

# 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

鈴木 佑樹<sup>†</sup> 寺田 実<sup>†</sup>

<sup>†</sup> 電気通信大学大学院 情報理工学研究科 情報・通信工学専攻 コンピュータサイエンスコース

〒 182-8585 東京都調布市調布ヶ丘 1-5-1

E-mail: <sup>†</sup>s1531055@edu.cc.uec.ac.jp, <sup>††</sup>terada.minoru@uec.ac.jp

あらまし Twitter 上に流れる情報は年々増え続けており、タイムライン上のトレンドは、日々新しいものによって変わっていく。フォローユーザ（フォロワー）を増やせばより多くの情報を得ることができるものの、そこで得られるすべてが有益な情報とは言い難い。大勢のユーザに拡散（リツイート）されるツイートは、一般的に他の人に伝えたい内容であることが殆どだが、それが自身にとって必ずしも有益なものであるとは限らない。ユーザ関連度の高いフォロワーのリツイートを抽出できれば、有益な情報を得ることが容易になるのではないかと考えた。本報告では、ユーザの興味を重視した推薦機構を自作し、有益なツイートを得られることができるかを検証する。実験では、推薦インターフェースを実際に使ってもらうことで、実際にいいね登録やリツイートをしたいツイートが得られたかを検証した。

キーワード Twitter, 情報推薦, フォロワー

## 1. 背景

Twitter は非常に多くのユーザに使われている SNS である。Twitter は「ツイート」という 140 文字以内のテキストメッセージの投稿をし、身の回り友人やネット世界での知り合い等、自分と繋がっているユーザと情報交換を行うことができるサービスである。情報収集の面においては、情報の速報性・多様性に優れ、社会現象や生活に役に立つ少ネタなど、枚挙にいとまがない。最近では、企業が自社のサービスを宣伝するために使ったり、芸能人が近況を発信したりなどと個人的な用途以外にも広く使われるようになった。

特定のツイートに対して行える返信以外のアクションとして、いいね（2015 年 11 月以前はお気に入り）とリツイートというものがある。いいねは、特定のツイートを自分のいいねリストに加え、あとでそれをまとめて見返すことができる。リツイートは、そのツイートを自分のタイムラインに表示し、自分のフォロワーに共有することができる。一般的にどちらも有益な情報に対して行われるアクションであるが、いいねは自分が後で見返す為のもの、リツイートは自分のフォロワーに共有したい知識・お得情報に対してなされる傾向がある。2016 年 6 月より、自分のツイートに対してもリツイートを行えるようになった。

フォローするユーザが増えると表示される発言の数も比例的に増え、有益な情報を見落としてしまう可能性が高まってしまふ。ユーザをグループ分けするための機能も提供されているが、それらはユーザが手動で行う必要があり、フォロワーが多い場合には面倒である。また、近年ではツイートに添付できるメディア URL や引用先のツイートなどが文字数制限にカウントされなくなり、1 ツイートあたりの情報の量が増えたことも、有益な情報を見落としてしまうことに拍車をかける。

また関連研究内では、ツイート自体のテキスト情報を解析することで、ツイートや発言ユーザをユーザに推薦するものもある。しかし、Twitter ではブログサイトの URL や添付された動

画・メディアの URL が有益だと思われ、多く拡散されるリツイートも存在する。そのようなリツイートは URL を除いた文字情報が極端に少ないこともあり、既存のテキストを解析する手法ではツイートの解析が困難である。

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興味・関心の高い内容を、拡散者の情報を基準に推薦を行うことである。フォローしているユーザグループが似ているユーザを自分と興味の近いユーザと考え、似た興味を持っているユーザほど、リツイートの内容も自分の興味があるものではないだろうか考えた。そこで本研究では Twitter のデータを利用し、ユーザ同士の関連度をフォロワー一致率という独自の尺度で捉え、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。

## 2. 関連研究

Twitter のデータを利用した推薦の研究は過去にも幾つか行われている。

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、ツイートとユーザの関心度との関連性を測定することによって、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モデルの提案を行った。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、その興味の基準が Chunliang の研究とは異なる。

渡部ら [2] による研究では、ユーザ間でのお気に入りの登録、被お気に入りの関係を使い、専用の推薦システムである「Twitter User Recommender」を使うことによって、特定の分野に関するユーザの推薦を行った。単純なフォロー関係を使うだけでは難しい、局所的にユーザ間評価の高いユーザを見つけることに成功している。本研究ではユーザではなく、フォロワー内のリツイートを推薦を行い、またその判断材料に使うのはお気に入りではなく、興味の一一致率である。

Huberman ら [3] は、リプライの送信先を用いたアクティブなユーザ同士のネットワークを抽出の研究を行った。この研究では Twitter から収集したユーザのデータを使い、特定のユーザへのリプライの数によってそのユーザが友達であるかを決定し、その友達の数がそのユーザの実際のフォロワー数、フォロワー数との比較を行った。これによって、大半のユーザはフォロワー数に対して友達の人数の割合が少ないことを発見した。この研究と本研究では、Twitter のデータを使ってユーザ同士の関係を判断するという点で共通するものの、本研究ではツイートの内容、送信先のデータを使わず、同一ツイートをリツイートしたか、どれほど同じユーザをフォローしているかという点を重視する。

近藤ら [4] 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。LDA とは潜在的ディリクレ配分法を意味する Latent Dirichlet Allocation の略であり、単純な単語の頻出回数ではなく、その単語が持つ潜在的な意味からトピックモデルを推定し、クラスタリングを行う手法である。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。本研究では、この研究と同様に推薦手法に協調フィルタリングを用いるという点で共通する。この研究では興味のあるトピックの推薦を行っていたが、本研究ではさらに具体的であるツイートそのものについて推薦を行うことができないかを検証する。

Twitter から有益な情報を抽出する研究としては、大原ら [5] による、情報の意外性に着目した手法がある。この研究では、通常のタイムラインではなかなか見ることのできない意外情報を有益な情報とした。ツイートの位置情報や投稿時間などの属性に着目したアトリビュート意外情報、ツイートの中身を見て判断するコンテンツ意外情報の 2 つの観点から意外情報を定義し、これらについてそのツイートが非日常的なものであるかを検証をした。本研究でもこの研究と同様に有益な情報を Twitter から抽出することを試みるが、ツイートそのものの情報でなく、そのツイートをリツイートしたユーザとの関係を重視して推薦ができないかを検証する。



図 1 公式から推薦される「おすすめユーザー」

ユーザへの推薦システムとして、Web 版 Twitter ではホーム画面のタイムラインの横に「おすすめユーザー」という項目が

表示される。これは自分のアカウントのフォロー情報やいいねの傾向等に基づいて、まだフォローしていないユーザを推薦するものである (図 1)。

### 3. 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。

#### 3.1 推薦

システムに推薦されるユーザ (以下、被推薦ユーザと呼ぶ) のフォロワーの最新のツイートを Twitter API にて取得し、その中からリツイートのみを抽出する。各リツイートに対して、それをリツイートしたフォロワーの被推薦ユーザから見たフォロワー一致率の値をそのまま推薦の際のスコアとした。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。

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た。評価実験では、これを考慮した場合、しなかった場合の二通りについて検証し、考慮する場合は、元の推薦スコアにサブスコアをかけ合わせたものを最終的な推薦スコアとした。

##### 3.1.1 フォロワー一致率

著者の先行研究の評価実験では、フォロワーの一致率とそのユーザとの関連度についての相関を示した。今回、そのユーザの興味的一致具合を求める尺度に、フォロワー一致率を利用できないかを検証する。ユーザ A から見たユーザ B とのフォロワー一致率は以下の式で求められる。

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

ここで  $f_A \cup f_B$  はユーザ A、ユーザ B のフォロワーの和集合、 $f_A \cap f_B$  はユーザ A、B のフォロワーの積集合である。

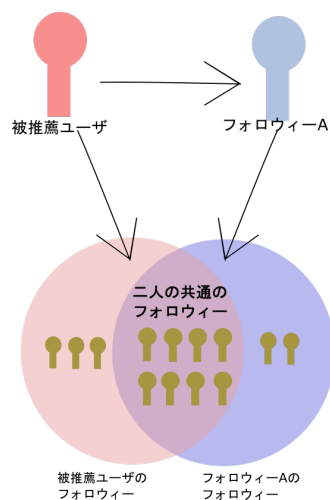


図 2 被推薦ユーザとフォロワー A のフォロワーの関係

### 3.1.2 協調フィルタリング

協調フィルタリングとは、ユーザの嗜好情報を利用して、似た嗜好を持つユーザが好むアイテムを推薦をする推薦アルゴリズムである。協調フィルタリングにはユーザ毎の嗜好の類似度を解析するユーザベースのものと、アイテム同士の類似度を解析するアイテムベースのものが存在する。本研究において、有益なリツイートが協調フィルタリングにおけるアイテムに該当するのであるが、アイテムベースの場合ツイート同士の類似度の測定が難しく、ユーザベースなら嗜好の類似度をフォロワー一致率というパラメータで測ることができるため、本研究ではユーザベースの協調フィルタリングを用いた。

### 3.1.3 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。リツイート数が少ないツイートの場合、狭いコミュニティでの内輪ネタや有名人のアカウントの日常ツイートなど有益な情報とは程遠い内容のものが多く、有益なリツイートの推薦には、不必要だと判断した。

## 4. 実装

本章では、システムの構成とその構成要素の詳細について述べる。

第3章で述べた提案手法を実現するために、推薦されるリツイートを表示するためのインターフェースを作成した。システムはWebアプリケーションであり、サーバ側のプログラムとクライアント側のプログラムがある。

### 4.1 開発環境

サーバ側には、JavaScriptのサーバ向けフレームワークであるNode.jsを使用した。使用したバージョンは0.12.0である。また本システムの開発にあたり、以下のNode.js向けパッケージを利用した。

サーバとクライアント間でのユーザ入力等によるデータのやり取りのためにSocket.IOを利用した。

本システムでは、APIによって取得したフォロワーやツイートのデータをユーザ毎に保存するために、MongoDB Inc.の開発するMongoDBを利用している。DB内のデータをスクリプトから動的に利用するために、JavaScript向けMongoDBライブラリであるMongooseを使用した。

また、本システムはWebアプリケーションであり、サーバ上でプログラムが動作しているが、それを再起動することなく、複数人が同時に利用をできなければならない。動的に生成されるセッション毎に使用するTwitterアカウントを分けるために、passportというパッケージを利用した。

## 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、またWebアプリケーションが正しく動作した上で、そのインターフェースの妥当であるかを検証するために実験を行った。

## 5.1 実験 1

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。この実験で推薦されたリツイートと被推薦ユーザの行ったリツイートの一致率が大きければ、フォロワー一致率を推薦スコアとした推薦アルゴリズムは有用であると言える。

被推薦ユーザは、以下の条件を満たすユーザを10名選出した。選出の基準は、2016年8、9月の特定のTwitterトレンドについて、何件かのリツイートをしているユーザを無作為に選んだ。その際、各被推薦ユーザの行ったリツイートのトレンド内容は被らないようにした。選出した被推薦ユーザの詳細は表1に示した。

- フォロワー数50以上、500未満(最低値68,最高値477)
- 最新ツイート200件の中に10件以上のリツイートをしている
- 最新ツイートが1週間以内のものである
- アカウントを半年以上使っていて、ツイート数が1000件以上

これらの推薦ユーザのフォロワーのフォロワー、フォロワーのリツイート情報、被推薦ユーザのリツイート情報をTwitter APIを用いて収集し、3.1.3節で触れたように、不要なデータをデータセットから取り除いた。

今回の実験では、フォロワーの行ったリツイートを被推薦ユーザが行う際、ユーザ関連度の高いフォロワーのリツイートを優先してリツイートをするか検証する為のものである。その為、被推薦ユーザの行ったリツイートのうち被推薦ユーザがリツイートを行った時刻より前に、被推薦ユーザのフォロワーが誰もリツイートしていないツイートについては、今回の実験ではデータセットに含めないようにした。非公開ユーザのフォロワーがリツイートして被推薦ユーザの目に留まったツイート、もしくはフォロワーから形成されるタイムライン以外で見かけたツイートが該当する。

10ユーザそれぞれについてサブスコアの考慮なし、あり場合の推薦を行い、その結果をそれぞれ表2,3に示した。「表中の被推薦ユーザRT数」はAPIで確認できた被推薦ユーザが実際に行ったリツイートの数、「推薦RT数」はシステムが被推薦ユーザに推薦したリツイートの数、「一致数」はその2つが一致した数をそれぞれ表している。また、これらの結果の評価指標には、適合率、再現率、F値を用いた。

結果から、サブスコアの考慮に関わらず、フォロワー数が多くフォロワー内のリツイートの数が増えるほど、推薦されるべきリツイートの数が増え、適合率が下がってしまうという結果が得られた。

これを受けて、APIで収集できたリツイートの数を推薦の際

表 1 実験 1 の被験者の内訳

被推薦	フォロワー数	収集 RT 数	被推薦ユーザ RT 数
A	68	87	22
B	131	152	8
C	140	172	18
D	170	188	27
E	177	269	15
F	219	247	18
G	254	382	17
H	288	291	26
I	340	350	20
J	477	397	24

のスコアの閾値に比例させ、適合率の安定を目指し、再び同様のデータセットを用いて推薦を行った。その結果を表 4, 5 に示した。

結果を見ると、閾値を考慮しない場合よりも適合率は全体的に向上したものの、値に関してはあまり変わらない結果となった。

表 2 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	一致数	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 3 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	一致数	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

## 5.2 実験 2

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 4 種の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。

Twitter をある程度使っている 6 人の被験者のそれぞれ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。クライアント側で表示されるリツイートを、本来の推薦スコアの

表 4 実験 1 においてサブスコア考慮なし、推薦閾値可変場合の結果

ユーザ	ユーザ RT 数	推薦 RT 数	一致数	適合率	再現率	F 値
A	22	12	5	0.42	0.23	0.29
B	8	8	3	0.38	0.38	0.38
C	18	15	3	0.2	0.17	0.18
D	27	24	10	0.42	0.37	0.39
E	15	14	4	0.29	0.27	0.28
F	18	10	2	0.2	0.11	0.14
G	17	15	5	0.33	0.29	0.31
H	26	26	7	0.27	0.27	0.27
I	20	22	5	0.23	0.25	0.24
J	24	26	7	0.27	0.29	0.28

表 5 実験 1 においてサブスコア考慮あり、推薦閾値可変場合の結果

ユーザ	ユーザ RT 数	推薦 RT 数	一致数	適合率	再現率	F 値
A	22	20	4	0.2	0.18	0.19
B	8	10	2	0.2	0.25	0.22
C	18	16	3	0.19	0.17	0.18
D	27	29	7	0.24	0.26	0.25
E	15	10	4	0.4	0.27	0.32
F	18	20	4	0.2	0.22	0.21
G	17	15	5	0.33	0.29	0.31
H	26	27	5	0.19	0.19	0.19
I	20	18	3	0.17	0.15	0.16
J	24	25	3	0.12	0.13	0.12

降順という形式ではなく、ランダムに表示し、そのリツイートの有用性を横に設置したラジオボタンで評価してもらう。被推薦ユーザには、「ツイートの発信者」「ツイートの内容」は分かるが、「そのツイートをリツイートしたフォロワーユーザ」「そのリツイートの推薦スコア」については伏せてある。これによって、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。

これを一人の被験者について 50 件のリツイートに対し評価を行う。評価内容は、以下の 4 項目から 1 つを選んでもらう形式である。

- 3, そのリツイートはいいねに加えたく、フォロワーにリツイートをしたい内容でもあった
- 2, そのリツイートはいいねに加えたい内容であった
- 1, そのリツイートはフォロワーにリツイートしたい内容であった
- 0, そのリツイートに対し、いいねもリツイートもしようとは思わなかった

これらの回答とそれぞれの有益度の対応について表 6 に示した。

今回の実験の被験者の内訳を表 7 に示した。

被推薦ユーザ毎の回答の個数について表 8 に示した。またこの結果を用いて、推薦のスコアを横軸にとり、いいねで見た有益度、リツイートで見た有益度、総合的に見た有益度の値をそれぞれ縦軸にとった時の相関係数をサブスコアの考慮なし、ありの

表 6 回答と有益度の対応

回答	いいね有益度	RT 有益度	総合有益度
3	1	1	1
2	1	0	0.5
1	0	1	0.5
0	0	0	0

表 7 実験 2 の被験者の内訳

ユーザ	フォロワー数
a	597
b	177
c	277
d	82
e	319
f	76

場合についてを求め、それぞれ表 9, 10 に示した。サブスコアを考慮した場合の結果は、そうでない場合の結果より全体的に相関係数の値が低い結果となった。

## 6. 考 察

実験 1 において、閾値を考慮しなかった場合に特に適合率が低い結果となった。適合率が低いのは、False Positive の部分が多く、システムの推薦した結果を被推薦ユーザが実際にリツイートしていなかったということである。しかし実験 1 の場合、被推薦ユーザがリツイートしていなかったとしても、「そのツイートがリツイートされた時間にタイムラインを見ておらず、

たまたまそのツイートを見ていなかった」ということも考えられ、必ずしも被推薦ユーザがそのリツイートを有益でなかったと感じたのではなかったとも考えられる。

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値が低くなっている。このような結果について、次のように考えられる。

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。今回のスコアの計算式の場合、「フォロワー一致率の低いユーザのツイートをフォロワー一致率の高いフォロワーがリツイートした場合」「フォロワー一致率の高いユーザのツイートをフォロワー一致率の低いフォロワーがリツイートした場合」これら 2 つのリツイートだと、サブスコアを考慮しなければ前者の方が優先的に推薦されるが、考慮した場合は後者の方が優先的に推薦される。今回の結果を見る限り、リツイートに関して言えば、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。

今回実験 2 において、ツイート自体の情報の量を測るべく、ツイートのメディアの数、URL の有無（メディア URL は除く）についての統計をとり、それぞれの被験者が評価したツイート群に対し、それぞれについて表 11, 12 に示した。また、それらを含まなかったツイートの個数についても回答毎に算出し、それを表 13 に示した。

表 11 実験 2 における有益度毎のメディア数

メディア数	3	2	1	0
4	5 / 13.9%	6 / 12.2%	6 / 27.3%	22 / 11.4%
3	8 / 22.2%	4 / 8.2%	2 / 9.1%	28 / 14.5%
2	6 / 16.7%	21 / 42.9%	5 / 22.7%	40 / 20.7%
1	10 / 27.8%	11 / 22.4%	3 / 13.6%	47 / 24.4%
0	7 / 19.4%	7 / 14.3%	6 / 27.3%	56 / 29.0%
合計	36 / 100.0%	49 / 100.0%	22 / 100.0%	193 / 100.0%

表 8 実験 2 のユーザの評価結果

ユーザ	3	2	1	0
a	10	10	8	22
b	3	8	3	36
c	8	8	1	33
d	2	9	0	39
e	7	2	6	35
f	6	12	4	28

表 9 実験 2 のサブスコアの考慮無しの場合の結果

ユーザ	総合有益度	いいね有益度	RT 有益度
a	0.505	0.447	0.344
b	0.325	0.307	0.177
c	0.604	0.453	0.576
d	0.449	0.384	0.229
e	0.434	0.36	0.39
f	0.472	0.296	0.417

表 12 実験 2 における有益度毎の URL の有無

URL	3	2	1	0
あり	10 / 27.8%	18 / 36.7%	16 / 72.7%	83 / 43.0%
なし	26 / 72.2%	31 / 63.3%	6 / 27.3%	110 / 57.0%
合計	36 / 100.0%	49 / 100.0%	22 / 100.0%	193 / 100.0%

表 10 実験 2 のサブスコアの考慮有りの場合の結果

ユーザ	総合有益度	いいね有益度	RT 有益度
a	0.397	0.346	0.276
b	0.259	0.21	0.184
c	0.565	0.432	0.528
d	0.225	0.259	0.002
e	0.337	0.22	0.354
f	0.383	0.18	0.409

表 13 実験 2 における有益度毎のメディア、URL がなかったツイートの個数と割合

回答	個数 / 割合
3	4 / 11.1%
2	6 / 12.2%
1	3 / 13.6%
0	46 / 23.8%

- [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] 渡部 将太, "Twitter User Recommender ~Twitter のお気に入り機能を用いたおすすめユーザ推薦システム~". DEIM Forum2012, B3-4, 2012.
- [3] Bernardo A. Huberman, "Social networks that matter: Twitter under the microscope". First Monday, 14(1-5), 2008.
- [4] 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.
- [5] 大原 啓詳, "Twitter 上のあるユーザの意外な情報抽出手法の提案". DEIM Forum2014, B6-3, 2014.

これらの結果より, 1 ツイート内の情報の量が有益度が高くなっているのが分かる. 有益なリツイートを推薦する際には, これらのパラメータを見てツイートの情報の量の多さを決定し, 推薦スコアに考慮するようにすれば, より良い推薦結果を得られると考えられる.

実験 2 で被験者に評価してもらったリツイートのうち, メディアが添付されている, もしくは何らかの URL が含まれているツイートのうち, URL を除いた文字数が極端に少ないツイートに関して, そのツイート群の統計について, それぞれ表 14, 15 に示した. 推薦スコアの上限は被験者によって異なるため, ユーザ毎に推薦スコアを正規化し, それを各有益度との相関をとった.

表 14 実験 2 におけるサブスコア考慮なしの場合の URL を除いた文字数が少ないツイートについての各有益度とスコアの相関

文字数	個数	総合有益度	いいね有益度	RT 有益度
0	14	0.603	0.371	0.487
1 ~ 10	38	0.516	0.235	0.407

表 15 実験 2 におけるサブスコア考慮ありの場合の URL を除いた文字数が少ないツイートについての各有益度とスコアの相関

文字数	個数	総合有益度	いいね有益度	RT 有益度
0	14	0.409	0.478	0.085
1 ~ 10	38	0.339	0.143	0.339

これらの結果を見るに, 従来手法では難しかった文字情報が極端に少ないツイートに対しても, ユーザ関係を見ることで推薦ができていけると言える.

## 7. まとめと展望

提案手法の推薦アルゴリズムによって, 2 つの実験によってその有用性を評価した. 結果として, 被推薦ユーザとフォロワー一致率が高いユーザのリツイートに有益な情報が含まれているということが分かり, これによって有益なリツイートの抽出が可能であることを示した. 文字情報の少ないツイートに対しても, 推薦が行えることを示した. また, 情報の量の少ないリツイートをノイズとして結果から取り除くことでより良い推薦結果を得られることを示した. しかし, 推薦結果には False Positive なものも多く見受けられ, これを改善する為の方法を模索する必要がある.

フォロワーが有益であると思った情報としてリツイートだけでなく, いいねについても推薦の候補に加えて, 推薦の幅を広げることも視野に入れたい. しかし, 候補を増やしすぎてしまうと前述のように False Positive が増えてしまう問題もあり, 情報過多のタイムラインから必要な情報のみを抽出するシステムとしては見逃せない問題である. いいね付きのリツイートとそうでないリツイートで有益度のスコア計算を変えるなどの補助的な用途に使いシステムの精度を上げることができないかを, 今後検証する予定である.