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．これはSNSの時代とも呼ばれ，Twitterはその時代の一画を担っている．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サービスが一日平均で5億件も送信されていることは多くのSNSの中でも調査する価値のある１つだと考えられる．（2014年9月現在）

　そしてこのマイクロブログサービスを購読するために用いられるフォローという機能は友人の情報や，興味のある情報を簡単に受け取る事ができる．この機能のおかげでフォローし，フォローされる事がそのユーザーの友人関係や趣味によるつながりをTwitter上で構成している可能性が高い．しかし匿名で登録可能なTwitterでは誰がどこに属している人なのか確認する事が難しい，そのためにフォロー関係からコミュニティを俯瞰的に確認することでユーザーのコミュニティを抽出することができると考えた．

2.　研究目的

　本研究の目的は，ユーザーのフォロー関係からフォローしている，あるいはされているユーザーとのコミュニティを見つけ出す．

　さらにこの研究を行い、Twitterからコミュニティを作成する事で、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使用できる事を目指す。

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．

　この研究はフォロー関係をクラスタリングし，マッピングしたものを理想のクラスタリング結果と比べる事で評価を行っている．まずユーザーをノード，フォローの有無をエッジとし，重みをユーザー間の類似度とした重みつき有向グラフとして取り扱っており，この研究で使用されたシステムはTwitterのユーザー名を入力することでTwitterAPIが動き，入力したユーザーのフォロー集合とその関係を取得し，このフォロー集合と関係に隣接行列を用いてグラフを作成し，エッジを明確にさせ，さらにノード間の関係の強さを調べるために類似度行列を作り，重さを明確にしている．最後にこの類似度行列を用いてスペクトラルクラスタリングを行っている．なおクラスタ間のカットはMCut（Min-maxCut）を用いている．

4.　成果物のイメージ

　成果物となるシステムを使うユーザー名を入力すると，TwitterAPIを使って入力したユーザーのフォローしている相手とされている相手を取得する，この情報を使って，ユーザーの持つグループを見つけ提案を行う．

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）形式で取得する事ができた．

7.　今後の計画

　ユーザーの情報を取得する事ができたが，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