

Twitter のツイートの自動タグ付けを行うライフログシステムの提案

1185036 小林 拓夢

(指導教員 速水 治夫 教授)

1. はじめに

ライフログとは、毎日の出来事などをデジタルデータに記録することである。近年、日本国内でもマイクロブログと呼ばれる Twitter が普及してきた。この Twitter に投稿されるツイートもライフログに活用できると考えられるため「Twitter のツイートを利用したライフログシステム」を試作し評価実験を行った。評価は良好であったが、「膨大なツイートの中から有用なツイートだけを容易に探し出したい」、「ツイートへのタグ付けを自動で行ってほしい」という意見が多かった。膨大なツイートを見返すときにはツイートを分野別に閲覧することが大切である。例えば「食」とタグ付けされたツイートを一覧表示することで、ユーザが過去に食べた料理などを時系列順に一覧で閲覧することができ、今晚の夕食の献立など今後の行動に役立てることが出来る。

そこで、取得した Twitter のツイートを自動でタグ付けできるライフログシステムを提案する。

2. Twitter のツイートを利用したライフログシ

ステムの提案^[1]

先行研究において Twitter のツイートをライフログとして活用するシステムを試作した。本研究は、このシステムを大幅に、改善・機能追加したことになる。先行研究で試作したシステムは Twitter のツイートを取得し、ライフログとして過去のツイートを閲覧できる Web アプリケーションである。

(1) ツイート取得部

TwitterAPI を使用し、Twitter からツイートやユーザ情報などを取得する。自動取得されるツイートは、ユーザ自身のツイートと、ユーザのメンションなど「ユーザに関連のあるツイート」である。その他ユーザが残しておきたいツイートを手動で取得することもできる。

(2) ツイート閲覧部

カレンダー機能をはじめとし、ツイートをライフログとして閲覧しやすいように表示する。ツイート単体には、左側にツイートしたユーザのアイコン画像、ツイートしたユーザ名とスクリーン名、宛先のユーザ名、日付時刻などを時系列順に表示させることで、いつ誰が投稿したのか確認しやす

くなっている。

また、取得されたツイートの検索機能も設け、取得した全てのツイートから指定したテキストで検索や、各利用ユーザで利用可能なツイートのタグ機能もある。

3. 提案システム

本システムは、ツイート取得部、ツイート閲覧部、ツイートの自動タグ付け部からなる。ツイート取得部とツイート閲覧部の一部は先行研究のシステムを改善したものを使用する。

3.1 ツイートの自動タグ付け

本システムでは、ツイートの自動タグ付けを行うために株式会社はてなの「はてなキーワード」を使用する^[1]。はてなキーワードを使用することで、ツイートへ 17 種類のカテゴリのタグを付与することができ、膨大なツイートの学習を行う必要もない。図 1 にツイートの自動タグ付け概要を示す。このツイートの例では、ツイート本文内にある「コマンドー」といったキーワードに対して「映画」がカテゴリタグとして登録される。

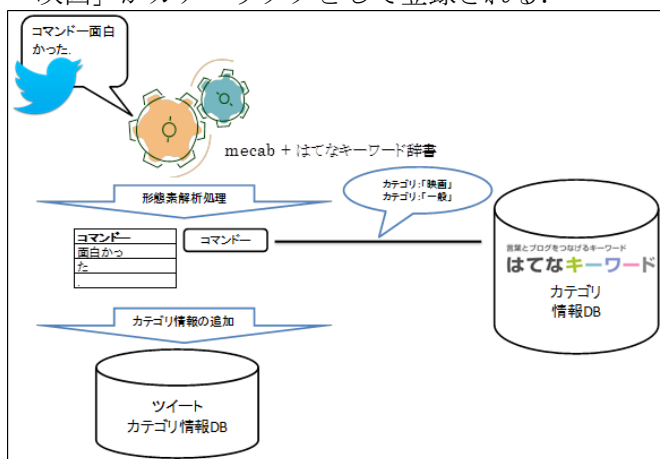


図1 ツイートの自動タグ付け概要

3.2 タグ付けされたツイートの閲覧

タグ付けされたツイートには、カテゴリタグとキーワードがツイート下部に表示される。図2にシステムのメイン画面、図3に自動タグ付けされたツイートを示す。このように、カテゴリタグとタグ付けされたキーワードを表示することでツイート本文内のどのキーワードがタグ付けされたのか分かる。



図2 システムのメイン画面

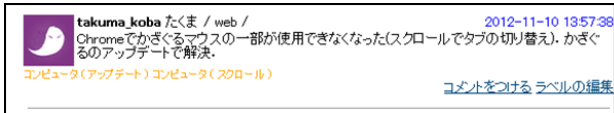


図3 自動タグ付けされたツイート

4. 評価実験

ツイートの自動タグ付けの正確性の評価と、6名の実験協力者による評価を行った。

4.1 自動タグ付けされたツイート数

全体の約22%のツイートがタグ付けされた。表1にツイート総数と自動タグ付けされたツイート数を示す。また、ツイート単体の文字数が多いほどタグ付けされやすいことが分かった。

表1 自動タグ付けされたツイート数

取得されたツイート総数	35629
自動タグ付けされたツイート数	8002

4.2 ツイートの自動タグ付けの正確性

500個のツイートをランダムで表示し、タグを付与すべきツイートの場合筆者がタグを指定し、それを「正解のタグ」とした。図4に自動タグ付けの正確性を示す。67%がシステムと人手で同一のタグ付け・タグ不要判定したが、33%のタグに対しては正解のタグと違うタグが付けられた。また、「タグ付け不要なツイートに、誤ってタグ付けした」の中には評価実験を行った筆者には理解できなかったが、発言者には理解できるキーワードが含まれていると考えられるため、23%より少なくなると考えられる。

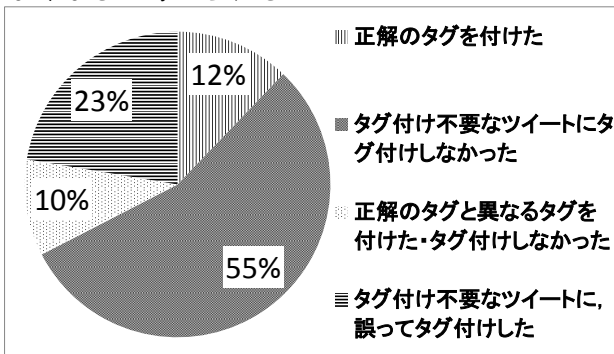


図4 自動タグ付けの正確性

4.3 自動タグ付けされたツイートのうちライフログとして有用なものの調査

実験協力者に任意の3つのタグ選んでもらい、6つの設定方法で検索してもらった。そして、検索結果総数と、総数のうちライフログとして有用なものをカウント(総数50以上のものは50ツイートまで)した数を記録した。

リンクスコア0設定で平均46%と、自動タグ付けされたツイートの半数近くがライフログとして有用だということが分かった。また、リンクスコアの数値を上げたり、2つ以上の同カテゴリあり設定にすると、有用なツイートが頻出するが、検索結果に出現するツイートの数は減る。

4.4 過去のツイートの検索所要時間の計測・比較

実験協力者にTwitter本家と本システムを使って、過去のツイートを探し出して目的のツイートが見つかるまでの時間を計測した。Twitter本家では目的のツイートが見つかるまで平均1分23秒かかっているが、本システムでは平均12秒と平均71秒早く検索することができた。表2に検索所持用時間を示す。

表2 検索所要時間

	本システム	Twitter 本家
検索所要時間平均	00:12.2	01:23.0
検索所要時間最大	00:29.6	04:56.4

5. おわりに

評価実験から、全体の67%のツイートを正確にタグ付け・タグ不要判定することができ、タグ付けされたツイートのうち、ライフログとして有用なツイートは約46%で、検索時にリンクスコアの数値などを上げることで更に有用なツイートを絞って検索することができることが分かった。また、過去のツイートを探し出すときにもTwitter本家で検索するよりも平均71秒早く検索することができ、過去のツイートを探し出すのに適していることが分かった。

今後は、正確にタグ付けされなかったツイートに対する処理の改良・改善や、ツイートの表示方法の改良、より詳細な評価実験を行いユーザの趣向やツイート数でのおすすめ検索方法を提示するなどの機能を追加していきたい。

参考文献

- [1] Twitter のツイートを利用したライフログシステムの提案, DICOMO2011 シンポジウム, pp.1265-1270(2010.7)
- [2] はてなキーワード, <http://d.hatena.ne.jp/keyword/>