

## Twitterにおけるセレンディピティのあるおすすめユーザの発見

徐哲林<sup>†</sup>周娟<sup>‡</sup>高田秀志<sup>‡</sup>立命館大学大学院情報理工学研究科<sup>†</sup>立命館大学情報理工学部<sup>‡</sup>

## 1 はじめに

近年、ソーシャルメディアの普及に伴い、人々がネット上で情報発信・情報共有を行うようになった。その中で、Twitterは、2018年のデータによると、世界で3億2600万人、日本国内で4500万人のアクティブユーザに利用されており、国内外に広く普及している。

Twitterでは、利用者がキーワード検索機能を利用してフォローしたいユーザを発見することができるが、キーワードが曖昧な場合、適切な結果が表示されないこともある。Twitter公式サイトでは、利用者が関心のありそうなユーザを推薦するサービスを提供している。しかしながら、表示されているおすすめユーザは、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る。また、ユーザは必ずしも同じジャンルのユーザだけを期待しているわけではないと考えられる。そのため、本研究では、ユーザの満足度を向上させるために、ユーザのツイートとリツイート間に現れた興味の偏りに着目し、2ホップまでのフォロー関係の中で、セレンディピティのあるおすすめユーザを発見する手法を提案する。

## 2 セレンディピティ(serendipity)

Resnickが推薦システム概念を導入したのは1997年である。しかし、現在に至っても、利用者の嗜好に一致した推薦が必ずしも、利用者の満足度を向上できるわけではないという問題を抱えている[1]。そのため、単に推薦精度に着目するのではなく、セレンディピティという推薦精度以外の評価指標により推薦システムを評価することが重要である[2]。また、ユーザにとっては、発見が難しく、かつ満足度の高いセレンディピティのあるユーザの推薦が望まれている[3]。

セレンディピティとは、利用者にとってどれだけ魅力的かつ意外性があるかという要素が加わった概念である[4]。そのため、セレンディピティの定義に基づくと、本研究ではセレンディピティのあるおすすめユーザとは利用者にとって有用な情報発信をしているユーザ、すなわちフォローしたいかつ意外な情報を発信するユーザである。

## 3 先行研究

## 3.1 セレンディピティユーザの推定手法

安部らはフォロー関係で繋がっているユーザの中で、利用者と興味が似ていないユーザ、すなわち利用者と類似度の低いユーザに着目し、セレンディピティのあるユーザを抽出することを試みた[3]。安部らの研究では、利用者とフォローしているユーザ間の類似度に着目した。一方で、本研究は、ユーザ自身のツイートとリツイート間に現れた興味の偏りに着目する。

## 3.2 セレンディピティに関する評価指標

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指標を定量的に評価する手法を提案している[4]。式(1)はセレンディピティを評価する式である。意外性のある推薦ユーザの集合(UNEXP)と利用者にとって有用な情報を発信する推薦ユーザの集合(USEFUL)の共通部分が提案したシステムの推薦ユーザの集合(N)に占めた割合がセレンディピティの値である。

$$\text{serendipity} = \frac{|UNEXP \cap USEFUL|}{|N|} \quad (1)$$

式(1)において、有用な情報を発信する推薦ユーザ(USEFUL)は利用者によって決定する。また、意外性のある推薦ユーザ(UNEXP)を評価するために、プリミティブ推薦(primitive prediction)を導入する。プリミティブ推薦とは利用者プロフィールや過去の利用者の行動履歴情報に基づいて推定された推薦ユーザの集合である。つまり、プリミティブ推薦は精度の高い推薦手法のことである。そのため、提案した推薦手法の結果(RS)とプリミティブ推薦の結果(PM)の差(式(2))が意外性のある推薦ユーザである。

$$UNEXP = RS \setminus PM \quad (2)$$

## 4 提案手法

## 4.1 提案手法の概要

本提案では、利用者のツイートに現れる興味の上位部分とリツイートに現れる興味の下位部分の間に偏りがあると仮定する。この興味の偏りの中で、リツイートで現れる興味は利用者の惹かれた興味だと考えられる。また、Golderらは、利用者と互いにフォローしているユーザがフォローしているユーザは利用者にとってフォローしやすいと述べている[5]。そのため、このような2ホップまでのフォロー関係の中で、惹かれた興味に基づいてセレンディピティのあるおすすめユーザを抽出する手法を提案する。

## 4.2 惹かれた興味

利用者自身が発信する内容より、フォローしているユーザによって発信される内容の方が多いので、利用者がフォローしているユーザのツイートとリツイートを見ると、それらに突然興味を引かれ、リツイートする可能性があると考えられる。そのため、このような利用者のリツイートに現れる重要度の低い興味、かつその利用者のツイートに現れない重要度の高い興味は、惹かれた興味だと考えられる。

図1は惹かれた興味である。惹かれた興味は利用者にとって魅力があるため、リツイートするので、利用者にとって有用性があると考えられる。また、利用者のすべての興味の中における出現頻度は高くないので、そのような興味に基づいて推薦すると、利用者にとっ

Discovery of recommended users with serendipity on Twitter

<sup>†</sup>Zhelin XU <sup>‡</sup>Juan ZHOU <sup>‡</sup>Hideyuki TAKADA<sup>†‡</sup>Faculty of Information Science and Engineering, Ritsumeikan University

て意外性を感じる可能性があると考えられる。そのため、惹かれた興味に基づいて、セレンディピティのあるユーザを抽出できるようになると考えられる。

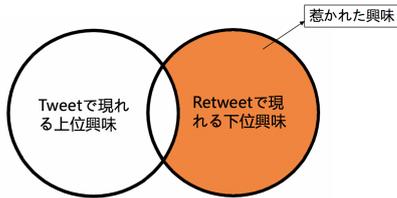


図1:惹かれた興味

### 4.3 惹かれた興味の検証

本研究では、利用者のツイートにおいて上位10%を占める名詞から現れる興味は利用者本来が持っている興味だと仮定した上で、リツイートから抽出された名詞のどの部分に惹かれた興味が存在するのかを検証するために、Twitterでランダムに4人の利用者を選択する。次に、利用者の最新のツイートとリツイートを150件ずつ抽出する。リツイートから抽出された名詞を上位から10%ずつに分けて、最下位から4つの部分を抽出する。

4人の結果を分析したところ、最下位の部分にはノイズが多く存在し、興味が一番現れる部分は最下位から第4部分であることが分かった。表1はある利用者の第4部分、表2はこのユーザのツイートの上位10%の名詞で現れる興味である。表3はツイートとリツイート間の偏りから抽出した惹かれた興味である。サンプル数が少ないが、惹かれた興味が存在することが見られた。

表1 Retweetの第4部分で現れる興味

	TF-IDFの値
けん玉	0.017901
ぬいぐるみ	0.017901
ひかる	0.017901
ふたりはアリキュア	0.017901
ふんだりけったり	0.017901
まどか	0.017901
アットホーム	0.017901
アドアーズ	0.017901
アニメ	0.017901
アニメイト	0.017901
イナズマイレブン	0.017901
イラスト	0.017901
オカルト	0.017901
カブトボーグ	0.017901
ガリアン	0.017901
キャラクターボイス	0.017901
キュアセレネ	0.017901
キュアソレイユ	0.017901
キュアハッピー	0.017901
キュアミルキー	0.017901
クス	0.017901
ゲーム情報	0.017901
サガ	0.017901
サマ	0.017901
シロクロ	0.017901
スクワット	0.017901

表2 tweetの最上位10%名詞で現れる興味

	TF-IDFの値
バイト	0.192269
カワコン	0.115361
ギター	0.115361
国木田花丸	0.082081
あいこや	0.076908
はなまる	0.076908
少女漫画	0.076908
山田涼介	0.076908
田丸篤志	0.076908
aqours	0.05472
baba嵐	0.038454
cm	0.038454
fukase	0.038454
honeyworks	0.038454
お菓子	0.038454
お雑煮	0.038454
きりしま	0.038454
こぶ	0.038454
さかな	0.038454

表3 惹かれた興味

	TF-IDFの値
けん玉	0.017901
ぬいぐるみ	0.017901
ひかる	0.017901
アットホーム	0.017901
アドアーズ	0.017901
アニメイト	0.017901
イナズマイレブン	0.017901
イラスト	0.017901
オカルト	0.017901
ガリアン	0.017901
クス	0.017901
ゲーム情報	0.017901
サガ	0.017901
サマ	0.017901
シロクロ	0.017901
スクワット	0.017901

### 4.4 提案手法の具体的なステップの概要

本提案の手法を図2に示す。具体的なステップは以下の通りである。

1. 利用者のツイートとリツイートを収集する。
2. 形態素解析ツール MeCab を用いてツイートとリツイートに現れる名詞を抽出する。
3. TF-IDF を用いて抽出された名詞の重要度を計算し、ツイートとリツイートで現れる興味をリストアップする。
4. ツイートとリツイートの中に現れる興味の偏りから、惹かれた興味を抽出する。
5. 利用者Pと互いにフォローしているユーザの中で、コサイン類似度を用いて利用者の惹かれた興味と類似度の高いユーザを抽出する。これは、図??におけるユーザSである。
6. 惹かれた興味と類似度の高いユーザがフォローしているユーザの中で、類似度の高いユーザを抽出する。図2に示すユーザRは、セレンディピティのあるおすすめユーザである。

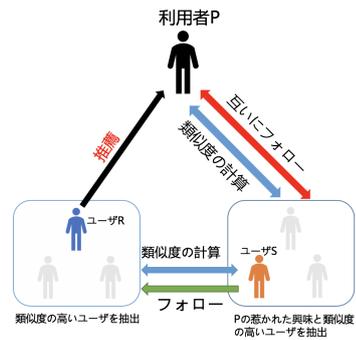


図2:本手法の概要図

### 5 おわりに

本稿では、利用者の惹かれた興味に基づき、セレンディピティのあるおすすめユーザを発見する手法を提案した。

今後は、Geらが提案したセレンディピティを評価する指標 [4] を用いて、提案手法で抽出されたセレンディピティのあるおすすめユーザの精度を検証し、評価を行う。

### 参考文献

- [1] 福島良典, 大澤幸生. ソーシャルメディアを利用したセレンディピティな情報推薦. 人工知能学会全国大会論文集 2012 年度人工知能学会全国大会 (第 26 回) 論文集. 一般社団法人人工知能学会, 2012: 3E1R66-3E1R66.
- [2] 土方嘉徳. 嗜好抽出と情報推薦技術. 情報処理, 2007, 48(9): 959-961.
- [3] 安部高城, 佐藤哲司. 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
- [4] Ge M, Delgado-Battenfeld C, Jannach D. Beyond accuracy: evaluating recommender systems by coverage and serendipity. Proceedings of the fourth ACM conference on Recommender systems. ACM, 2010: 257-260.
- [5] Golder, Scott A., and Sarita Yardi. Structural predictors of tie formation in twitter: Transitivity and mutuality. IEEE Second International Conference on Social Computing (SocialCom), 2010.