

Twitter におけるフォロー基準 合致判定支援システムの検討

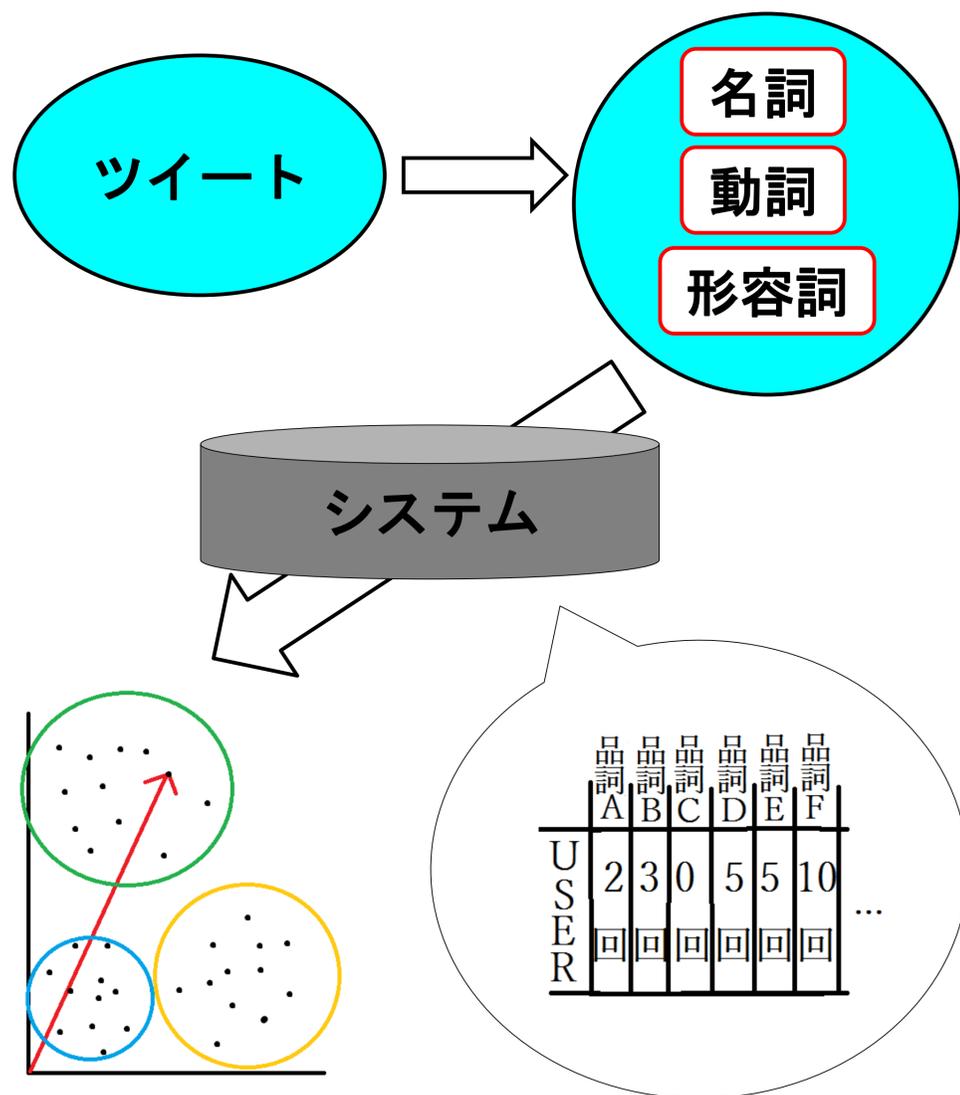
M0113363 林健輝

研究の背景

SNSの利用者増加により素性の分からない人と関わる機会が増えている。相手がどのような人か調べる場合はネットの情報しかなく、近年未成年者がネットの情報から判断できず、事件の被害者になってしまうことがある。そのため相手の特徴を提示するシステムがあれば良いと考えた。

研究の方法

あるユーザーのツイートを取得、分析して、そのユーザーの特徴を提示するシステムの検討のために、ツイートに含まれる品詞に注目した。品詞の頻出回数によってユーザーをベクトル化して特徴を出す方法を提案する。特徴を提示するために使う品詞のピックアップは、無差別に収集したユーザーのツイートに含まれる品詞からランダムに選んだ。使う品詞は名詞・形容詞・動詞で約1000単語ずつ用意する。そこからベクトル化するシステムを制作し、それに無差別で選んだユーザーを約1000人通し沢山のベクトルを作り、散らばったベクトルごとにクラスタリングする。そして気になるユーザーの1000ツイートを対象に、先ほどのシステムに通して出たベクトルに近いクラスタを提示することで特徴を提示できると考える。



進行状況

特徴を見つけないユーザーのツイートをTwitterAPIで取得するシステムを制作中である。ツイートの取得は出来るが、ツイートの取得数が少ないので、より多く取得できるようにコードを書いている。

今後の予定

判断したいユーザーのツイートを品詞ごとに抽出するシステムが完成次第、大量のベクトルをクラスタリングするシステムの制作に入る。