

Twitter 上で任意の検索語句に対するネガポジ度を判定し 可視化するアプリケーションの開発と研究

村上奈緒^{†1} 尼岡利崇^{†1}

近年普及してきたマイクロブログと呼ばれるサービスの中でも、Twitter は世界的に大きな人気を博している。即時性のある無加工の情報群は、評価情報の収集や感情分析などに非常に役に立つ。しかし膨大な情報群のため、文字のみでの表示では煩雑になる。そこでまず始めに、一目でおおまかな評価情報を得るために、指定した検索語句に対するネガポジ度を解析し、分類わけをした。本研究では、評価情報とシンプルなキャラクターの棒グラフ状の可視化を組み合わせることで、情報可視化とエンタテインメント性を融合した手法を提案し、見やすく親しみやすい評判分析ツールの開発をする。

Development and research of application for visualizing the negative-positive degree of any search phrases on Twitter

NAO MURAKAMI TOSHITAKA AMAOKA

Among the service called micro blogging have become popular in recent years, Twitter has gained great popularity worldwide. The group of real-time raw data is very useful for the Sentiment collection and analysis. But Big data in the text-only view become cumbersome. Therefore at first, analyzed and do classification a negative-positive degree for the search phrase to obtain at one view evaluation. In this study, we combine big data analysis with simple character shaped bar chart visualization to propose the technique of information visualization with entertainment value, and develop the sentiment analysis tool more visible and friendly.

1. はじめに

近年、Twitter の使い方は、マイクロブログにとどまらず、多様化してきている。例えば、NHK の NEWS WEB[1] というニュース番組では、#nhk24 というハッシュタグのついたツイートの数を可視化する「つぶやきカウンター」やその日つぶやかれたツイートに含まれている単語をランキング化した「つぶやきビッグデータ」というコンテンツがある。また、ファミリーマートでは、Twitter 上でアイデアを募集し、実際に商品として売り出すという手法[2]や、パズル&ドラゴンズでは公式アカウント宛にツイートを送ってアンケートをとる[3]など、最近では、Twitter はユーザーの声を集めるために利用されている例が多い。

テレビを見ながら意見をつぶやく、イベント会場や話題の店に足を運んで、その場で感想やレポートをつぶやくなど、投稿されたツイート群には評価情報が豊富に含まれている。そのため、Twitter は意見抽出や感情分析の情報源として有効であることが検証されている[4]。

このように、Twitter の使用方法が多様化していく中で、意見、評価の投稿という利用方法が広がっているにも関わらず、それらを手軽に閲覧できるアプリケーションはまだまだ少数である。

Twitter は、ユーザーの生の声そのまま投稿されている

場合がほとんどで、加工されていない情報を知るには適しているツールである。そこで、Twitter から情報を収集し可視化することで、加工されていない意見や評価を知る機会が増えるのではないかと考えた。また、内閣府の調べによると、日本におけるスマートフォンの世帯普及率は、2014年4月現在で54.7%、タブレット端末は20.9%となっている。このような端末の普及は、Twitter 等のマイクロブログサービスの人気の要因の一つとなっている。

そこで本研究では、手軽さとユーザーの使用率の高さから、スマートフォン上で動作する、評価情報収集のアプリケーションを作成する。評価情報の最上位層には、ネガティブ度、ポジティブ度（以後、ネガポジ度とする）の2つを設定する。これにより、ユーザーは自身の検索した語句に対する意見がポジティブなものか、ネガティブなものかの評価がわかる。さらに、初めにネガポジのどちらかを選択できることから、必要な情報を効率よく収集できると考えた。また、自分のもっている意見は世間では少数派なのか、多数派なのかを知ることが可能となる。

可視化手法には、エンタテインメント性を与えることで、ユーザーが情報収集だけでなく、Twitter から抽出した情報に興味を持って接することが出来る工夫を行った。オブジェクトをイラスト調にし、軽い動きをつけ、画面を明るくカラフルにするなど、かわいらしくシンプルなデザインにすることで、大量の情報をユーザーに見やすく使いやすいアプリケーションになるようにする。

^{†1} 明星大学 情報学部
School of Information Science, Meisei University

これらによって、個人が発信する生の情報へのアクセスを容易にし、手軽に評価情報の収集を可能とする。また、情報検索の結果をエンタテインメント性をもって可視化するアプリケーションの制作を行う。

2. 先行研究

2.1 感情分析に関する研究

ヤフージャパンが提供する、つぶやき感情分析[5]は、Twitter のリアルタイム検索を用いた感情分析ツールである。図 1 に示すように、ユーザーが検索ボックスに語句を入力し検索すると、Twitter からその語句を含むツイートを集集し、ストリーミング形式で表示する。



図 1 リアルタイム検索結果[5]

右上には時間帯別のツイート数、右中にはネガポジの割合、右下にはトレンド語句が表示される。なお、スマートフォンで動作する、「リアルタイム検索」というアプリケーション版も発表されている。アプリケーション版では、Web 版の機能に追加し、よく検索する単語を登録することで、検索中に話題に大きな変化があった時に通知することができる。さらに通常の検索と同様に、表記揺れに対応しているため、取得漏れが少ない。

2.2 感情の可視化に関する研究

図 2 で示す We feel[6]は CSIRO, The Black Dog Institute が共同開発した、Twitter に投稿されたツイートの感情を、グラフ化と地球儀へのマッピングで可視化した Web サイトである。左は各感情の割合を示しており、中央は時間別のツイート数を表している。どちらも、色分けされた感情の部分をクリックすると、その感情語の中で更に細分化された語句とグラフが表示される。

感情語は surprise, joy, love, sadness, anger, fear, other が設定されているが、それぞれ 3 つずつ階層をもつ等、細かい分析が可能になっている。右の地球儀では、地域を設定できるようになっている。色の濃いエリアほどツイート数が多くなっている。図 3 の下の棒グラフには男女が表示され、こちらもグラフに反映させることができるが、ユーザー入力 of データに基づき表示される項目であるため、不

明である割合が多い。また、table builder という機能で、期間や感情、地域などを指定して、時間と数値の表を作成するモードもある。

しかし、これらのサービスでは、日本語/英語のどちらかにしか対応していない、関連語句単体での再検索ができない、表示する情報量が多すぎて見にくいという点がある。

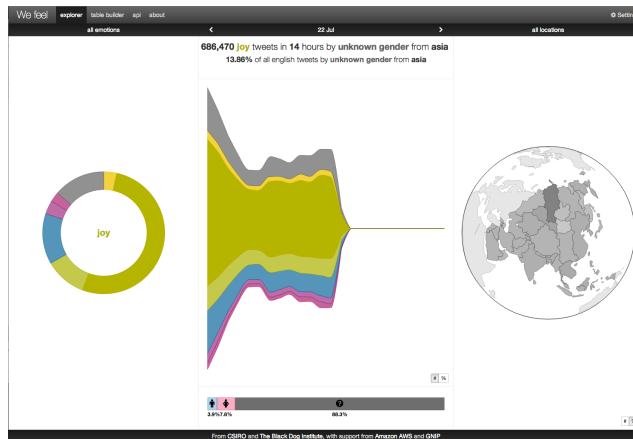


図 2 We feel 初期画面[6]

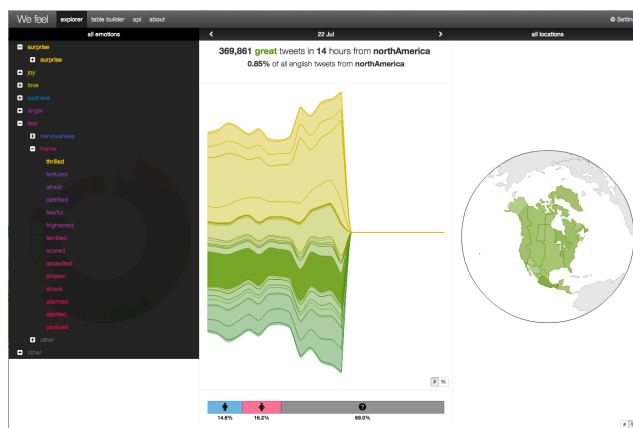


図 3 We feel 感情・地域選択後画面[6]

2.4 本研究の特徴

そこで、これらを改善、実装したスマートフォン向けアプリケーションを、本研究では作成する。一画面で多くの情報を表示させるのではなく、ユーザーに選択を委ね、詳細画面を表示させる。検索予測ではなく関連語句で検索することにより、ユーザーは検索結果を得られるだけでなく、検索語句に関連する新たな語の発見を可能にする。また、日本語/英語での検索の切り替え機能を実装する。これにより、同じ検索語句に対し、異なる 2 つの言語圏における、ネガポジ度の差異を知ることが可能となる。

この 3 点を実装することによって、感情分析と可視化、それぞれを活かした情報検索ツールを提案出来ると考える。

3. 研究概要

3.1 可視化およびインタラクティブ性について

まず始めに、図4の初期画面について説明する。ユーザーは③の検索ボックスに検索語句を入力する。初期画面では、他には特に設定することはなく、検索結果が表示される。①はネガポジ度を合計100%とした時の割合を示した棒グラフで表示している。この棒グラフにタッチパネルを介し触れると生き物のようにふるふると動くようになっており、ユーザーに可視化ツールだけでなく、エンタテインメント性を持たせた癒し系ゲームのような感覚で情報に関わる方法を提供している。②は収集したツイート群で、アイコン、ユーザー名、ユーザーID、ツイート内容が含まれている。画面上部には④その他設定エリアがあり、そのエリアの選択項目により、日本語/英語の切り替えや、マニュアル表示ができるようになっている。

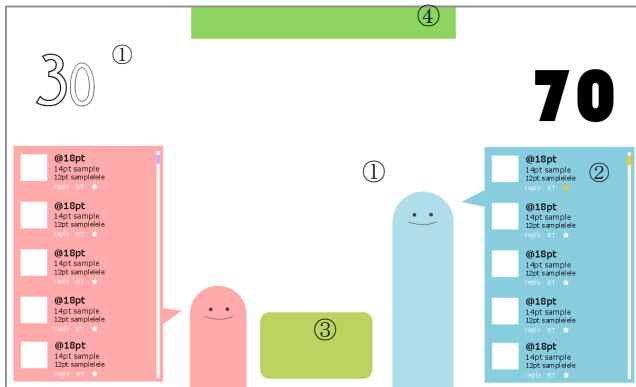


図4 初期画面 (①ネガティブ度・ポジティブ度の割合、②ツイート群、③検索ボックス、④その他設定)

なお、検索ワードを入力する前には、自動的にトレンドワードを取得し、その結果が表示されるようになっている。表示後、①のネガティブ度・ポジティブ度の割合を示す棒グラフを選択すると、画面が図5に切り替わり、詳細が表示される。

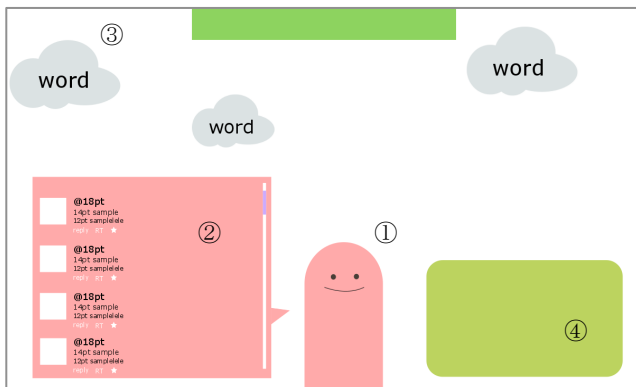


図5 詳細画面 (①ネガティブ度 (またはポジティブ度) の割合、②ツイート群、③関連語句、④使用されている感情語)

図5の③の雲のオブジェクトは関連語句を表示するグラフィックスであり、選択すると、語句に関するネガポジ度を再検索し、図4のように表示する。これにより検索語句と関連語句を比較することが可能となり、連鎖的に関連語句の検索が出来ることにより、評価情報の質の向上に役立つ。④は、それぞれのツイート群で、ネガポジ度の評価に使用する感情語のうち、どのような感情語が使われているのかの内訳を表示する。検索語の設定後、割合を可視化することで、調べた語句の評価情報を一目で分かるようにしており、そこから詳細画面に移動し、内訳や関連語句など、追加情報を閲覧する。図6は、ユーザー操作の一連の流れを示している。ユーザーは、検索語句の入力、閲覧、選択という3つの操作のみで評価情報を調べることができることから、簡単な操作によりTwitterの可視化情報が得られるアプリケーションを実現しているといえる。

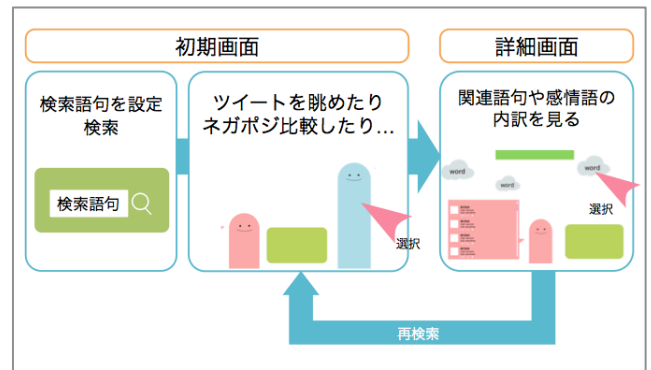


図6 ユーザーの操作のイメージ

3.2 オブジェクトと画面のデザイン

本研究では、キャラクターを用いた可視化をおこなっている。これは、キャラクターを用いることにより、単に可視化ツールというだけでなく、エンタテインメント性を与えることでユーザーが情報に親しみやすく接することができる考えたためである。また、これらキャラクターを赤(ポジティブ)と青(ネガティブ)に色分けして表示することで、直感的にネガポジをとらえられるようにし、一目で情報の傾向をとらえられるように工夫している。さらに、表示にはJavascriptのsvg描画ライブラリのRaphaelを使用し、棒グラフの生物的な動きを実現している。

3.3 ネガポジ度の分析および関連ワードの抽出方法

本研究は、PhoneGap[b]を使用し、スマートフォン用アプリとして実装を行う。使用APIは、Twitter API、ネガポジAPI[c]、キーフレーズ抽出APIの3つである。まず始めにTwitter APIを用いて、検索語句を含むツイート群を収集し、収集したデータをネガポジAPIにより、ネガポジ度を評価する。次に、キーフレーズ抽出APIにより、関連ワードの

a) HTML5+CSS+JavaScriptで書くクロスプラットフォームなアプリの開発環境のこと。

b) (株)メタデータが提供する感情解析API

抽出を行う。ここでは、ネガポジ分析と関連キーワード分析についての詳細について述べる。

3.3.1 ネガポジの分類

(1) ネガポジ分析

ある事物の評判、レビューを知りたい際、感情分析が行われるが、その中でもよく使われているのがネガポジ分析である。ネガポジという区分である程度大まかに分類することで、ユーザーは最初に調べたい情報の評価の高低を選択でき、Twitter上の評価情報の閲覧と整理がしやすくなる。本研究ではTwitterを利用するため、データ量は膨大である。そこでネガポジ分析をおこなうことで、大量の情報を簡単にまとめ、ユーザーが欲しい情報、本研究では検索語句に対する評価情報をシンプルに表示させることにした。検索語句に対するTwitter上の評価が、よい評価なのか悪い評価なのかだけを知りたい場合、初期画面のみで欲しい情報が手に入るという視認性の良さやシンプルな結果表示の両方を実現している。さらに、ネガポジをメインに表示し、そのシンプルな表示方法を用いたことで、関連語句との比較が容易になるという利点もある。

また、Twitterは「自分のことをつぶやき、共感／議論し、つながるツール」だといえる。本研究を利用して、自分と同じ感情の人がどの位いるのかを調べる、また、異なった感情をもつ人の意見を知ることで、多視点から一つの事物をとらえることが可能となる。さらに、お気に入り登録やリツイート、発言者をフォローする、等の使い方もでき、Twitterの拡張機能としての役割も果たせると考えている。特に、本研究で自身の意見が多数派なのか少数派なのかを知ることは、世論調査が頻繁に行われる政治問題や社会問題等といったトピックスだけでなく、より広範囲な目的での使用が可能であると考えられる。例えば、ある言動に対して、「あなたはアリ／ナシどちらですか」などといったような企画やアンケートなどもこれまで行われている。最近では、ユーザーが投稿したお題に対してネガポジの投票をおこなう“FLAG”というアプリケーションも発表されている。このように、ネガポジ分析はエンターテインメントの分野でも使用されており、Twitterというユーザー同士が簡単につながることが出来るサービスでは、このような使い方も、ゲーム性があり有用ではないかと考えた。このようなTwitterの使い方における可視化方法として、本論文の提案手法が有効であると考えられる。

(2) ネガポジ API

ネガポジの判定には、(株)メタデータの提供する高精度ネガポジAPIを使用する。ネガポジAPIは、文章に7段階のネガポジ属性 excellent(素晴らしい), very good(非常に良い), good(良い), neutral(中立), no good(良くない), bad(悪い), very bad(非常に悪い)]を、数万以上の形容表現・価値判断表現類から付与し、-3.00 から+3.00 までの 601 段階で判定す

る API である。

このAPIを利用し、収集したツイート群のツイートを分析する。7段階のうちの neutral を除く、very bad, bad, no good をネガティブ、good, very good, excellent をポジティブなツイートとし、二つにまとめ、図4の①を表示するための値に変換する。また、ツイート内でそれぞれにどのような感情語が使われているのかも同APIで抽出が可能であることから、使用されている感情語の処理結果を取得し、図5の④で使用する。分類されたツイート群は保存し、関連ワードの抽出や、ツイートの表示に使用する。

3.3.2 関連ワード抽出

関連語句の抽出にはヤフージャパンが提供するキーワード抽出APIを使用する。ネガポジそれぞれのツイート群をキーワード抽出APIで分析すると、ツイート群の中で多く使われる重要語句を最大20単語まで抽出可能である。しかしここで使用するツイート群は、検索してから保存されたものである。よって、検索してしばらく経ち、ある程度のツイート数になってからでないといふ重要語句の判定材料が少なく、うまく抽出ができない可能性がある。抽出した関連語句は、図5の③で使用する。

図7では、分析して振り分けられたツイート、割合、関連語句、使用語句をPhoneGap内で受け渡し、図5、6のように可視化するまでの処理を示している。

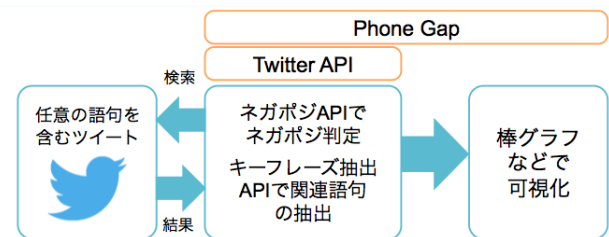


図7 処理の流れ

また、関連語句で再検索された場合も同様の処理をする。

4. おわりに

本研究では、Twitterにおける検索語句に対する評価をネガポジ度で可視化するシステムを提案した。評価をネガポジという2つの大きな評価基準により可視化することで、評価結果を一目でわかるようにした。また、ツイート情報の表示、関連ワードの表示ならびに関連ワードでの再検索等、検索語句から連鎖的に語句の評価を得られる仕組みを提案した。さらに、棒グラフをキャラクター化し、可視化方法にインタラクティブ性を与え、生物的な動きを付与したことで、エンターテインメント性を持った情報可視化方法を提案し、アプリケーションを作成した。

今後の改善点としては、ネガポジ分析の際にネガポジAPIを利用しているが、呼び出しの回数が100回/日であるため、代わりに単語感情極性対応表[7]や、日本語評価極

性辞書[8]を利用することも検討している。単語感情極性対応表は約 5.5 万語の日本語の感情極性値を-1~1 で表示し、英語版は約 8.8 万語を収録している。日本語評価極性辞書は、用言編と名詞編があり、ネガポジ、評価、経験の 4 項目で分類分けされている。しかし、どちらも活用形に対応していない部分があるためツイートの形態素解析が必要、収録語句が膨大であるなどの理由から、処理に時間がかかる、精度が落ちる等が予想される。

また、キャラクターに動きをつけるだけでなく、キャラクターと情報、さらにユーザーの全てがつながるような仕掛けを付与し、ユーザーがより検索したくなるような作りに発展させる必要がある。

将来的には閲覧のみではなく、収集・共有することで、Twitter の強みを活かせると考え、Twitter のアカウント認証をし、お気に入り登録やリツイートなどの Twitter 連動の機能も追加したいと考えている。他にも、位置情報やタグなど、Twitter には利用可能な情報が含まれるため、可視化方法にこれらの要素を取り入れることで、さらに本研究を発展させられると考えている。

参考文献

- 1) NHK NEWS WEB <http://www3.nhk.or.jp/news/newsweb/about.html>
- 2) ファミリーマート みんなで作るおむすび選手権
<http://www.family.co.jp/campaign/omusubi2013/>
- 3) ガンホー パズル&ドラゴンズ Twitter 連動アンケート
http://mobile.gungho.jp/news/pad/140704_anke.html
- 4) Alexander Pak, Patrick Paroubek “Twitter as a Corpus for Sentiment Analysis and Opinion Mining” In LREC2010, pp.1320-1326, 2010
- 5) Yahoo! Japan リアルタイム検索 <http://search.yahoo.co.jp/realtime>
- 6) CSIRO, The Black Dog Institute We feel...<http://wefeel.csiro.au/>
- 7) 高村大也, 乾孝司, 奥村学
“スピンモデルによる単語の感情極性抽出”, 情報処理学会論文誌ジャーナル, Vol.47 No.02 pp. 627--637, 2006.
- 8) 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一. 意見抽出のための評価表現の収集. 自然言語処理, Vol.12, No.2, pp.203-222, 2005.
東山昌彦, 乾健太郎, 松本裕治, 述語の選択選好性に着目した名詞評価極性の獲得, 言語処理学会第 14 回年次大会論文集, pp.584-587, 2008.