

話題の新鮮度を考慮したマイクロブログ推薦手法の提案

Recommendation of Microblogs by Consideration of Topic Freshness

横山 慎[†] 馬 強[†]
Makoto Yokoyama Qiang Ma

1. はじめに

現在, Twitter をはじめとした SNS が普及しつつある [1] が, その利用目的はコミュニケーション, マーケティング, 情報収集などを始め多岐にわたる. 人々は自らの利用目的に応じてそれに適したサービスを選択し, 利用している. テレビや新聞といった既存メディアに加え, インターネットや SNS を利用した情報収集が増加しつつあり, 特に 29 歳以下の若者の間では情報収集手段としてソーシャルメディアが新聞の利用率を上回るとい調査もある [2].

一方, SNS を利用してニュースを閲覧する場合, 体系的にニュースを整理して表示する他の利用したニュースサービス (Yahoo!ニュースや各種ニュースアプリなど) と異なり, 情報の網羅性や不偏性が保障されず, 得られる情報に過不足がある場合が多い.

Twitter を対象とする推薦手法は数多く提案されている [3, 4, 5]. 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. これらは多くの場合ユーザの興味に基づく推薦であり, 普段からタイムラインを介して得られる情報と似通ったものしか得られず, 多様な情報入手したいというニーズに十分に答えられない場合がある.

本研究ではユーザにとっての情報の真新しさを示す尺度として新鮮度を提案し, それを用いたツイートの推薦を試みる. さらにその応用として, ツイートの話題性やユーザの興味を考慮した多様な情報を閲覧できる Twitter クライアントを提案する. 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.

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



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

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興味ある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, 新鮮度に加えて, ツイートの話題性を表す流行度やユーザの興味との関連性を表す興味度といった尺度を利用した Twitter クライアントを提案する.

2. 関連研究

2.1 Twitter における情報推薦

Li ら [3] はフォロ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『FRec』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 [6, 7].

Yang ら [4] も Twitter においてトピックモデルを構築してユーザの興味対象を推定する手法を提案している. トピックが時間とともに変遷していくことを仮定し, 単

[†] 京都大学, Kyoto University

位時間ごとに異なるトピック分布を与えるトピックモデルを提案している。

さらに, Uysalら [5] 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 情報の新鮮度

Kotkovら [8] 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。

Maら [9] 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・流行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。

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とツイート群を入力に, ユーザにとっての新鮮度を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, 推薦を行う。この計算のためにLiら [3] 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る。

3.2 単語・トピック分布の推定

Liら [3] 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。

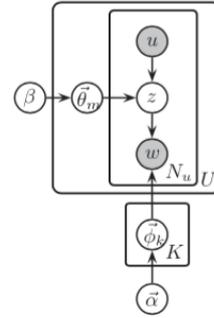


図 2: User-Topic Model

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

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

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^- 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_u^- はユーザ文書 u の単語数である。

UTモデルの詳細は文献 [3] を参照されたい。

3.3 新鮮度計算手法

本節ではTwitterにおけるトピックレベルでの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の例に沿って説明すると, 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。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」「ありがちな話題」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

「京大」と同じトピックである可能性の高い単語を多く含む (低新鮮度)



「京大」と同じトピックである可能性の低い単語を多く含む (高新鮮度)

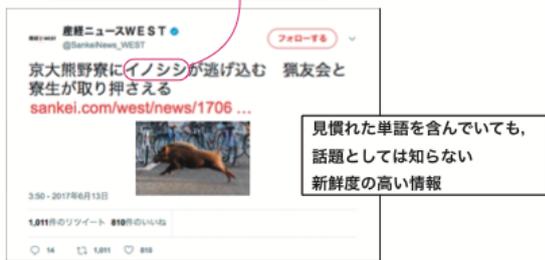


図 3: トピックレベルで新鮮度の低いツイート, 高いツイートの例

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気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。2017年7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、ツイートのお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、トピックレベルでのユーザ 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。

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

入力とタイムライン上のツイート群に対しトピックを推定し、入力のうち見慣れないトピックを持つツイートを推薦



図 4: 提案手法によるツイート推薦の例

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。

4. 応用：ツイートの推薦

4.1 概要

本研究では新鮮度に基づくツイート推薦手法の提案に加え、より広範な情報を収集できるような情報収集システムとして、通常のタイムライン上のツイート群とともに新鮮度レベルでの推薦により得られたツイート群を同時に閲覧できるような Twitter クライアントを提案する。

このような Twitter クライアントを実現するにあたり、より広範でかつ価値の高い情報を得るために新鮮度 (freshness) を用いた推薦手法は有益であると考えられるが、図1のB,Cのような情報のニーズの存在を考えると、ユーザの嗜好 (interest) や全体的な流行 (popularity) を推定し、これに関連する内容を含む情報を推薦することもまた有用であり、有用な情報をより広範に収集するという目的のためには、Twitter クライアントはこれらにより推薦されるツイート群も含めて閲覧できるようにすべきである。一例を図5に示す。

本研究では、ツイートのユーザ嗜好との関連性と、ツイートの話題性を示す流行度をそれぞれ興味度と流行度として推定する。

4.2 興味度

Liら[3]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、提案手法のひとつにキーワード群を入力に推薦ユーザを出力するアルゴリズムがあり、これは $\tilde{p}(u|W)$ を計算することで実現される。ユーザ u にとってのツイート t の価値はユー



図 5: Twitter クライアントの例

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 に対する興味度 $f_{UT}(t, u)$ を以下のようにして求める。

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

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 $f_{UT}(t, u)$ 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。

4.3 流行度

Twitter 全体で話題になっている内容を発信しているツイートは流行度が高く推薦するべきである。そのためと判断するための計算手法を提案する。

直感的には、一定期間内の任意の多数のツイートとの類似度の高いようなツイートは、期間内で話題となっていた内容を含むと考えられる。本研究ではこのアイデアに基づき単語レベルでのツイートの流行度を計算する。

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、

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

あるツイート t について、本研究では単語レベルでのツイートの流行度 $f_{rp}(t, u)$ を、ごく最近投稿されたツ

weet群中の各ツイート t_i ($i = 1, 2, \dots, N$) との類似度の平均をとることで与える。

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

本研究では単語の tf-idf 値の代わりにストップワードを省いた tf 値のみから文書をベクトル化することとした。

5. 評価実験

5.1 目的

本論文で提案しているツイートの三つ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 (5.2 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。
- 興味度と流行度の評価実験：Twitter クライアントでツイートを推薦するための尺度、興味度と流行度の有効性を確認する。

5.2 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、4.3 で定義したツイートの類似度 (式 (8)) 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用いる。

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 $f_{rf}(t, u)$ を類似度から計算することができ、

$$f_{rf}(t, u) = 1 - \frac{1}{N} \sum_{i=1}^N sim(t, t_i) \quad (10)$$

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。

なお、単語レベルでの流行度と新鮮度是对になる概念ではあるが、類似度を計算するツイート群が異なるため、 $f_{rf}(t, u) = 1 - f_{rp}(t, u)$ のようにはならない。これは比較対象とするツイート群が異なるためであり、図9のように、新鮮度計算時は比較ツイート群にタイムライン上のツイート群 t_{rf} を、流行度計算時は比較ツイート群にフォロー内外の新しいツイート群 t_{rp} を用いる。

5.3 実験の実施概要

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++ [10] 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、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、流行度、興味度及び単語ベースの新鮮度をそれぞれ計算し、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

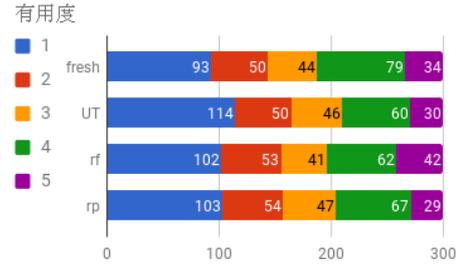


図6: 手法毎の有用度に関するアンケート結果

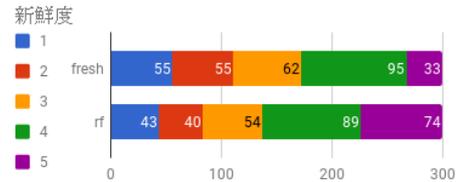


図7: 2手法の新鮮度に関するアンケート結果

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の有用度・新鮮度・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・流行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。

5.4 結果・考察

図6は、有用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。なお以降の図表にて各推薦手法について、トピックに基づく新鮮度による推薦をfresh、単語に基づく新鮮度による推薦をrf、興味度による推薦をUT、単語に基づく流行度による推薦をrpと略記する。

また、トピックと単語に基づく新鮮度による推薦により得られたツイート群について新鮮度に関するアンケート結果を図7に、興味度による推薦と単語に基づく流行度による推薦により得られたツイート群について流行度に関するアンケート結果を図8にそれぞれ示す。

図7によると新鮮度による推薦は正しく新鮮度の高いツイートを推薦できていることがわかる。同様に図8によると、興味度はユーザとそのフォロワー間のみにおける流行度とみなせるので、興味度、流行度による推薦についても正しく流行度の高いツイートを推薦できていることがわかる。一方で図6によると推薦されたツイート

*<https://dev.twitter.com/rest/public>



図 8: 2 手法の流行度に関するアンケート結果

表 1: 被験者毎の有用度の高いツイートの新鮮度流行度別の分布

| 被験者 | 新鮮度の高いツイートの有用率 | 流行度の高いツイートの有用率 |
|-----|----------------|----------------|
| 1 | 0.4508196721 | 0.5853658537 |
| 2 | 0.5517241379 | 0.3086419753 |
| 3 | 0.2335766423 | 0.2916666667 |
| 4 | 0.275 | 0.5961538462 |
| 5 | 0.6923076923 | 0.8421052632 |
| 6 | 0.1923076923 | 0.3529411765 |
| 平均 | 0.3863179074 | 0.4591836735 |

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。

5.4.1 評価尺度の妥当性

図 7, 8 により、それぞれの手法は想定通りに新鮮度や流行度の高いツイートを推薦できていることが確認できた。

表 1 に、流行度、新鮮度の高いと判定されたツイートのうち、有用度も高いと判定されたツイートの割合を被験者毎に示す。表によると、この 2 値の振る舞いは被験者によって様々であり、両方とも高い・低い値である場合や片方のみが高い場合が存在する。これはツイートの価値判断基準がユーザによって様々であることを示唆している。興味度、流行度による推薦や多くの既存研究は自分や世間における流行度の高いツイートしか推薦できない手法であるため、新鮮度の高い有用なツイートを推薦する手法はそのような手法と同等に価値があると考えられる。

5.4.2 新鮮度に基づく推薦手法の比較

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 2 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

表 2: 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 |

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に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、以下のようになる。

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

1,2はトピックレベルで新鮮度が高く, 1,3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。

5.4.3 興味度, 流行度に基づく推薦手法

前述の通り, 図8によると, 興味度, 流行度による推薦についても正しく流行度の高いツイートを推薦できていることがわかる。一方で推薦された流行度の高いツイート群に関して有用度も高いとは言い難い。図9は各手法の有用度評価値におけるROC曲線である。興味度, 流行度による推薦の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。

5.4.4 混合手法

複数の推薦手法について, それぞれにより推薦されたツイート群毎に推薦タイムラインを作成する代わりに, これを複数組み合わせ合わせた混合手法により1つの推薦タイ

表 3: 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 | UT | rp | fresh-UT | fresh-UT-rp |
|-----|---------|---------|---------|----------|-------------|
| 1 | 0.37960 | 0.49621 | 0.50850 | 0.50280 | 0.50280 |
| 2 | 0.61818 | 0.48810 | 0.40000 | 0.45018 | 0.45046 |
| 3 | 0.35511 | 0.46853 | 0.50000 | 0.57106 | 0.57106 |
| 4 | 0.54135 | 0.58421 | 0.53846 | 0.48546 | 0.48527 |
| 5 | 0.63889 | 0.63043 | 0.49275 | 0.45974 | 0.45974 |
| 6 | 0.59568 | 0.37255 | 0.47436 | 0.55481 | 0.55481 |
| 平均 | 0.52147 | 0.50667 | 0.48568 | 0.50401 | 0.50402 |

表 4: ツイートの新鮮度, 流行度についての評価値分布

| | fresh | UT | rf | rp | 合計 |
|----------------|-------|-----|-----|-----|------|
| $f > 3, p > 3$ | 18 | 10 | 23 | 16 | 67 |
| $f > 3, p < 3$ | 89 | 67 | 125 | 96 | 377 |
| $f > 3, p=3$ | 21 | 6 | 15 | 11 | 53 |
| $f < 3, p > 3$ | 59 | 110 | 51 | 81 | 301 |
| $f < 3, p < 3$ | 28 | 38 | 22 | 33 | 121 |
| $f < 3, p=3$ | 23 | 17 | 10 | 14 | 64 |
| $f=3, p > 3$ | 7 | 4 | 7 | 6 | 24 |
| $f=3, p < 3$ | 32 | 24 | 28 | 22 | 106 |
| $f=3, p=3$ | 23 | 24 | 19 | 21 | 87 |
| 合計 | 300 | 300 | 300 | 300 | 1200 |

ムラインを作成することができれば, Twitter クライアントにより表示されるすべてのツイート群における有用度の高いツイートの割合がより高まることが期待される。本研究ではデータ統合アルゴリズム CombMNZ を用いてこれらを統合することを検討する。CombMNZ は Foxら [11] 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 [12, 13]。

新鮮度と興味度, 新鮮度と興味度と流行度についての 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・流行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3は, ランキングの有用度に関する ROC 曲線の AUC 値であり, それぞれ被験者合計における混合前の手法, 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, 新鮮度と興味度・流行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4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, 新鮮度・流行度の評価値は一方が高く, 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。このようにそれぞ

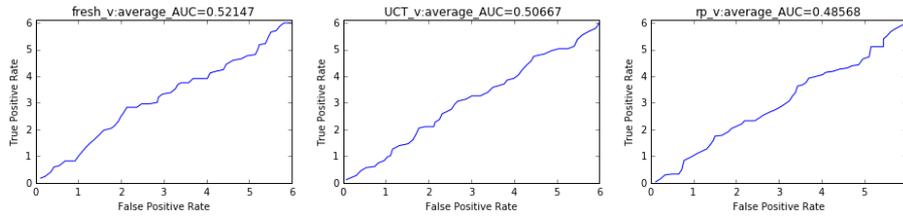


図 9: 被験者合計での手法毎の有用度評価値における ROC 曲線

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の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、各尺度で総合的に価値の高いツイートではなく 1つの尺度において価値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。

6. おわりに

本研究では新鮮度に基づきツイートを推薦する手法と、これを用いたより網羅性の高い情報収集に役立つ Twitter クライアントを提案している。

日常的に Twitter を利用する被験者 6 人に対し各ツイート推薦手法について実証実験を行い、新鮮度に基づく情報推薦の有用度が確認できた。また、新鮮度に基づく推薦に際し、単語レベルで新鮮度を計算する手法に比べてピックレベルで新鮮度を計算する手法が優れていることが示唆された。更に、Twitter クライアントの情報収集効率を高めるため新鮮度・流行度及び新鮮度・流行度・興味度における混合推薦手法を検討した。

今後は実験の結果を踏まえモデルや手法を見直しつつ、これを用いた Twitter クライアントの実装を進めていく予定である。

7. 謝辞

本研究の一部は、科研費（課題番号 16K12532）による。

参考文献

- [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] 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.
- [4] 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.
- [5] Ibrahim Uysal and W. Bruce Croft. User oriented tweet ranking: a filtering approach to microblogs. In *CIKM 2011*, pp. 2261–2264, 2011.
- [6] Daniel D Lee and H Sebastian Seung. Algorithms for non-negative matrix factorization. In *Advances in neural information processing systems*, pp. 556–562, 2001.
- [7] Shipeng Yu, Kai Yu, and Volker Tresp. Soft clustering on graphs. In *NIPS*, pp. 1553–1560, 2005.
- [8] Denis Kotkov, Shuaiqiang Wang, and Jari Veijalainen. A survey of serendipity in recommender systems. *Knowledge-Based Systems*, Vol. 111, pp. 180–192, 2016.
- [9] Qiang Ma, Shinya Miyazaki, and Katsumi Tanaka. Webscan: Discovering and notifying important changes of web sites. In *DEXA*, pp. 587–598, 2001.
- [10] 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++>.
- [11] Edward A. Fox and Joseph A. Shaw. Combination of multiple searches. *NIST SPECIAL PUBLICATION SP*, pp. 243–253, 1993.
- [12] Joon Ho Lee. Analyses of multiple evidence combination. *SIGIR Forum*, Vol. 31, No. SI, pp. 267–276, July 1997.
- [13] Daqing He and Dan Wu. Toward a robust data fusion for document retrieval. In *NLP-KE'08*, pp. 1–8, 2008.