに対し、モデルから推定される $p(z|u), p(w|z)$ の値などから新鮮度、ユーザ嗜好度及び混合スコアをそれぞれ計算しランキングを作成した。

4.3.2 タイムラインのサンプリング

期間中にフォロワーにより投稿されたツイートは5000件を大きく超える分量であった。被験者の記憶能力と本来のタイムライン上のツイートの多様性を考慮し、本研究では以下の方法

(注1) : <http://www.lancers.jp>

(注2) : <https://help.twitter.com/ja/using-twitter/account-suggestions>

(注3) : <https://dev.twitter.com/rest/public>

によるサンプリングを行った。各方法によりサンプリングし作成したタイムラインを用いて同条件の実験を行い、サンプリング方法によりランキングや評価値が変化するか確認した。

- (1) 投稿時間の新しいツイートに順に 5000 件収集した。
- (2) 投稿時間によらず、一様にランダムに 5000 件収集した。
- (3) 期間中に投稿された全ツイートについて、投稿時間の新しい順に忘却確率に基づくフィルタリングを行い、この結果 4712 件のツイートが収集された。

ここで、投稿時間により決定される忘却確率はエビングハウスの忘却曲線に従うものとした。このとき、 t 秒後の忘却確率 $p(t)$ は、定数 $k = 1.84, c = 1.25$ を用いて、式 10 により与えられる。

$$p(t) = \frac{k}{c \log t + k} \quad (10)$$

4.3.3 結果・考察

まず、おすすめツイート群についてトピックを考慮した新鮮度と単語レベルでの新鮮度を計算し、この値を用いてランキングを作成した。次に、被験者ごとに異なる情報新鮮度に関するラベルをもとに Precision@k, Mean Average Precision (以下 MAP), ROC 曲線の AUC 値を算出した。各評価値の被験者平均を表 1,2,3 に記す。

いずれのサンプリング手法においても、上位数件程度までの評価においては提案手法が高い評価値を、ランキング全体を考慮した評価においては単語レベルでの新鮮度計算手法が高い評価値を示した。この結果により、ランキング全体、特に中下位を含めた数十件の新鮮度計算ではベースラインが優れる一方で、提案手法はランキング上位の数件程度の推薦に用いる上では優れていることが示唆された。

新鮮度のみに基づくツイートの推薦は真新しい情報を推薦することでユーザに新たな発見を促す一方で、過剰な推薦は真新しいもののユーザにとって有用な情報の割合を低下させる可能性を孕んでいる。したがって少量かつ精度の高い推薦を行うことが重要であり、このような推薦のための新鮮度計算には提案手法を用いるべきであると言える。

表 1 サンプリング手法 (1) による実験結果

被験者平均	Precision			MAP	ROC-AUC
	@1	@5	@10	all	
提案手法	0.610	0.520	0.505	0.559	0.504
単語レベルでの新鮮度計算手法	0.560	0.542	0.503	0.612	0.510

表 2 サンプリング手法 (2) による実験結果

被験者平均	Precision			MAP	ROC-AUC
	@1	@5	@10	all	
提案手法	0.500	0.468	0.430	0.489	0.468
単語レベルでの新鮮度計算手法	0.480	0.428	0.440	0.506	0.497

表 3 サンプリング手法 (3) による実験結果

被験者平均	Precision			MAP	ROC-AUC
	@1	@5	@10	all	
提案手法	0.580	0.540	0.487	0.565	0.508
単語レベルでの新鮮度計算手法	0.570	0.526	0.521	0.579	0.553

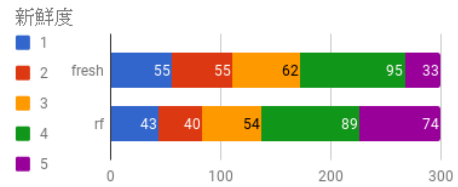


図 5 2 手法の新鮮度に関するアンケート結果

4.4 ユーザアンケートを用いた新鮮度評価実験

4.4.1 概要

平成 29 年 6 月 16 日から 30 日にかけて、日常的に Twitter を利用している被験者 6 名に対しアンケート実験を行った。モデル生成用に被験者のアカウント及びフォローしているアカウントと、それらの投稿したツイートをアカウント毎に 2000 ツイートずつ、Twitter REST API を用いて収集した。これを用いて、被験者ごとに User-Topic model (UT モデル) を生成した。これに際する下処理やハイパーパラメータは 4.3 節の実験に準拠した。

UT モデル生成の後、各被験者のアカウントに Twitter 社から提供されるおすすめユーザ 100 アカウントから発信される合計 20000 ツイートに対し、モデルから推定される $p(z|u), p(w|z)$ の値などから新鮮度、ユーザ嗜好度及び単語ベースの新鮮度等をそれぞれ計算し、それぞれについて高い値を示したツイートを 50 個ずつ抽出し、これについて有用度などを問うアンケートを被験者に対し実施した。情報推薦という観点から、ここで推薦するツイートは 3 語以上のもの・他者へのリプライでないものとした。アンケートの体裁・内容としては、3 つの質問はそれぞれ 1 ツイートの持つ被験者にとって「自分の見たい内容を含んでいるか (有用度)」、「最近の自分の TL には流れない話題であるか (新鮮度)」、「自分の身の回りで最近よく聞く、目にする内容を含んでいるか (流行度)」(以降、図表にてそれぞれ v, f, p と略記する) を問う内容である。「このツイートはあなたにとって有用度・新鮮度の高いものであるか」と問い、1「全くそう思わない」から 5「とてもそう思う」までの 5 段階でもっともふさわしいと思う評価値を回答してもらった。

4.4.2 結果・考察

図 5 は、新鮮度に関して各手法により推薦されたツイート群が獲得した回答である。なお、トピックを考慮した新鮮度による推薦を fresh, 単語レベルでの新鮮度による推薦を rf と略記する。

図 5 より、十分多いツイート群から多数のツイートを新鮮度に基づき推薦する場合、単語レベルでの新鮮度による推薦の方がより精度が高いことがわかった。これは 4.3.3 節の実験結果とも同じ傾向であり、異なるユーザ、タイムラインにおいても同様の傾向が示唆されたと言える。

4.4.3 新鮮度に基づく推薦手法の有用度を考慮した比較

次に、新鮮度に加え有用度を考慮し、提案手法と単語レベルでの新鮮度による推薦手法を比較する。

有用度、新鮮度についてのそれぞれ評価値を集計したものが表 4 である。これによると、推薦したツイートに関して有用度の高いものの個数・内訳には大きな差はなく、差があるのは推

表4 新鮮度計算手法の比較

	fresh	rf	fresh - rf
$v > 3, f > 3$	59	59	0
$v > 3, f < 3$	34	25	9
$v > 3, f = 3$	20	20	0
$v < 3, f > 3$	54	81	-27
$v < 3, f < 3$	60	49	11
$v < 3, f = 3$	29	25	4
$v = 3, f > 3$	15	23	-8
$v = 3, f < 3$	16	9	7
$v = 3, f = 3$	13	9	4
合計	300	300	0

薦したものの有用度が低いと判定されたツイートのうち、新鮮度が高いものであることがわかる。つまり、トピックを考慮した新鮮度の計算手法は従来の単語ベースの手法と比べて、ユーザにとって読む価値の高い、新しい情報を含むツイートをより高い精度で推薦できる。

単語とトピック、ユーザにとっての新鮮度の大小という区分を任意のツイートにあてはめると、任意のツイートはユーザにとって、

- トピックレベルでも単語レベルでも新鮮度が高いと計算される
- トピックレベルでは新鮮度が高いが単語レベルでは新鮮度が低いと計算される
- トピックレベルでは新鮮度が低い単語レベルでは新鮮度が高いと計算される
- トピックレベルでも単語レベルでも新鮮度が低いと計算される

のいずれかであると分類する事ができる。

トピックレベルで新鮮度が高いとは、同じトピックに属する可能性の低い単語群が1ツイート中に含まれていることであり、単語レベルで新鮮度が高いとは、単に出現頻度の低い単語がツイートの含まれていることである。これに基づき先の要件を言い換えると、以下ようになる。

- (1) 見慣れない単語を含み、同じトピックを持つ確率も低い
- (2) 単語自体の出現頻度にはよらないものの、同じトピックを持つ確率は低い
- (3) 同じトピックを持つ確率にはよらないものの、単語自体の出現頻度は低い
- (4) 単語自体の出現頻度も同じトピックを持つ確率も比較的高い

(1)と(2)はトピックレベルで新鮮度が高く、(1)と(3)は単語レベルで新鮮度が高いと言える。したがって、有用度が低く新鮮度が高いツイート群の個数差は、(2)と(3)に基づく推薦の性能差に由来すると考えられる。

各手法において有用度が低く新鮮度が高いと判定されたツイート群について文章内容を調べたところ、単語レベルでの新鮮度に基づく推薦ツイート群について、広報・宣伝ツイートの割合が高くなっていることがわかった。新たに宣伝する商品名

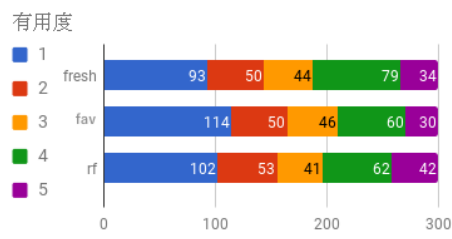


図6 手法毎の有用度に関するアンケート結果

や作品名といった固有名詞はそれまでの出現頻度は低いと考えられるため、このようなツイートは単語頻度で見ると新鮮度が高く、一方でトピックレベル新鮮度を計算することを考えると、これはツイートがいくつか未知の単語を含んでいたとしてもそれ以外の単語から推定したトピックのみに依存するので新鮮度が高いとみなされるとは限らない。

また、各手法で推薦されたツイート群のうち有用度の高かったものに対しても同様の傾向があると考えられるが、これに該当するツイート数には手法間の差はなかった。このため、それぞれの手法でしか推薦できないツイートが同程度存在し、そのようなツイート群の傾向として、単語レベルでの新鮮度に基づく推薦ツイート群については広報・宣伝ツイートが多く、トピックを考慮した新鮮度に基づく推薦ツイート群については3.3の内容から単語自体は見覚えがあるものの話題としては未知なツイートが多いと推定される。

以上の観点から、先の表から定量的に判断すれば、新鮮度が高いツイート群において、有用度の再現率としてはトピックを考慮した新鮮度に基づく推薦手法は単語レベルでの新鮮度に基づく推薦手法より優れているものの、それぞれの手法でしか推薦できないツイートが一定数存在するため、改善の余地がある。

4.5 新鮮度とユーザ嗜好度を用いたツイート推薦の評価実験

4.5.1 概要

4.4.1節に記載の通り、日常的にTwitterを利用している被験者6名について、被験者毎に新鮮度、ユーザ嗜好度及び単語ベースの新鮮度等による情報推薦を行い、推薦されたツイートについて有用度などを問うアンケートを実施した。本小節ではアンケートの回答結果から提案手法、ユーザ嗜好度及び単語ベースの新鮮度計算手法による推薦を比較した後、新鮮度とユーザ嗜好度を同時に考慮した混合推薦手法について検討する。

実験の詳細な実施概要に関しては4.4.1節を参照されたい。

4.5.2 ユーザ嗜好度に基づく推薦手法

図6は、有用度に関して、提案手法、単語レベルでの新鮮度を用いた推薦手法、ユーザ嗜好度を用いた推薦手法により推薦されたツイート群が獲得した回答である。なお、トピックを考慮した新鮮度による推薦を fresh、単語レベルでの新鮮度による推薦を rf、ユーザ嗜好度による推薦を fav と略記する。

図6によると推薦されたツイートの有用度はいずれの尺度においても同じような結果を示し、それぞれの差は少ないということがわかる。

表5 被験者毎の手法別での有用度評価値における ROC-AUC 値

被験者	fresh	fav	fresh-fav
1	0.37960	0.49621	0.50280
2	0.61818	0.48810	0.45018
3	0.35511	0.46853	0.57106
4	0.54135	0.58421	0.48546
5	0.63889	0.63043	0.45974
6	0.59568	0.37255	0.55481
平均	0.52147	0.50667	0.50401

表6 ツイートの手法別評価値分布

	fresh	fav	rf	合計
f > 3, p > 3	18	10	23	67
f > 3, p < 3	89	67	125	377
f > 3, p=3	21	6	15	53
f < 3, p > 3	59	110	51	301
f < 3, p < 3	28	38	22	121
f < 3, p=3	23	17	10	64
f=3, p > 3	7	4	7	24
f=3, p < 3	32	24	28	106
f=3, p=3	23	24	19	87
合計	300	300	300	1200

また、図7はユーザ嗜好度に基づく推薦手法の有用度評価値における ROC 曲線である。

ユーザ嗜好度による推薦の ROC 曲線の AUC 値は 0.5 程度であり、改善の余地があると言える。

4.5.3 混合手法

新鮮度のみに基づくツイートの推薦は多様な情報の収集を支援できる一方で、ユーザにとって有用な情報の割合を低下させるリスクを孕んでいる。また、ユーザの嗜好のみに基づくツイートの推薦はユーザにとって有用な情報を推薦できる可能性は高い一方で、すでに持っている情報とほとんど同じものしか得られない可能性がある。本研究ではデータ統合アルゴリズム CombMNZ を用いてこの2つの推薦手法を統合し、それぞれの推薦手法の抱える問題を同時に解決することを検討する。

CombMNZ は Fox ら [18] により提案されて以来多くの研究で利用されてきた確立されたデータ混合手法の一つで、多くの論文により他手法との比較検証もなされている [19], [20]。

新鮮度とユーザ嗜好度についての CombMNZ による混合手法を定義し、先の実験により得られた有用度・新鮮度についての評価値を持つツイート 1200 件についてランキングを再度作成した。表5は、ランキングの有用度に関する ROC 曲線の AUC 値であり、fresh, fav はそれぞれ被験者合計における混合前の手法、fresh-fav は混合手法によるものである。これによると、有用度に関して既存手法に比べ精度が向上しているとは言い難い。

理由として、新鮮度とユーザ嗜好度は互いにトレードオフの関係にあることが考えられる。表6に実験により評価値の得られた 1200 件のツイートについての評価値の分布を示す。

これによると、提案手法やベースラインは普段あまり見かけない内容を含むツイートを、ユーザ嗜好度による推薦は特に最近目にする内容を含むツイートを多く推薦する傾向にあり、そ

れぞれの推薦手法は適切に真新しい情報や興味のある情報をよく推薦できていると言える。しかし一方で、評価値の一方が高い時は、他方が低くなる傾向が強いことがわかる。

CombMNZ はランキング毎の非ゼロ要素数を重みとした重み付き和を混合評価値とするので、このようにそれぞれにおける評価値を足しあわせると、混合前のいずれの手法においても評価値が低いにもかかわらず混合手法において評価値が高いようなツイートが存在する可能性があり、このようなツイートが推薦され結果としてランキングのノイズとなっていると考えられる。これを解決するには、各尺度で総合的に評価の高いツイートではなく1つの尺度において評価が特に高いようなツイートを推薦することが考えられ、具体的には各手法の評価最大値を用いた推薦などが考えられる。

5. おわりに

本研究では情報の真新しさを表す尺度である新鮮度と、これを用いたツイートの推薦手法を提案した。複数の実験を行い、

新鮮度に基づく推薦に際して単語レベルで新鮮度を計算する手法に比べてトピックを考慮して新鮮度を計算する手法が、少量かつ高精度な推薦が必要な場合や、有用度を考慮すべき場合に優れていることが示された。また、ユーザ嗜好度を用いた推薦を提案し、提案手法と比較した上でそれぞれが想定通りに真新しい情報や興味のある情報をよく推薦する傾向にあることを確認した。更に、新鮮度・ユーザ嗜好度を同時に考慮した混合推薦手法を提案し、新鮮度のみによる推薦、ユーザ嗜好度のみによる推薦との性能比較を行ったが、有用度に関して推薦手法の混合によって精度が向上するとは言い難い結果となった。

今後の課題としては、推薦手法の精度向上を見据えた採用するトピックモデルを再検討、計算手法の見直しなどが挙げられる。

6. 謝 辞

本研究の一部は科研費 (16K12532) と総務省 SCOPE(172307001) による。

文 献

- [1] 総務省. 第2部第2節 ソーシャルメディアの普及がもたらす変化. 平成27年版 情報通信白書, pp. 199–214, 2015.
- [2] 森田真樹子. 『情報源に関する意識・実態調査報告書』. Technical report, 経済広報センター, 2013.
- [3] Su Mon Kywe, Ee-Peng Lim, and Feida Zhu. A survey of recommender systems in twitter. In *SocInfo*, pp. 420–433, 2012.
- [4] Lei Li, Wei Peng, Saurabh Kataria, Tong Sun, and Tao Li. Recommending users and communities in social media. *TKDD*, Vol. 10, No. 2, pp. 17:1–17:27, 2015.
- [5] Jianshu Weng, Ee-Peng Lim, Jing Jiang, and Qi He. Twitterrank: finding topic-sensitive influential twitterers. In *WSDM 2010*, pp. 261–270, 2010.
- [6] Ibrahim Uysal and W. Bruce Croft. User oriented tweet ranking: a filtering approach to microblogs. In *CIKM 2011*, pp. 2261–2264, 2011.
- [7] Min-Chul Yang and Hae-Chang Rim. Identifying interesting twitter contents using topical analysis. *Expert Syst. Appl.*, Vol. 41, No. 9, pp. 4330–4336, 2014.

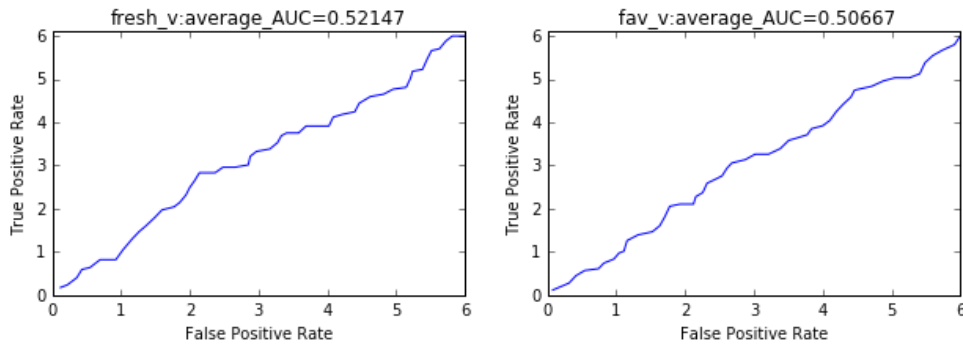


図7 被験者合計での手法毎の有用度評価値における ROC 曲線

- [8] Daniel D Lee and H Sebastian Seung. Algorithms for non-negative matrix factorization. In *Advances in neural information processing systems*, pp. 556–562, 2001.
- [9] Shipeng Yu, Kai Yu, and Volker Tresp. Soft clustering on graphs. In *NIPS*, pp. 1553–1560, 2005.
- [10] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. In *WWW*, pp. 161–172, 1998.
- [11] Denis Kotkov, Shuaiqiang Wang, and Jari Veijalainen. A survey of serendipity in recommender systems. *Knowledge-Based Systems*, Vol. 111, pp. 180–192, 2016.
- [12] Laurence AF Park and Simeon Simoff. Second order probabilistic models for within-document novelty detection in academic articles. In *SIGIR*, pp. 1103–1106, 2014.
- [13] Charles LA Clarke, Maheedhar Kolla, Gordon V Cormack, Olga Vechtomova, Azin Ashkan, Stefan Büttcher, and Ian MacKinnon. Novelty and diversity in information retrieval evaluation. In *SIGIR*, pp. 659–666, 2008.
- [14] Mohammad Al Hasan, W Scott Spangler, Thomas Griffin, and Alfredo Alba. Coa: Finding novel patents through text analysis. In *SIGKDD*, pp. 1175–1184, 2009.
- [15] Qiang Ma, Shinya Miyazaki, and Katsumi Tanaka. Webscan: Discovering and notifying important changes of web sites. In *DEXA*, pp. 587–598, 2001.
- [16] 馬強, 角谷和俊, 田中克己ほか. 放送型情報配信システムのための時系列性を考慮した情報フィルタリング. 情報処理学会論文誌データベース (TOD), Vol. 41, No. SIG06, pp. 46–57, 2000.
- [17] 京都大学大学院 情報学研究科黒橋・河原研究室. JUMAN++. <http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN++>.
- [18] Edward A. Fox and Joseph A. Shaw. Combination of multiple searches. *NIST SPECIAL PUBLICATION SP*, pp. 243–253, 1993.
- [19] Joon Ho Lee. Analyses of multiple evidence combination. *SIGIR Forum*, Vol. 31, No. SI, pp. 267–276, July 1997.
- [20] Daqing He and Dan Wu. Toward a robust data fusion for document retrieval. In *NLP-KE'08*, pp. 1–8, 2008.