

Twitter における言語解析を用いたユーザのグルーピング

阿原 正弥 杉浦 稔介 野田 悠介 横井 優斗 濱川 礼
中京大学 情報理工学部 情報システム工学科

1. 概要

本研究は、ソーシャルネットワーキングサービス(以下 SNS とする)である Twitter において、ユーザにバッジによるタグ付けをし、グルーピングを行うことで、Twitter においてのユーザ検索の効率化を図り、ユーザ間のコミュニケーションの拡大を補助するものである。

2. 背景

Twitter([1])は SNS として需要が急速に拡大し、アクティブユーザは 1 億人を超える([2]). Twitter はフォローしたユーザとツイートを交換する情報サービスであるため、フォローするユーザを検索する必要がある. 1 億人を越える利用者の中から任意のユーザを検索する方法は ID, プロフィール, ツイートからの文字列検索であるため、現在の検索方法では意図しない検索結果が検出される。

例として、サッカーが趣味のユーザを検索する際に、対象文字列を「サッカー」にした場合、ID, ツイート, プロフィールに「サッカー」の文字列が含まれていれば検出されるため、1 度だけツイートをした人や、プロフィールで苦手なスポーツとしてサッカーを挙げている人も検索結果として検出されてしまう. その他 Twitter を利用する既存のアプリケーションにおいても文字列検索を利用するものが多い([3]~[6]).

3. 提案手法

ユーザにバッジを付与、バッジが付与された回数と付与したユーザを記録したデータベース(以下「DB」とする)を作成する. 文字列からユーザを検索するのではなく、この DB からバッジの付与回数一覧や、他のユーザからのバッジ付与による評価を利用することで意図したユーザ検索を実現する。

3.1. バッジ

バッジとは、ユーザの特徴、興味、状況についての情報を画像に置き換えたものである。

バッジは名称、枚数、バッジを付与した人で構成

Grouping users with a language analysis
on Twitter

Masaya Ahara, Ryosuke Sugiura, Yusuke
Noda, Yuto Yokoi and Rei Hamakawa

される. 枚数は付与された回数を表す. バッジの枚数が多くなれば、ユーザがバッジに関連する要素に興味関心が高いことを意味する. 加えて付与した人を表示する事でそのユーザが他のユーザからどのように評価されているのかを判断することができる。

バッジを用途から 3 種類に分類した[図 1].

1. 基本情報バッジ.

名称のみ設定される. 性別や位置情報など、1 つあれば十分なものがこれに当てはまる。

例えば、「男性」、「学生」、「愛知県」である。

2. ツイートの文字解析による自動取得バッジ.

バッジに対応したツイートをすると付与される. 名称と枚数が設定されており、枚数が多いほどバッジに対応した内容に興味関心が高いことを表現する. 例えば、「テニス」、「カラオケ」、「映画」、「料理」である。

3. 他のユーザから付与されるバッジ.

ツイートからでは表現できない情報を他のユーザからの目線で付与することで、客観的な評価を表現することができる. 例えば、「イケメン」、「食いしん坊」、「リア充」である。

バッジの種類は独自に判断した 250 種類を用意した。




- 基本的な情報を表現するバッジ (最新の情報より取得)
 属性:愛知県 枚数:なし つけた人:なし
- 自動取得されたバッジ (ツイートより取得)
 属性:F1 枚数:30 つけた人:なし
- 評価を表すバッジ (他ユーザからの評価によって取得)
 属性:リア充 枚数:45 つけた人:ユーザ名

図 1: バッジの種類

3.2. バッジ検索

検索画面からバッジを選択し、そのバッジを持つユーザの一覧を取得する. 文字列ではなくバッジの個数や付与したユーザを参考に気になるユーザを選択することで、意図にあったユーザの検索を実現する[図 2]. これにより、既存の検索による意図しない結果を改善することができる。

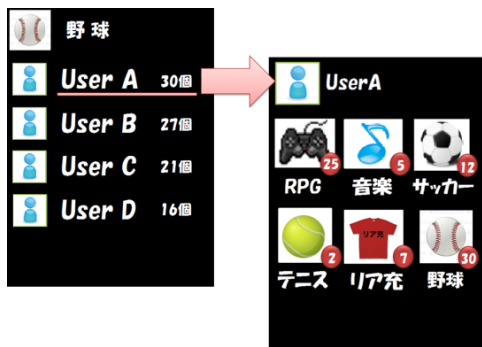


図 2：表示画面

4. 処理の流れ

本研究では提案する内容を図 3 で示したアプリケーションとして実装した。ここでは、実装したアプリケーションの内部処理について説明する。本アプリケーションでは「クライアント部」、「通信管理部」、「サーバ管理部」、「文字解析部」の 4 つで構成されている。

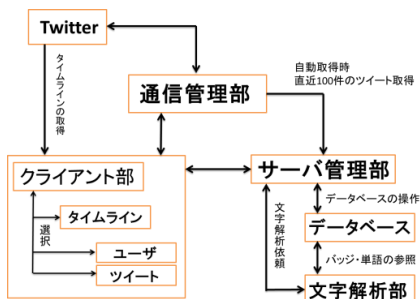


図 3：処理の手順

クライアント部

TL の表示と、プロフィール情報とバッジ情報の表示、ツイートの投稿を行う。

表示された TL より気になるツイートを選択し、ツイートの詳細画面の表示をおこなう。この画面からツイートに対しての評価と、他のユーザからの評価を確認することができる。また、この画面からツイートを投稿したユーザ、評価したユーザのプロフィール画面に移動する。プロフィール画面はユーザが過去投稿したツイートだけでなく、ユーザの付与されたバッジの履歴を確認することができるようになっている。

通信管理部

Twitter と本アプリケーション間のツイートの取得、および送信と、「クライアント部」と「サーバ部」のユーザ情報とバッジ情報の通信管理を行う。

通信でやり取りされる情報は、Twitter 上でのツイート内容及び関わったユーザの情報だけでなく、付与するバッジの種類、バッジの付与の命令及び削除の命令である。

サーバ管理部

バッジ情報の DB と、ユーザ情報の DB の管理を行う。 バッジ情報での DB はクライアントから

送られてきたツイートの ID, 投稿したユーザ ID, 付与を行ったユーザ ID, 及びバッジの種類で構成される。この DB よりツイートの ID から検索を行うことで、ツイートの評価を行ったユーザを検索することができる。また、投稿したユーザ ID より検索をかけることで、ユーザが過去付与されてきたバッジの種類及び枚数の情報を取得することができる。

文字解析部

ツイートの文字列にバッジに対応する文字列があるか検索する。検索のために独自で作成した、文字列の DB をもち、バッジと対応する文字列は 1 対多の関係にある。検出した文字列に対応するバッジを全てサーバ管理部にバッジ情報として送る。このバッジ情報をサーバ管理部の DB で管理、検索で利用することでグルーピングを実現する。

5. 評価

評価は以下の項目で行う。

1. 既存の検索より使用しやすいか
2. バッジの数は適当か
3. バッジの付与は適当か
4. 表示画面は見やすいか
5. 処理速度は適当か

現在、評価中である。

6. 考察

バッジの種類と対応する単語の充実が課題となる。また「クライアント部」の表示の最適化と処理速度の向上も必要になると考えられる。

7. 参考文献

- [1] Twitter
<http://twitter.com>
- [2] 特集:「Twitter とソーシャルメディア」
情報知能学会誌, Vol.27 No.1 pp.34-74 (2012)
- [3] Twittag (関連サービス)
<http://twittag.org>
- [4] Twitter (アプリケーション)
<https://market.android.com/details?id=com.twitter.android>
- [5] twicca (アプリケーション)
<https://market.android.com/details?id=jp.r246.twicca>
- [6] TweetDeck (アプリケーション)
<https://market.android.com/details?id=com.tweetdeck.android.app>