

Twitter上の多様な情報を統合利用するための情報ダッシュボードの構築

北田 哲也[†]立命館大学大学院情報理工学研究科[†]高田 秀志[‡]立命館大学情報理工学部[‡]

1 はじめに

Twitterは個人のつぶやき以外に企業広告、天気予報、ニュースなど、様々な目的で利用されている。また、Internet of Thingsの発展により、電子機器が自律的にTwitterに投稿するようになっている。

Twitterはツイートをタイムラインとして表示する機能を持ち、ユーザはフォローしたアカウントのツイートを時系列順に閲覧することができる。しかし、Twitterが様々な目的で利用されるにつれてフォローしているアカウントが増加すると、タイムライン上に大量のツイートが表示されてしまい、どれが有用な情報なのかの判別が困難になるといった問題点がある。Twitterが用意している手段であるリストを利用する方法もあるが、ユーザが閲覧したいアカウントを手動で振り分けなければならない。

そこで本稿では、ツイートを分類してカテゴリ化し、情報の有用性を容易に判別できるような情報ダッシュボードを提案する。

2 Twitterにおける自然言語処理

以前から自然言語で書かれた内容をコンピュータで処理する研究は数多く存在し、Twitterにおいても自然言語処理を用いてツイート内容を解析する研究が行われている。

例えば、近藤らの研究[2]では、はてなキーワードを利用してツイート文からユーザの特徴推定を行っている。また、久米らの研究[3]では、はてなキーワードAPIを用いて、ツイート文に含まれる単語のカテゴリ情報を取得し、ユーザの興味領域の計算に利用している。

3 情報ダッシュボード

3.1 カテゴリ分類

ツイートのカテゴリは、ツイートの内容、あるいは、ツイートのアカウントによって判別する。ツイート内容による分類には、はてなキーワード自動リンクAPI[1]を用いる。はてなキーワード自動リンクAPIは、文章内にはてなキーワードに対応する単語が存在する場合にリンクタグを挿入するAPIである。このAPIにはリンク対象となる単語データを抽出する機能が搭載されている。抽出された単語データには単語名の他にカテゴリ情報とリンクスコアが含まれる。カテゴリ情報は一つだけでなく複数含まれる場合もある。リンクスコアは単語の文字数が多い場合や、単語が固有名詞である場合に高くなる傾向がある。

本手法ではこれらの機能を用いて、カテゴリ分類を行う。基本的な手順として、まず、Twitterサーバからツイートを取得する。次にツイートを上述のAPIに送信して単語データのリストを受け取る。単語データのリストの例を表1に示す。次に、カテゴリごとにリンクスコアの合計を求め、最も高いリンクスコアの合計をそのツイートのカテゴリとして設定する。表1の

データに対してリンクスコアの合計を求めたものを表2に示す。この場合のカテゴリは「ゲーム」になる。

はてなキーワードが付与するカテゴリの一つに「一般」がある。「一般」カテゴリと他カテゴリを含んでいる単語は一般的な意味で使われているのか、カテゴリ特有の意味で使われているのかを判断できないため、同じカテゴリの単語が他に存在する場合のみリンクスコアの合計に反映する。また、「はてな」や「地理」など分類にふさわしくないカテゴリは対象外とする。

表1: 単語データのリストの例

単語	カテゴリ	リンクスコア
テニス	スポーツ	15
ゲーム	スポーツ, ゲーム	16
ファミコン	ゲーム	62

表2: カテゴリごとのリンクスコア合計の例

カテゴリ	リンクスコア合計
スポーツ	31
ゲーム	78

3.2 ツイートの提示

カテゴリ分類したツイートを提示する手段として情報ダッシュボードを構築する。ここでのダッシュボードとは原義の自動車の計器盤のように、一目で複数の情報が得られるものを想定している。

情報ダッシュボード(図1)はツイートを表示する複数のパネルとメニューバーで構成される。パネル(図2)はカテゴリ情報を持ち、対応するツイートを表示する。また、視認性の向上のために色分けや文字サイズの変更を行うことができる。メニューバーはAddボタンとUpdateボタンを持つ。Addボタンを押すと、パネルが情報ダッシュボード内に追加される。複数のパネルが追加されると領域が分割される。Updateボタンを押すと、パネルのツイートが最新のものに更新される。これによりユーザは情報ダッシュボードを通してカテゴリ分類されたタイムラインを閲覧できるようになる。

4 実装

本情報ダッシュボードを実現するシステムの構成要素を図3に示す。サーバ部ではカテゴリ分類機能を主に実現し、アプリ部では情報ダッシュボードの操作機能および閲覧機能を実現する。

4.1 サーバ部の構成

サーバにおいては、一定時間おきにツイートをtwitterサーバから取得しデータベースに保存する機能と、サーバのデータベースに保存されているデータをアプリで解読できるように変換して出力する機能が実装されている。実装に使ったサーバ環境はさくらサーバであ

An Information Dashboard for Integrated Use of Various Information Sources on Twitter

[†]Tetsuya Kitada [‡]Hideyuki Takada

[†]Graduate School of Information Science and Engineering, Ritsumeikan University

[‡]College of Information Science and Engineering, Ritsumeikan University

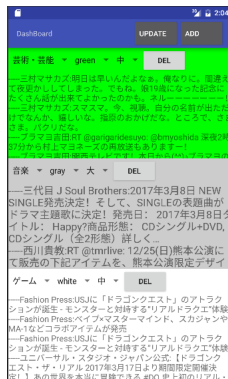


図 1: 情報ダッシュボードの外観

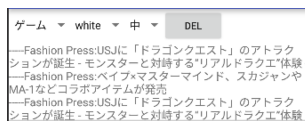


図 2: パネルの外観

る。開発言語には Python を使い、外部ライブラリとして tweepy を導入している。また、データベースには MySQL を使用している。

サーバのプログラムはツイート収集 CGI と出力用 CGI で構成される。ツイート収集 CGI は定期的に自動で動作し、Twitter サーバからツイートを収集してカテゴリ分類を行い、結果をデータベースに保存する。出力用 CGI は、アプリ部からリクエストが行われた時に実行され、データベースからツイート収集 CGI で保存したツイートやカテゴリ分類結果を XML に変換して出力し、アプリ部に返す。

4.2 アプリ部の構成

情報ダッシュボードを提示するアプリケーションは Android 向けに提供される。統合開発環境として Android Studio を利用している。パネルや情報ダッシュボードは ViewGroup を用いて構成され、カテゴリ分類結果の表示に使用される。

5 評価実験

構築した情報ダッシュボードの利便性を検証するために行った評価実験について述べる。

5.1 実験内容

被験者は 6 人であり、情報ダッシュボードを利用した後に、使い勝手をアンケート形式で回答した。本実験で用いられるツイートは実験用に取得した専用のアカウントから収集している。被験者は、次の項目に対して五段階評価で回答し、任意で評価の理由についても回答した。

質問 1 操作しやすかったかどうか (1: わかりにくい ~ 5: わかりやすい)

質問 2 構築した情報ダッシュボードの UI やレイアウトはよかったか (1: 悪い ~ 5: 良い)

質問 3 ツイートがカテゴリ化されていることが特にカテゴリ化されてないタイムラインと比べてわかりやすいと感じたか (1: わかりにくい・効果がなさそう ~ 5: わかりやすい・効果が有る)

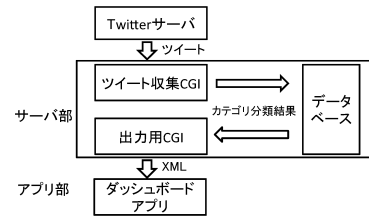


図 3: 実装

質問 4 カテゴリに表示されるツイートが適切であったか (1: 適切でない ~ 5: 適切である)

5.2 結果と考察

アンケートの結果は表 3 の通りである。

表 3: アンケート結果

質問番号	評価				
	1	2	3	4	5
質問 1	0 人	1 人	1 人	4 人	0 人
質問 2	1 人	2 人	1 人	1 人	1 人
質問 3	0 人	0 人	0 人	6 人	0 人
質問 4	0 人	0 人	1 人	4 人	1 人

評価の理由として、質問 1 では「操作自体は単純である」といったものがあつた。質問 2 では「テキストだけで見づらい」という意見が複数見られた。質問 3 では「カテゴリ分けされると知りたいカテゴリを中心に情報が収集できて効率が良い」との答えや、「カテゴリ化されていると何の情報なのかすぐにわかる」といった答えがあつた。質問 4 では「重複したツイートが一部あつた」との意見が得られた。

質問 2 の情報ダッシュボードの UI に関しては評価が低めであるため改善が必要である。質問 3、質問 4 のカテゴリ分けに関する質問には高い評価が得られている。

これらの結果から、ツイートをカテゴリ分類することは情報の有用性を容易に判別できるようになり、利便性につながっていると考えられる。

6 まとめ

本稿では、ツイートを分類してカテゴリ化し、情報の有用性を容易に判別できるような情報ダッシュボードを提案した。評価実験によって情報の有用性を容易に判別できることが示された。今後は、アプリケーションの操作性と視認性をより高める必要がある。

参考文献

- [1] はてなキーワード自動リンク API: <http://developer.hatena.ne.jp/ja/documents/keyword/apis/autolink>
- [2] 近藤 直人, 内田 理: Twitter を用いた個人の特徴抽出とその情報提供ポータルサイト構築への応用, 言語処理学会第 20 回年次大会, pp.586-589
- [3] 久米 雄介 打矢 隆弘 内匠 逸: 興味領域を考慮した Twitter アカウント推薦, マルチメディア、分散、協調とモバイル (DICOMO2014) シンポジウム, pp.627-632