2017 年度卒業研究概要

黒歴史ツイート検出システムの構築

大谷　紀子　研究室
1472066　中尾太一

1. 研究の背景・目的

今日ではスマートフォンの普及により，多くの人や企業がソーシャル･ネットワーキング･サービスを利用している. 現在普及している SNS に Twitter がある. Twitter は 2006 年に開始されたサービスで，140 字の字数制限と投稿の拡散力，リアルタイム性が特徴である．Twitter での投稿のことをツイートと呼び，投稿の手軽さから日常の些細なできごとや考えを広く発信できる．発信された投稿の中には時間が経ってから見返したときに恥ずかしいと感じる投稿が含まれていることがある．投稿の内容を恥ずかしいと感じるようになってからも，投稿は削除しない限り残っているので，ユーザ指定や期間指定の検索をすることで簡単に見つけることが可能である．恥ずかしい投稿を見られないようにするには，過去のツイートを見返して削除する必要がある．しかし，総ツイート数が多い場合，ツイートを見返すことに時間と手間がかかるという問題がある．

本研究では，時間が経って見返したときに恥ずかしいと感じるツイートを黒歴史ツイートと定義する．黒歴史ツイートの削除の効率化を目的として，過去の投稿から黒歴史ツイートを自動で検出し，該当ツイートの削除を支援するシステムを構築する．

2. システム概要

本システムは Python によって開発し，CGI として動作させる．過去ツイートの取得には Twitter が提供する“Twitter Rest API”を利用した．まず，ユーザは Twitter の ID とパスワードを入力する．次に，黒歴史ツイートを検出する期間とリツイート･リプライの除外の有無を設定し，検出開始ボタンを押す．最後に，提示された黒歴史ツイートの候補をユーザが任意で選択し削除する．

システムは 3 つの方法で黒歴史ツイートを検出する．1 つ目は NG ワードを設定して黒歴史ツイートを検出する方法である．まず，取得した過去のツイートを形態素解析し，名詞と動詞を抽出する．形態素解析とは文を言語で意味を持つ最小の単位に分割して，単語の品詞などを判別する処理である．抽出した単語があらかじめ作成した NG ワードリストに含まれていた場合に黒歴史ツイート候補として提示する．2 つ目は新しいツイートの特徴と古いツイートの特徴を比較する方法である．直近のツイートの内容は現在のユーザの感覚を反映していると考えられる．したがって直近のツイートをベクトル化して特徴を抽出し，過去ツイートの特徴と比較することで現在の感覚にない特徴を持ったツイートを検出する．ツイートの特徴ベクトルの生成には Doc2Vec を用いる．Doc2Vec とは文章をベクトルに変換する機械学習アルゴリズムである．本システムでは直近 50 件のツイートの特徴ベクトルと過去ツイートの特徴ベクトルを比較し，類似度が低い順に 5 つを黒歴史ツイート候補として提示する． 3 つ目はナイーブベイズ分類器を用いて黒歴史ツイートを検出する方法である．ナイーブベイズ分類器とは，未分類のデータがどのカテゴリに属するのかを判定する教師あり学習の分類器である．マイナビが実施したアン

ケート[1][2]の結果から，回答者の多くが恥ずかしいと感じる“ポエム投稿”，“リア充アピール”，“人の悪口”，“政治に関する意見”を黒歴史ツイート候補のカテゴリとする．黒歴史ツイート候補のカテゴリを含む全 28 カテゴリの分類器を作成し，話題分類器が黒歴史ツイート候補の 4 つのカテゴリのいずれかだと判定したツイートを検出結果として提示する．

3. 評価実験

日頃から Twitter を使用している大学生 13 名を被験者として本システムを使用させ，アンケートにて，本システムで検出された黒歴史ツイートの結果，黒歴史ツイートが検出されるまでの時間，システムの使いやすさ，システムのデザインについて 5 段階で評価させる．また，検出結果中の削除したいツイートの有無，過去のツイートから恥ずかしい，または攻撃的だと感じたツイートを削除した経験の有無，削除したいと感じるツイート内容のカテゴリ，本システムをまた利用したいかについても回答させる．さらに，自由記述欄として本システムを使用して感じたことや意見を集めた．

実験の結果，肯定的な評価をした被験者は検出結果について 13 人中 11 人，動作時間について 10 人，システムのデザイン性について 12 人，使いやすさについて 11 人であった．また，過去に恥ずかしいツイートや攻撃的なツイートを削除したことがあると回答したのは 7 人，本システムをまた使用したいと回答したのは 9 人であった．削除したいと感じるツイートの内容は上位から順に“悪口”，“恋愛”，“詩”，“政治”という結果で，マイナビが実施したアンケートの結果と同様であった．被験者の意見の一部を以下に示す．

- 概ね過去の衝動的なツイートがリストアップされており，便利に感じた．
- デザインもわかりやすく、ユーザビリティが良いシステムだと思った．自分の言葉が汚いので，ばか楽しいとかの「ばか」の部分に反応してポジティブなツイートが多く検出された．
- システムによって選ばれたツイートに対するリプライ，いいね，日付などの詳細情報について，削除する前に確認できればより便利だと思った．

4. 考察

本研究では，過去のツイートの傾向や内容をもとに黒歴史ツイートを自動で検出するシステムを構築した．アンケートでは，システムの検出結果，動作時間，デザイン，使いやすさに関して肯定的な評価を得ることができた．しかし，被験者からの意見を受け，システムを改善する必要がある．ポジティブな内容のツイートが多数検出された問題については，NG ワードリストにある単語を含んだツイートをすべて検出結果として提示することが原因であると考えられる．NG ワードの有無のみを考慮するのではなく，係り受けを考慮することで問題の改善が可能であると考えられる．また，システムによって選ばれたツイートに対する詳細情報について確認できるとより便利になるという意見については，“Twitter Rest API”によって詳細情報を取得し，検出結果と合わせて表示することで実装が可能である．同様の方法で投稿に添付された画像の表示も可能になる．

参考文献

[1] マイナビニュース，“【女性編】「これは恥ずかしい」と思う他人の Twitter 投稿ランキング”，https://news.mynavi.jp/article/20140828-a178/

[2] マイナビニュース，“【男性編】「これは恥ずかしい」と思う他人の Twitter 投稿ランキング”，https://news.mynavi.jp/article/20140823-a062/