Twitterのユーザー関係に基づくコミュニティ抽出

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（以下，SNSと呼ぶ）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．

　このTwitterのマイクロブログを購読するために用いられるフォローという機能を使うことでユーザーは友人の情報や，興味のある情報を簡単に受け取ることができる．この機能によってフォローし，フォローされること（以下フォロー関係と呼ぶ）がそのユーザーの友人関係や趣味による関係がTwitter上で形成されている現実のコミュニティにとても近いものだと思われる．このコミュニティを抽出することは，フォローユーザーをまとめる際のリストを手動で作成するといった手間を省くことができるため必要である．

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人物のTwitter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

　Twitterに関する研究は盛んに行われているが，ユーザーのフォロワーとのつながりについて研究しているのは，フォロー関係に基づくTwitterユーザーの分類の研究[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評価したものを研究の成果物とする．

この研究方法は，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と考える．さらに取得したユーザーの一致を距離と考え，その距離の近いユーザーをコミュニティとしそれを結果とする．

5.　評価方法

　成果物の評価方法は実際の正解データとなるリストを複数作成しそのリストと，成果物のシステムの出すリストとの一致する数を調べ，再現率と精度を出すことで評価を行う．

6.　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（Java Script Object Notation）形式で取得することができた．

7.　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