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SNSの中でもTwitterはそれを代表する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．

そのため，コミュニティを抽出するためにユーザー間のフォロー関係の情報を取得することでフォロー関係を俯瞰的に確認しようと考えた．

2.　研究目的

　本研究の目的は，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できることを目指す．

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し，その研究方法や評価方法を研究の成果物とする．

5. 研究方法

コミュニティの抽出方法は，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からその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．これを最初に取得したユーザーの中から複数回繰り返すことで，調べたユーザーの持つフォロー関係を見つけ出すことができる．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

　研究の評価方法は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．なお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を再現率とし，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を精度とする．

7.　進捗状況

　現在の進捗状況は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形式で取得することができた．

8.　今後の計画

　ユーザーの情報を取得することができたが，これらのデータをどのように分析していくのか決まっていない．まずは分析を行うための方法を明確にする，それと同時に正解データとなるリストを作っていく．

　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