

マイクロブログにおける「惹き付ける興味」の推測による セレンディピティなユーザの推薦

徐哲林¹ 周娟² 高田秀志²

概要: マイクロブログの一つである Twitter 公式サイトにおけるお薦めユーザシステムでは、予測の正確さを重視しているため、ユーザがフォローしている人と同じジャンルのユーザを多く推薦する。しかしながら、単に正確さを重視しただけでは、推薦する範囲が狭くなるなどの問題が発生し、ユーザの満足度に悪い影響を与える可能性がある。そのため、本研究では、フォロー関係で繋がっているユーザの中で、ユーザベース協調フィルタリングに基づいて、個人を惹き付ける興味を推測することにより、セレンディピティを感じるユーザを推薦する手法を構築し、ユーザの満足度を向上させることを目指す。また、セレンディピティのあるお薦めユーザに関する既存研究と比較し、有効性を検証する。

Serendipitous User Recommendation by Inferring "Attracting Interests" in Microblogs

XU ZHELIN¹ ZHOU JUAN² TAKADA HIDEYUKI²

Abstract: In Twitter which is a microblog service, user recommendation on its official site recommends many users of the same genre as the people you follow because it emphasizes the accuracy of predictions. However, simply focusing on accuracy may cause problems such as narrowing the recommended range, which may adversely affect user satisfaction. Therefore, in this paper, we aim to improve satisfaction by a method for recommending users with serendipity feeling by inferring the attracting interests based on user-based collaborative filtering among users who are connected in the follow relationship. In addition, the effectiveness is verified by comparing with an existing method on recommending serendipitous users.

1. はじめに

近年、インターネットの発展に伴い、Web 上のデータ量は膨大になっている。それによって、多くの情報の中からユーザは自分が好きなものもしくは価値のある情報を探すのが難しくなるという問題が発生している。そのため、このような問題を解決するために、データマイニング技術を応用した推薦システムが登場している [1]。

推薦システムでは、一般的に予測の正確さ (Prediction Accuracy) という評価指標を重視している。このようなシステムによって推薦されるアイテムはユーザに対して直感的であり、しかも常に予測結果が正しいことを目指している。そのため、ユーザのプロファイル (たとえば、行動履歴

や好きなアイテムなど) と類似しているアイテムしか推薦しないという傾向がある [2]。しかしながら、推薦されるアイテムは本人がすでに知っている可能性が高いため、推薦しなくても自身が見つけられ、しかも、同じジャンルのアイテムを多く推薦すれば、推薦の範囲は狭くなると考えられる。結局、予測の正確さを重視しすぎると、ユーザの満足度に悪い影響を与える可能性がある。そのため、このような問題を起こさないために、セレンディピティという評価指標が注目されている。しかしながら、現在に至るまで、セレンディピティにある要素は様々であるため、定義はまだ統一されていないという現状である。セレンディピティに関する要素については 2.1 節で詳しく紹介するが、本研究では、Kotkov ら [3] と Ge [4] らが提案した要素の定義に基づいて、セレンディピティのあるアイテムは予期していなかった、かつ、好きなアイテムであると考えられる。このよ

¹ 立命館大学院情報理工学研究所

² 立命館大学情報理工学部

うなアイテムを推薦すれば、意外性を感じることに伴い、興味の範囲も広げることができる。さらに、JL Herlockerらは満足度に大きい影響を与えるセレンディピティという評価指標を重視することを指摘し [5]、他の既存研究も予測の正確さ以外のセレンディピティを重視する必要があることを支持している [3,6-8]。

本研究では、マイクロブログの一つである Twitter に着目する。Twitter^{*1}とは人々がネット上で情報発信、情報共有を行うサービスであり、国内外に広く普及している。Twitter 公式サイトにおけるお薦めユーザシステムは予測の正確さを重視するので、推薦するユーザは本人がフォローしているユーザと同じジャンルのユーザがほとんどであり、フォローしたいユーザ、すなわち価値のあるユーザが多くないと考えられる。それに対して、興味あるかつ意外性を感じるセレンディピティのあるお薦めユーザを発見すれば、ユーザがフォローしている人と同じジャンルではないユーザを推薦するので、お薦めユーザの範囲を広げ、意外性を感じるとともにフォローしたいユーザが多くなり、満足度を向上させると考えられる。

これまでに、Twitter においてセレンディピティのあるユーザを推薦する手法はいくつか提案されている。Scottらはユーザが発信した tweet に着目し、tweet にある興味とハッシュタグによってユーザと類似しているユーザがもつ興味、添付された url にある興味を個々に推測した上で、ハッシュタグによって推測した類似している人がもつ興味に基づいてセレンディピティのあるお薦めユーザを推測できる可能性があることを指摘している [9]。しかしながら、ハッシュタグを使うユーザは多くない、また、tweet から抽出したハッシュタグは少ない可能性があるため、ユーザと類似している人は十分に推測できないことが考えられる。一方で、本研究では、Twitter で最も特徴的な機能である tweet と retweet に着目する。これらの機能はユーザが常に利用しているので、tweet と retweet によってユーザと類似している人を推測しやすいと考えられる。また、安部らはフォロー関係で繋がっているユーザの中で、ユーザと興味が類似していない人はフォローしているので、多少興味あり、かつ、意外性を感じる可能性が高く、このようなユーザと類似している人はセレンディピティのあるユーザであると指摘している [10]。しかしながら、類似していない人に着目すれば、極めて少ない情報を発信するユーザと公式アカウントを抽出する可能性が高くなるため、これらはセレンディピティのあるユーザにならないと考えられる。

これらに基づき、本研究では、Twitter における、フォロー関係で繋がっているユーザの中で、ユーザベース協調フィルタリングに基づいて、ユーザとユーザがフォローしている人の tweet と retweet によってユーザと類似してい

る人を推測した上で、惹かれた興味と関心がありそうな興味という特徴を導入し、これらの興味から推測される「ユーザを惹き付ける興味」を多くもつ人はセレンディピティのあるユーザとして推薦する。また、予備実験によって、提案手法を用いてセレンディピティのあるユーザが推測できる可能性があり、安部らの既存研究よりセレンディピティの評価値が高いことを示す。

本論文の構成は以下に示す通りである。第 2 章では既存研究について述べる。第 3 章では Twitter におけるセレンディピティのあるお薦めユーザを推測する手法を提案する。第 4 章はセレンディピティを評価する手法を示す。第 5 章では提案手法を評価する予備実験について、セレンディピティを評価する手法を用いて、既存研究と比較し、考察を行う。最後に、第 6 章はまとめと今後について述べる。

2. 既存研究

2.1 セレンディピティ (serendipity)

現在に至るまで、様々な要素が考えられるため、セレンディピティの定義はまだ統一していない。既存研究では、セレンディピティに関する要素には主に相関性 (Relevance)、意外性 (Unexpectedness)、新規性 (Novelty) の 3 つがある。また、要素の定義は各筆者によって異なる。

2.1.1 相関性 (Relevance)

Kotkov らによれば、ユーザ自身がすでに好きになったアイテム、または関心がありそうなアイテムは相関性のあるアイテムである [3]。一方で、Su らによれば、ユーザと関係ある人から推測されるユーザが興味がありそうなアイテムは相関性のあるアイテムである [11]。

2.1.2 意外性 (Unexpectedness)

Ge らによれば、ユーザのプロフィール (興味、行動履歴) と類似していないアイテムは意外性があると指摘されている [4]。さらに、Zheng らによれば、ユーザのプロフィールと類似するかどうかを考慮した上で、流行ではないアイテムは意外性があると指摘されている [2]。

2.1.3 新規性 (Novelty)

Kapoor らによれば、ユーザが以前気になっていたが、今忘れたアイテムは新規性のあるアイテムである [12]。一方で、Kotkov らによれば、ユーザが全く知らないアイテムは新規性のあるアイテムであると提唱されている [3]。

2.2 本研究におけるセレンディピティ

本研究では、Kotkov らと Ge らが提案している相関性と意外性という要素の定義に基づいて、セレンディピティとはどれだけ魅力的か、かつ、意外性があるかという要素が加わった評価指標、すなわち、意外性を改良する評価指標であると考えられる。どれだけ魅力的かとはユーザが従来から関心があるアイテム、またはこれから関心がありそうなアイテムである。意外性があるかとはユーザが予期してい

*1 <https://twitter.com/>

なかったアイテムである。Twitter においては、ある興味に関する情報を発信する、すなわち、ある興味をもつお薦めユーザに対して本人が「関心があるかどうか」と「意外性を感じるかどうか」の両方を満足すれば、セレンディピティのあるユーザになると考えられる。

図 1 はセレンディピティのあるお薦めユーザの例である。ユーザ B は火鍋が好きなので、火鍋に関する情報を発信する。ユーザ A は常にスイーツに関する情報を見ているが、火鍋に関する情報をもたらるのは予期していなかったことなので、意外に感じるとともに好きになれば、ユーザ B はセレンディピティのあるお薦めユーザになる。



図 1 セレンディピティのあるお薦めユーザの例

2.3 ユーザベース協調フィルタリング

ユーザベース協調フィルタリングとは、ユーザと類似している人が好きなアイテムはユーザも好きになる可能性が高いと想定した上で、彼らの好きなアイテムをユーザに推薦する手法である [11]。ユーザベース協調フィルタリングによってユーザと類似している人からユーザが関心ありそうな興味を推測でき、しかも、類似している人がもつ興味は好きになる可能性が高いので、どれだけ魅力的かという要素を満足する。そのため、本研究ではユーザベース協調フィルタリングに基づいてセレンディピティのあるお薦めユーザを推測する。

3. 提案手法

本研究におけるセレンディピティの定義を満たすためには、お薦めユーザが発信する情報、すなわち、お薦めユーザがもつ興味はユーザにとって関心がある、かつ、意外性を感じるの両方を満足する必要がある。図 2 は提案手法の全体像である。フォロー関係で繋がっているユーザの中で、ユーザ P の惹かれた興味 (赤い部分) と P の関心がありそうな興味 (青い部分) という特徴を導入し、P を惹き付ける興味を推測する。P を惹き付ける興味は、P と類似している人がもつ興味であり、かつ、自身の興味ではないので、「どれだけ魅力的か」と「意外性があるか」という要素を同時に満足すると考えられ、ユーザ P と類似している人がフォローしているユーザの中で、P を惹き付ける興味を含む情報を常に発信しているユーザ、すなわちこのような

興味をもつユーザをセレンディピティのあるユーザとしてユーザ P に推薦する。

以下、3.1 節は惹かれた興味という特徴を提案する。3.2 節は関心がありそうな興味を推測する手法を提案する。3.3 節はこれらの二つの興味を合わせて惹き付ける興味を推測する手法を提案する。また、惹き付ける興味の重要度を計算する方法を述べる。3.4 節は惹き付ける興味をもつユーザを推測する手法を提案する。

3.1 惹かれた興味

retweet は興味のある tweet に対して行われるので [13]、retweet によく出てこない興味はユーザの下位興味であると捉える。それによって、下位興味もユーザの興味であるが、よく出てこないの、ユーザ自身の興味に含まれていない下位興味はもう忘れた可能性が高いと考えられ、ユーザにとって意外性を感じると考えられる。また、このような興味は従来からもっている興味であるが、忘れた可能性もあるので、再推薦しても、全部興味があるわけではないと考えられる。そのため、ユーザベース協調フィルタリングに基づく、ユーザと類似している人の興味で、下位興味に含まれるものは、類似していない人からの興味よりも再びに好きになる可能性が高いと考えられる。本研究では、「惹かれた興味」とはユーザと類似している人が発信した情報を見てから、自身の興味と類似していない興味を惹かれ、その興味を含む情報をリツイートし、時間が経ってから忘れそうな興味であると考えられる。また、惹かれた興味は興味あるかつ意外性を感じるものであり、セレンディピティの二つの要素を同時に満足すると考えられる。

図 2 の赤色部分は惹かれた興味である。惹かれた興味を推測するために、まず、ユーザと類似している人の興味を推測する。次に、ユーザ自身の興味と下位興味を抽出した上で、惹かれた興味を推測する。

3.1.1 ユーザと類似している人の興味を推測する手法

本研究では、tweet と retweet に着目し、ユーザがフォローしている人の中でユーザと類似している人を推測した上で、類似している人がもつ興味を推測する。

興味は基本的に名詞で現れるので、Tweepy^{*2}を用いて、ユーザとユーザがフォローしている人の Timeline から tweet と retweet を抽出した上で、MeCab^{*3}を用いて名詞を抽出する。MeCab は現在、最もよく使われている日本語形態素解析器である。また、本研究では mecab-ipadic-NEologd という IPA 辞書を拡張したシステム辞書を利用し、Twitter にある新語の一部に対応できるようにする。ここで、ユーザの Timeline にはリプライもあるが、眩きやユーザ間の雑談が多いと考えられるため、本研究ではリプライに着目しない。

*2 <https://www.tweepy.org/>

*3 <https://taku910.github.io/mecab/>

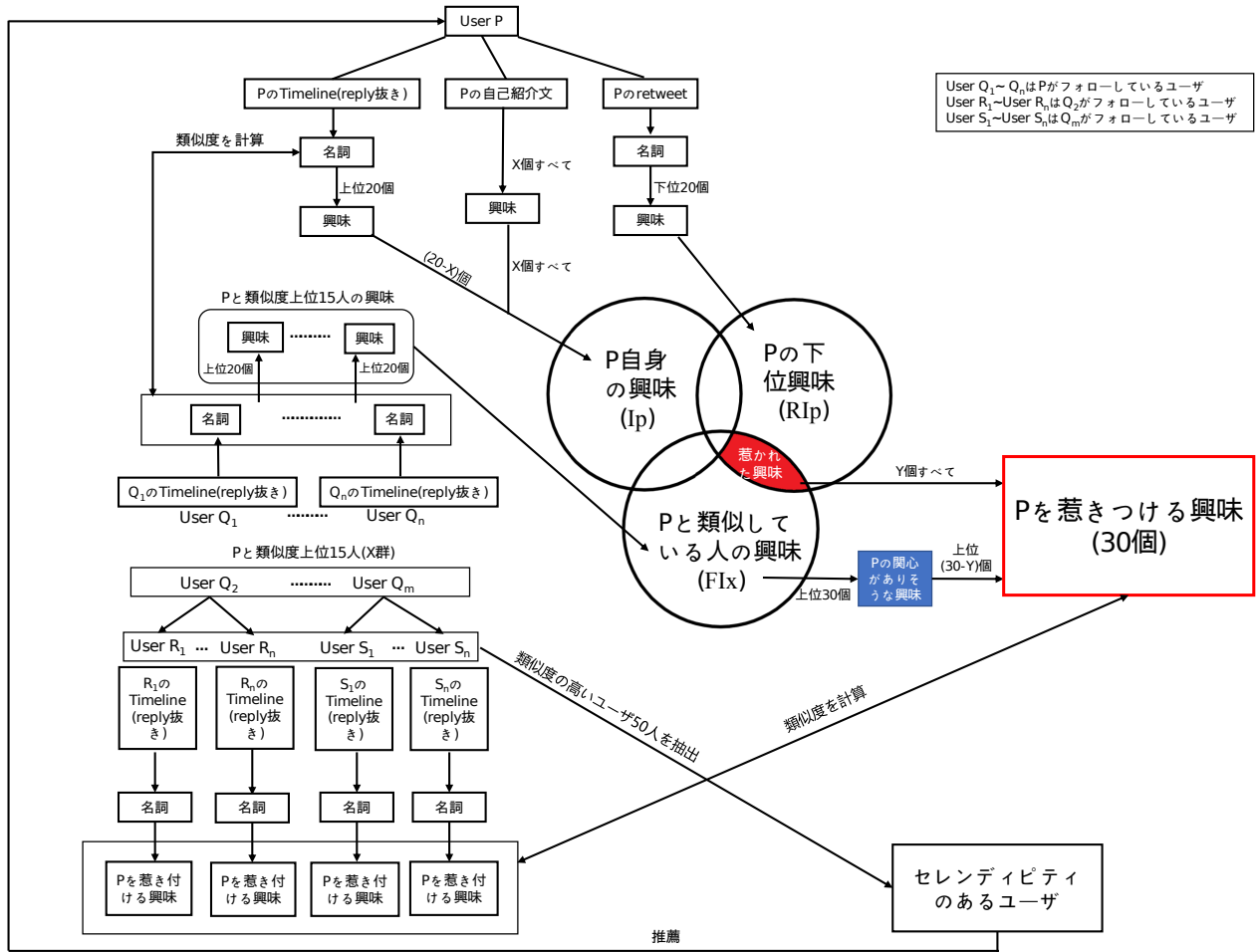


図 2 提案手法の全体像

次に、一つのユーザから抽出した名詞を一つの文書としてコーパスに入れた上で、各文書にある名詞の TF-IDF 値を計算する。TF-IDF は一つの文書にある単語の重要度を示すための手法である。

最後に、コサイン類似度を用いてユーザ同士の類似度を計算する。本研究では、BOW ではなく、名詞の TF-IDF 値に従って各ユーザのベクトルを作成する。また、ユーザと類似度の高いユーザがもつ興味は好きになる可能性が高いと考えられ、ユーザ P との類似度の最上位から 15 人をユーザと類似している人として X 群に入れる。これらのユーザの tweet と retweet から抽出した名詞を TF-IDF 値に基づいて降順に並べ替え、最上位から 20 位までの名詞を各人の興味とする。

3.1.2 ユーザ自身の興味と下位興味

ユーザ P の tweet と retweet から抽出した名詞の TF-IDF 値に基づいて降順に並べ替え、最上位から 20 位までの名詞を興味とする。また、自己紹介に書いた興味は Timeline から推測する興味より重要度が高いので、自己紹介にある興味と Timeline から推測する興味を合わせて 20 個までの興味をユーザ P 自身の興味とする。

下位興味は retweet によく出てこない興味であるとしているので、ユーザ P の retweet から抽出される名詞の TF-IDF 値によって最下位から 20 個までの興味を下位興味とする。

3.1.3 惹かれた興味を推測する手法

まず、ユーザ P の下位興味は忘れそうな興味であると考えられ、自身の興味を除いて残った興味は意外性を感じると考えられる。そのため、式 (1) に従ってユーザ P が意外性を感じる興味 (UI_P) を抽出する。ここで、 RI_P は P の下位興味、 I_P は P 自身の興味である。

$$UI_P = RI_P \setminus I_P \tag{1}$$

次に、ユーザ P は自分と類似している人から興味を惹かれるので、式 (2) に従って P の惹かれた興味 (AI_P) を推測する。ここで、 FI_X は P と類似している 15 人 (X 群) がもつ興味集合である。

$$AI_P = UI_P \cap FI_X \tag{2}$$

このようにして、ユーザ P の惹かれた興味を推測するが、ユーザ P と類似している人は同じ興味をもつ可能性があるため、 FI_X にある興味が多くなく、 UI_P との共通部分も多くない可能性がある。したがって、ユーザ P と類似している人 (15 人) から P の「関心がありそうな興味」を推測し、惹かれた興味を合わせてユーザ P を惹き付ける興味を推測する。

3.2 関心がありそうな興味を推測する手法

ユーザベース協調フィルタリングによって、ユーザ P と類似している人がもつ興味はユーザが好きになる可能性が高い。しかしながら、P にとって FI_X のすべてに興味があるわけではないと考えられる。そのため、まず、ユーザ P 自身の興味と惹かれた興味を FI_X から除いて、残った興味の重要度を計算し、重要度の高い興味は関心がありそうな興味としてユーザが好きになる可能性が高いと考える。また、ユーザ自身の興味を含まないので、意外性を感じると考えられる。そのため、セレンディピティにある二つの要素を同時に満足する。

式 (3) に従って FI_X に残った興味の重要度を算出し、重要度の最上位から 30 個までの興味を関心がありそうな興味にする。ここで、 \hat{r}_{Pj} は興味 j の重要度、すなわちユーザ P の興味 j に対して予測される評価値である。 $N_j(P)$ はユーザ P と類似している人の中で興味 j をもつユーザの集合である。 w_{PV} はユーザ P と V の類似度である。 r_{Vj} はユーザ V がもつ興味 j の評価値、すなわち TF-IDF 値である。つまり、類似している人がもつ興味の評価値 (TF-IDF 値) とユーザ P との類似度の両方を考慮した上で、ユーザ P の興味に対する予測の評価値、すなわち重要度を推測する。

$$\hat{r}_{Pj} = \frac{\sum_{V \in N_j(P)} w_{PV} \cdot r_{Vj}}{\sum_{V \in N_j(P)} |w_{PV}|} \quad (3)$$

3.3 惹き付ける興味を推測する手法

推測したすべての惹かれた興味 (Y 個) とユーザ P の関心がありそうな興味 (重要度上位 (30-Y) 個) を合わせて P を惹き付ける興味にする。さらに、惹かれた興味はユーザ P が従来からもつ興味なので、P の関心がありそうな興味よりユーザが好きになる可能性が高く、重要度も関心がありそうな興味より高くすべきであると考えられる。そのため、式 (4) によってユーザ P の惹かれた興味 $K(\hat{r}_{Pk})$ の重要度を算出する。ここで、 \hat{r}_{Pi} は推測したユーザ P の一番関心がありそうな興味 i の重要度である。 r_{Pk} は惹かれた興味 k の元の重要度、すなわちユーザ P の下位興味 k の TF-IDF 値である。つまり、惹かれた興味を関心がありそうな興味より重要度が高くなるために、惹かれた興味の元

の重要度を一番関心がありそうな興味の重要度に加えて、惹かれた興味の新しい重要度にする。

$$\hat{r}_{Pk} = \hat{r}_{Pi} + r_{Pk} \quad (4)$$

3.4 惹き付ける興味をもつユーザを推測する手法

まず、ユーザ P と類似している 15 人がフォローしているユーザに対して、各ユーザの tweet と retweet を含む名詞の TF-IDF 値を計算する。次に、各ユーザの名詞に P を惹き付ける興味があれば、それを抽出した上で、計算した TF-IDF 値を重要度として、コサイン類似度によって P を惹き付ける興味 (30 個) との類似度を計算し、類似度の高いユーザ 50 人を惹き付ける興味をもつユーザにする。最後に、このようなユーザはセレンディピティのあるユーザとしてユーザ P に推薦する。

4. セレンディピティを評価する手法

本章ではセレンディピティを評価する手法を述べる。Geらは意外性と有用性を考慮した上で、セレンディピティという評価指標を定量的に評価する手法を提案している [4]。同じお薦めユーザに対して意外性を感じるかどうかは被験者によって異なるので、評価するのが難しいという問題があるため、客観的な評価のために、彼らは推薦されたユーザの中で予測の正確さを重視する手法によって推薦されないユーザは意外性のあるお薦めユーザであると指摘している。これに基づき、本研究では意外性を式 (5) によって評価する。ここで、 PM (Primitive Prediction Model) は予測の正確さ (Prediction Accuracy) を重視する推薦手法を用いて推測されるお薦めユーザの集合である。たとえば、ユーザが多くの歌手をフォローすれば、他の歌手を推薦することは PM である。また、 RS (Recommender System) は本研究の提案手法を用いて推測されるお薦めユーザの集合である。

$$UNEXP = RS \setminus PM \quad (5)$$

本研究におけるセレンディピティの定義では、意外性だけでなく、ユーザが興味があるかどうかも重視するので、セレンディピティは式 (6) に従って評価する。ここで、推薦するユーザに対して興味があるかどうかは実験者によって評価し、その集合を $USEFUL$ で表す。また、 N は RS に含まれるユーザ数である。

$$serendipity = \frac{|UNEXP \cap USEFUL|}{|N|} \quad (6)$$

5. 予備実験

5.1 実験のデータと前処理

本手法の有効性を確認するために、被験者一人を対象に簡易な実験を行った。この被験者がフォローしているユーザの一部はロックしている状態なので、結果として、被験者がフォローしているユーザ 83 人を抽出した。次に、被験者と被験者がフォローしているユーザの Timeline から最新の tweet と retweet を合わせて 200 件抽出した。また、惹かれた興味を推測するために、被験者の最新の retweet も 200 件抽出した。

抽出された tweet と retweet には様々なノイズとなる情報がある。例えば、イベントのページの URL やユーザ ID などのような分析できない情報はノイズとなる情報である。このようなノイズがあると、ユーザ間の類似度と各ユーザの興味をうまく推測できないと考えられる。そのため、抽出されたデータから以下の前処理によってノイズを除去する。

- URL を除去する
- EMOJI を除去する
- @user ID を除去する
- 単語は全部小文字に正規化する
- 半角に統一する
- 数字を全部 0 に置き換える
- SlothLib^{*4} を利用し、情報のない Stop-Word を除去する

5.2 実験の設定

前節で述べた前処理を行ったとしても、MeCab を用いて抽出した名詞にはまだノイズがある。たとえば、「笑」や「下」など興味ではない名詞が存在する。そのため、被験者と被験者と類似しているユーザ (15 人) に対して計算する名詞の TF-IDF 値によって、人手で最上位から興味ではない名詞を削除することによって、各ユーザの興味を抽出する。また、実験者の retweet にある名詞を TF-IDF 値によって最下位から興味ではない名詞を人手で削除することによって、実験者の下位興味を抽出する。

次に、抽出した被験者と類似している 15 人がフォローしているユーザの中で、被験者がフォローしているユーザを削除した上で、ランダムに 500 人を選ぶ。また、各人の最新の Timeline から tweet と retweet を合わせて 200 件を抽出し、提案するセレンディピティのあるユーザを推測する手法によって 500 人からお勧めユーザ 50 人を推測し、推薦リスト 1 を作成する。

最後に、本手法との比較対象として、安部らの提案手

法 [10] に基づいてお勧めユーザ 50 人を推測し、推薦リスト 2 を作成する。その後、二つの推薦リストにあるユーザの Timeline と自己紹介を実験者に見てもらった上で、アンケートに記入してもらう。アンケートには二つの質問がある。質問 1 は「推薦するユーザに対して興味ありますか?」である。回答には 5 段階評価を利用し、1 は興味なし、5 はとても興味ありとする。また、閾値を 3 に設定し、3 以上は被験者が推薦するユーザに対して興味をもつとみなす。質問 2 は「推薦するユーザに対して意外性を感じますか?」である。回答には 5 段階評価を利用し、1 は全く意外なし、5 はとてもびっくりするとする。また、閾値を 3 に設定し、3 以上は被験者が推薦するユーザに対して意外性を感じるとみなす。

5.3 評価方法

本研究では、Ge らによる意外性を評価する方法 (式 (5)) に加えて、アンケートで実験者から主観的な意外性を評価する手法を利用し、式 (6) に基づいて安部らが提案した手法と比較する。評価手法 1 は、式 (5) に従って客観的な意外性の評価値を用いてセレンディピティの評価値を算出する。式 (5) において、PM は前節で説明したランダムに抽出した 500 人の中で、被験者自身の興味と類似度の最上位からの 50 人である。USEFUL はアンケートにある質問 1 によって興味があるとみなされたユーザの集合である。

評価手法 2 は、アンケートの質問 2 によって被験者が意外性を感じたユーザを UNEXP とする。また、USEFUL は評価手法 1 と同じデータを利用する。

5.4 実験結果

図 3 に式 (5) を用いて二つの推薦リストにあるユーザの意外性を客観的に評価した結果を示す。横軸は推薦するユーザの人数である。縦軸は意外性を感じたユーザの人数である。

また、図 4 に客観的な意外性を利用してセレンディピティを評価した結果を示す。縦軸はセレンディピティの評価値である。

さらに、図 5 にアンケートによって意外性を評価した場合の結果を示す。最後に、図 6 に主観的な意外性を用いてセレンディピティを評価した結果を示す。

5.5 考察

図 3 に示されているように、式 (5) による意外性の評価では、既存研究ではユーザと類似していないユーザに着目しているため、推測するお勧めユーザと PM によって推測するお勧めユーザの共通部分はなく、意外性のあるユーザの集合が大きくなる傾向がある。しかしながら、お勧めユーザに対して興味がなければ、セレンディピティのあるユーザにならない。図 4 に示されているように、お勧め

^{*4} <http://svn.sourceforge.jp/svnroot/slothlib/CSharp/Version1/SlothLib/NLP/Filter/StopWord/word/Japanese.txt>

ユーザが 20 人の場合、二つの手法はセレンディピティの評価値は同じであるが、他の場合には提案手法の方が既存研究よりセレンディピティの評価値が高い傾向がある。そのため、意外性という一つの要素の評価値が大きくても、セレンディピティの評価値を向上させるわけではないと考えられる。

一方で、図 5 に示されているように、意外性を実験者によって評価した場合には、提案手法では既存研究より意外性のあるお薦めユーザが多くなる。ここで、ユーザ間の類似度は tweet と retweet から抽出した名詞によって推測するので、常に同じジャンルの情報を発信している公式アカウントでは抽出した名詞の多様性が少なく、様々な情報を発信する実験者との類似度が低くなる。既存研究は類似していないユーザに着目するので、抽出したお薦めユーザは個人ユーザではなくて公式アカウントになる可能性が高く、ユーザがすでに知っている可能性が高いと考えられ、主観的な意外性が低くなる。そのため、ユーザと類似していない人に着目してセレンディピティのあるユーザを推測する場合には、式 (5) に従って意外性のあるユーザを評価する手法は、セレンディピティの評価値を実際以上に上げてしまう可能性があると考えられる。さらに、図 6 に示されているように、提案手法は既存研究よりセレンディピティの評価値が高いことが見られる。また、二つの手法の結果には小さくない差が見られる。

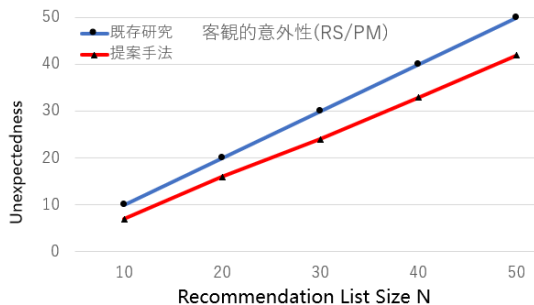


図 3 客観的な意外性

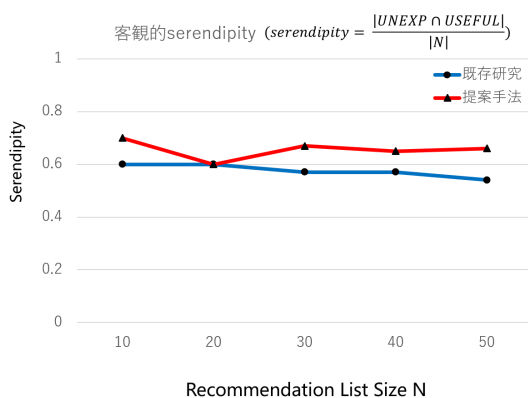


図 4 客観的なセレンディピティ

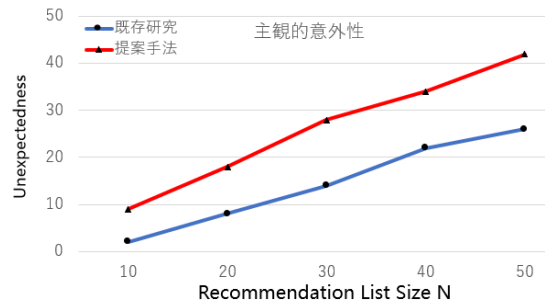


図 5 主観的な意外性

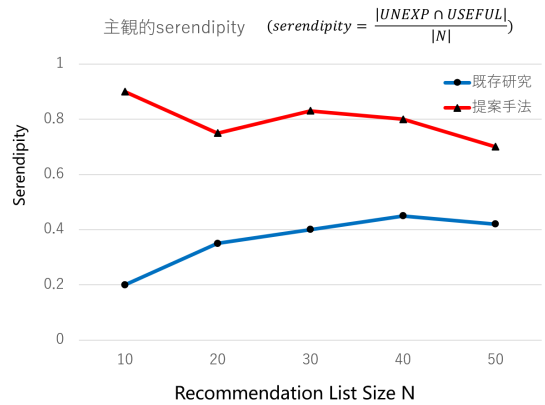


図 6 主観的なセレンディピティ

6. おわりに

本論文では、ユーザベース協調フィルタリングに基づいて、惹かれた興味と関心がありそうな興味という特徴を導入し、ユーザを惹き付ける興味を推測する手法を提案した。さらに、ユーザを惹き付ける興味に基づいて、ユーザと類似している人がフォローしているユーザの中で、セレンディピティのあるお薦めユーザを推測する手法を示した。また、提案手法によって推測されるお薦めユーザは、いずれの実験手法でも既存研究よりセレンディピティの評価値が高くなることが示された。

今後は、ユーザ間の類似度推定の性能を上げるために、コサイン類似度に加えてピアソンの相関係数を利用することを試みる、また、セレンディピティを評価する指標を吟味し、より多くの被験者により、実験を行う。

参考文献

- [1] Chiu Y S, Lin K H, Chen J S. A social network-based serendipity recommender system[C]//2011 International Symposium on Intelligent Signal Processing and Communications Systems (ISPACS). IEEE, 2011: 1-5.
- [2] Zheng, Qianru, Chi-Kong Chan, and Horace HS Ip. An unexpectedness-augmented utility model for making serendipitous recommendation. Industrial conference on data mining. Springer, Cham, 2015.
- [3] Kotkov, Denis, Jari Veijalainen, and Shuaiqiang Wang. A serendipity-oriented greedy algorithm for recommendations. WEBIST 2017: Proceedings of the 13rd

- International conference on web information systems and technologies. Volume 1, ISBN: 978-989-758-246-2. SCITEPRESS Science And Technology Publications, 2017.
- [4] Ge M, Delgado-Battenfeld C, Jannach D. Beyond accuracy: evaluating recommender systems by coverage and serendipity[C]//Proceedings of the fourth ACM conference on Recommender systems. ACM, 2010: 257-260
 - [5] Herlocker J L, Konstan J A, Terveen L G, et al. Evaluating collaborative filtering recommender systems[J]. ACM Transactions on Information Systems (TOIS), 2004, 22(1): 5-53.
 - [6] Tacchini E. Serendipitous mentorship in music recommender systems[J]. 2012.
 - [7] Sarwar B M, Karypis G, Konstan J A, et al. Item-based collaborative filtering recommendation algorithms[J]. Wwww, 2001, 1: 285-295.
 - [8] Said A, Fields B, Jain B J, et al. User-centric evaluation of a k-furthest neighbor collaborative filtering recommender algorithm[C]//Proceedings of the 2013 conference on Computer supported cooperative work. ACM, 2013: 1399-1408.
 - [9] Scott Piao, and Jon Whittle. A feasibility study on extracting twitter users' interests using nlp tools for serendipitous connections. 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing. IEEE, 2011.
 - [10] 安部高城, 佐藤哲司 (2015) 「ソーシャルメディアにおけるセレンディピティを考慮したユーザー推薦手法の提案」[DEIM Forum 2015] ,B2-61
 - [11] Su X, Khoshgoftaar T M. A survey of collaborative filtering techniques[J]. Advances in artificial intelligence, 2009, 2009.
 - [12] Kapoor, Komal, et al. I like to explore sometimes: Adapting to dynamic user novelty preferences. Proceedings of the 9th ACM Conference on Recommender Systems. ACM, 2015.
 - [13] Boyd D, Golder S, Lotan G. Tweet, tweet, retweet: Conversational aspects of retweeting on twitter[C]//2010 43rd Hawaii International Conference on System Sciences. IEEE, 2010: 1-10.