

ユーザ関係の分析による Twitter コミュニティ抽出

プロジェクトマネジメントコース 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。自分以外のユーザのツイートを読むためには、そのユーザのページにアクセスするか、そのユーザをフォローする必要がある。フォローしているユーザのツイートは、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 におけるユーザのつながりの一つの表現である。仮に Twitter 上でユーザ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

検索したいユーザの Twitter 上に持っているフォロー関係から、そのユーザの持つ実際のコミュニティを Twitter の機能であるリツイートを解析することによって見つけ出す。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ため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較することで研究の評価を行っている。

4. 研究方法

この研究を説明する際に使用する用語を以下の表 1 にまとめた。

表 1 使用する用語

リツイート A	観測されたリツイート
ツイート A	リツイート A のオリジナルのツイート
ユーザ A	ツイート A のユーザ
ユーザ B	リツイート A のユーザ
指標 1	ユーザ B がユーザ A をフォローしていない確率
指標 2	ユーザ B がツイート A をリツイートした人たちがフォローしていない確率

まずこの研究では指標 1 と指標 2 を導き出す。そのために TwitterAPI の User streams の機能を使い、自分のフォローしているユーザの行ったリツイートのみを抽出し保存する。この保存されたリツイートの中からリツイート A を使用し研究を行ったとする。この場合はユーザ B のフォローしている相手とツイート A をリツイートしたユーザ B 以外の相手を取得する。取得したリツイートの全体と指標 1、指標 2 の関係は以下の図 1 のようになる。

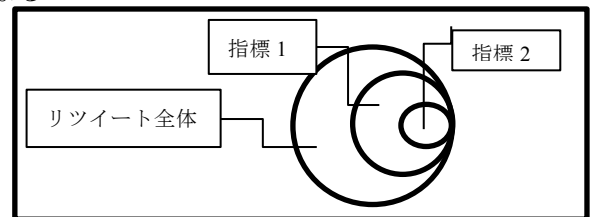


図 1 リツイート全体、指標 1、指標 2 の関係

5. 研究結果

この研究では成果物を研究の結果から得られた 2 つの指標とする。指標 1 は 60%であった。指標 2 は 40%であった。

6. 考察

指標 1 と指標 2 の結果から確認できるように、多くのユーザ達がフォローして無いユーザのツイートも確認し、リツイートを行っていることが分かった。これはフォロー関係によって繋がっていると思われていた従来の Twitter ユーザ同士のネットワークの形ではない別の形のネットワークがある可能性が示された。

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 ユーザの分類」『第 75 回全国大会講演論文集』pp.107-109 情報処理学会