

# ユーザー関係の分析による Twitter コミュニティ抽出

PM コース 矢吹研究室 1142123 渡邊雄大

## 1. 研究背景

コミュニケーションツールとして、Social Networking Service (SNS) を使用している人はとても多くいる。その中でも Twitter は、SNS を代表する 1 つである。なぜなら Twitter はアクティブユーザー数が 2 億 3 千万人もいることだけでなく、ツイートと呼ばれるマイクロブログが一日平均で 5 億件も送信されているからだ(2014 年 9 月現在)。そのため Twitter は、調査する価値のある SNS の 1 つであると考えられる。

Twitter はツイートと呼ばれる短い文字列を投稿するためのサービスである。自分以外のユーザーのツイートを読むためには、そのユーザーのページにアクセスするか、そのユーザーをフォローする必要がある。フォローしているユーザーのツイートは、ひとまとめにされ、タイムラインを形成する。誰が誰をフォローしているかという情報(フォロー関係)は、Twitter におけるユーザーのつながりの一つの表現である。仮に Twitter 上でユーザーがコミュニティを形成していたとすれば、フォロー関係にもそれが反映されていると思われる。そのコミュニティを抽出することができれば、フォローすべきユーザーの発見が容易になるなど、Twitter のユーザビリティが大きく向上することが期待される。

## 2. 研究目的

検索したいユーザーの Twitter 上に持っているフォロー関係から、そのユーザーの持つ実際のコミュニティを見つけ出す。さらに、この研究を行うことで、プロジェクトを円滑に行う為の人的資源マネジメントとして活用することを目指す。

## 3. 類似研究

ユーザーのフォロワーとのつながりについての研究として[1]が挙げられる。この研究では、フォロー関係からクラスタリング分析を行い、その結果からクラスタを出力できるシステムを作成している。この研究はシステムが結果として出力したクラスタの内容と、手作業で作成した理想のクラスタを比べることで研究の評価を行っている。

## 4. 成果物のイメージ

Twitter のフォロー関係を使って、調べたいユー

ザーの持つ周辺のコミュニティを抽出する。抽出・評価方法を研究の成果物とする。

## 5. 研究方法

まず調べたいユーザーが、Twitter で使用しているユーザー名を入力することで、TwitterAPI から調べたいユーザーのフォローしている相手と、フォローされている相手のユーザー名を取得する。この作業を取得したユーザーの中でも繰り返すことで、調べたいユーザーの持つフォロー関係を見つけ出す。それらのフォロー関係の内容を詳しく分析することでコミュニティが抽出できると考える。

## 6. 研究の評価方法

実際の正解データとなるコミュニティのユーザーリストを複数作成し、そのリストを比較対象として精度と再現率を出し、評価を行う。正解データのユーザー数の中から抽出したユーザーの正解数の割合が再現率、抽出したユーザー数の中から抽出したユーザーの正解数の割合が精度である。

## 7. 進捗状況

Linux のディストリビューションの 1 つである Ubuntu 上で、Python を使って TwitterAPI を動かし、指定したユーザーのフォロー、フォロワー情報を JSON 形式で取得した。

## 8. 今後の計画

今後の計画は以下のものとする。

日程	内容
10 月	分析のためのユーザー情報の収集 正解データとなるリストの作成
11 月	収集したデータからリストの抽出 作成したリストの評価
12 月	論文の執筆
1 月	発表資料の作成

## 参考文献

- [1] 山下拓也・佐藤晴彦・小山聡・栗原正仁(2013)「フォロー関係に基づく Twitter ユーザーの分類」『第 75 回全国大会講演論文集』pp.107-109 情報処理学会