Twitter ユーザの印象選好の可視化

鈴木 智也 * 熊本 忠彦 *

†, ‡千葉工業大学 情報科学部 情報ネットワーク学科 〒275-0016 千葉県習志野市津田沼 2-17-1 E-mail: ‡ kumamoto@net.it-chiba.ac.jp

あらまし Twitter には、面識のない人でも気軽にフォローできるという特徴があり、他者とつながるための手段として有用と言える。しかしながら、その一方で、匿名の人からフォローされたり、すでにフォローしているユーザのリツイート等により、知らない人のツイートを目にしたりする機会も多く、そのような人たちをフォローすべきか否か迷うことがある。そこで本論文では、フォロー候補となるようなユーザの印象選好(普段、どのような話題の、どのような印象のツイートを見たり投稿したりしているか)を可視化することで、フォローの是非の判断を支援するシステムを提案する。本システムを用いることで、有益な情報源かどうかといった単なる損得勘定だけでなく、自身の感性に合ったユーザかどうかという視点からの判断も可能となる。

キーワード Twitter, ユーザモデリング, 印象マイニング, Sentiment Analysis, つながり構築支援

1. はじめに

近年,スマートフォンやタブレット端末の普及により,いつでもどこでもインターネットに接続できるようになった.その影響もあり,Twitterや Facebookといった様々な SNS(Social Networking Service)が盛り上がりを見せている.特に,Twitterは,友人・知人だけでなく,芸能人や有名人といった面識のない人とも気軽につながることができるという点において,他のSNSよりも優れており,2012年第4四半期のアクティブユーザは約3億人とも言われている[1].

Twitter では多くのユーザによって日夜様々な話題のツイート(つぶやき)が投稿されており、その話題は政治・経済から身の周りの出来事まで多岐にわたる。このような情報を日常的に得るためには、その発信元であるユーザをフォローする必要があるが、そのユーザが普段どのようなツイートをしているかは、実際にそのユーザの(少なくない量の)ツイートを精読してみないとわからない、中には、悲しいことばかりをつぶやいている人もいれば、怒ってばかりの人もおり、そのような人たちのツイートは見たくないと思っている人も多い。

そこで本論文では、フォローしても良いか、しない方が良いかの判断を支援することを目的として、フォロー候補となるユーザの印象選好(そのユーザが普段、どのような話題の、どのような印象のツイートを見たり投稿したりしているか)を可視化する手法を提案し、Webアプリケーションシステムとして実装する.

具体的なシステムの処理としては、本システムの利用者が任意の Twitter ユーザ名を入力すると、システムはそのユーザが投稿したツイート群 (ユーザタイムライン) とそのユーザがフォローしているユーザらが投稿したツイート群(ホームタイムライン)を Twitter API [2]を用いて収集し、著者らが開発した印象マイニング

システム[3]を用いて各ツイートの印象を数値化するとともに、それぞれのツイート群で特徴的なキーフレーズを Yahoo!キーフレーズ抽出 API[4]を用いて抽出する.システムは以上の結果を GoogleChartTools[5]を用いて可視化し、利用者に提示する.なお、印象マイニングシステムは、入力された文章を「楽しい⇔悲しい」、「うれしい⇔怒り」、「のどか⇔緊迫」の3つの印象軸で評価し、それぞれの印象軸における印象の程度(印象値)を $1\sim7$ の7段階評価スケール「(左側の印象を)感じる(1点)、わりと感じる(2点)、やや感じる(3点)、(どちらの印象も)感じない(4点)、(右側の印象を)やや感じる(5点)、わりと感じる(6点)、感じる(7点)」に準じた実数値として出力する.

また、本手法をキーワード検索に適用し、任意のキーワードの印象文脈(そのキーワードがどのような印象のツイートにおいて用いられているか)を可視化する手法も提案する. 具体的には、Twitter API を用いてそのキーワードを含むツイート群を収集し、各ツイートの印象を数値化することで、どのような印象のツイートが多いかを円グラフにして提示する.

2. 関連研究

Twitter を利用した研究は数多くなされており、トレンド分析[6]やニュース記事推薦[7]、リアルタイムイベント抽出[8]などに応用されている。また、Twitter ユーザどうしのつながり構築を支援するという目的で、フォロー候補となるユーザを推薦するという研究も盛んであり、総合的に影響力の高いユーザを推薦する研究[9]や実世界における友達関係を推定してユーザを推薦する研究[10]、潜在的な興味が似ているユーザを推薦する研究[11]など様々なアプローチが採用されている。本研究では Sentiment Analysis 技術の一つである印象マイニングを応用している点が新しい.

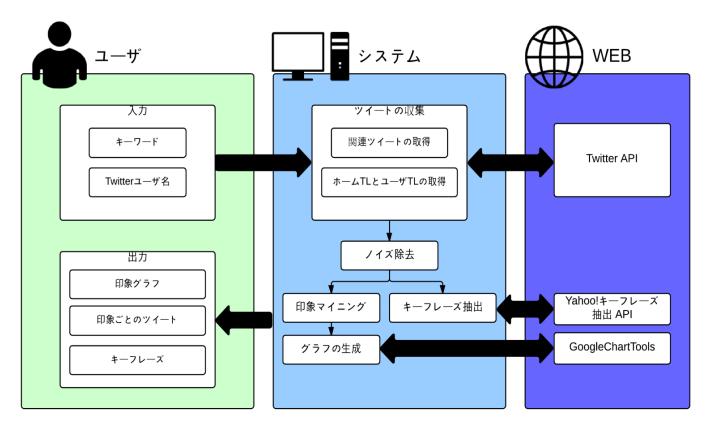


図1 システムの全体構成

Sentiment Analysis の研究分野では、評判や感情、印象といった主観的な情報をテキスト(レビュー、ニュース記事、Webページ、小説など)から抽出するための研究が盛んであり、評判分析[12][13]や情報可視化[14]、情報推薦[15]、異メディアコンテンツ生成[16]、印象タグ付与[17][18]、情報の信頼性評価[19]といった様々な分野で応用されている。しかしながら、ほとんどの研究がクラス分類問題あるいはアノテーション問題としての扱いであり、読み手が感じる印象の程度を数値的に求め、利用するという研究はまだ少ない[20][21][22].

3. 印象可視化システムの設計と開発

3.1 システムの全体構成

本システムの全体構成を図1に示す.

本システムの利用者がキーワードもしくは Twitter ユーザ名を入力すると、システムは Twitter API[2]を用いてツイートを取得する. 入力がキーワードの場合、システムはそのキーワードを含むツイート群を取得し、入力が Twitter ユーザ名の場合、システムはその Twitter ユーザがフォローしている人らのツイート群 (ホームタイムライン) と、その Twitter ユーザ自身が投稿したツイート群 (ユーザタイムライン) を取得する.

次に、取得したツイート群から分析する際にノイズ

となる URL や Twitter ユーザ名を取り除き, 既存の印象マイニングシステム[3]を用いて各ツイートの印象値を算出するとともに, Yahoo! JAPAN 社が提供している Yahoo!キーフレーズ抽出 API[4]を用いてそれぞれのツイート群からキーフレーズを取得する.

最後に、Google 社が提供している GoogleChartTools [5]を用いて、キーワード分析の場合は円グラフを、ユーザ分析の場合は散布図を生成し、キーフレーズとともに利用者に提示する.

以下の節で、本システムの各モジュールについて詳しく説明する.

3.2 ユーザ登録・ログインモジュール

利用者はまず、ユーザ登録画面においてユニークな ID とパスワードを入力し、ユーザ登録を行う必要がある. パスワードは、暗号化され、ID とともにデータベースに保管される.

既に登録済みの利用者は、登録した ID とパスワードを用いて、ログイン画面より本システムにログインすることができる.

3.3 ツイート収集モジュール

利用者は、システムのキーワード分析画面もしくは ユーザ分析画面より入力を行う.

3.3.1 関連ツイートの取得

キーワード分析の場合は、入力されたキーワードを 含むツイート群を Twitter API を用いて取得する. なお、 取得するツイート数の初期値は100ツイートであるが、システムの設定画面より利用者が独自に設定することも可能である.

3.3.2 ホームタイムライン・ユーザタイムラインの取 得

ユーザ分析の場合は、ユーザタイムライン(ユーザTL)とホームタイムライン (ホーム TL) を取得する. ユーザ TL とは指定したユーザ本人が投稿したツイート群のことであり、ホーム TL とは指定したユーザがフォローしている人らが投稿したツイート群のことである.

ユーザ TL は、ユーザ TL 取得用の API が提供されているので、これをそのまま用いて取得する.一方、ホーム TL 取得のために Twitter が提供している API では、認証されたユーザ自身のホーム TL のみが取得できるような仕様になっており、任意のユーザのホーム TL を取得できるようにはなっていない.そこで本システムでは、複数の API を組み合わせて任意のユーザのホーム TL を擬似的に再現し、取得している.

まず、指定されたユーザがフォローしている人らのID一覧をAPIで取得し、さらに、それぞれのIDにおける最新投稿日時をチェックする.ホーム TL は今ユーザが見ているタイムラインのことであるため、この最新投稿日時が別途定める基準日より古い ID は、ホーム TL には表示されないものとして分析対象から除外する.本システムでは基準日の初期値を1日としているが、システムの設定画面より利用者が独自に設定することも可能である.ここまでの処理で、ユーザが今実際に見ているホーム TL に登場するユーザ集合と近いユーザ集合(ID集合)が選別される.

次に、選別された各 ID からそれぞれのユーザ TL を 取得することで、擬似的にホーム TL を再現すること になる. 但し、ID の取得時と同様に、基準日以前のツ イートは取得しない.

3.4 ノイズ除去モジュール

取得したツイート群に顔文字や URL といった分析に適さない文字列が含まれていると、正確な分析結果を得ることができない、そこで、事前にこのような文字列を取り除くことにする.

具体的には、正規表現を用いて文字列検索を行い、マッチする文字列を取り除くことにする. 現在、定義している、取り除く文字列と各文字列を取り除くために用いた正規表現を表 1 にまとめる. すなわち、httpもしくは https から始まる英数文字列を URL (Uniform Resource Locator)を表しているものとして取り除き、@から始まる英数文字列を Twitter ユーザ名を表しているものとして取り除く. さらに、文字「w」が 2 個以上連続したもの(笑いを表す)や顔文字も取り除く.

表 1 ノイズとして取り除く文字列の詳細

除去文字列	正規表現
URL	/(https?: Y/Y/[Yx21-Yx7e]+)/i
Twitter ユーザ名	/@(\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\
顔文字	/(¥(.*?¥))/
2個以上連続する「w」	/(w w){2,}/i

3.5 印象マイニングモジュール

本システムでは、著者らの印象マイニングシステム [3]を用いて各ツイートの印象値を算出する。この印象マイニングシステムは、「楽しい⇔悲しい」、「うれしい ⇔怒り」、「のどか⇔緊迫」という 3 種類の印象軸を対象としており、それぞれの印象軸における印象の程度 (印象値)を $1\sim7$ の7段階評価スケール「(左側の印象を)感じる (1点)、わりと感じる (2点)、やや感じる (3点)、(どちらの印象も)感じない (4点)、(右側の印象を)やや感じる (5点)、わりと感じる (6点)、感じる (7点)」に準じた実数値として出力する。

この印象マイニングシステムを開発するに当たり, まず,「ある印象を有する単語はその印象を表現する印 象語群と共起しやすく,逆の印象を表現する印象語群 とは共起しにくい」という仮定のもと、新聞記事デー タベースに現れる任意の単語と対比的な印象を有する 2 つの印象語群との共起の仕方を調べ、どちらの印象 語群とより共起しやすいかを数値化することにより, 印象辞書(単語が記事の印象に及ぼす影響を数値化し たもの)を構築した. さらに, この印象辞書を用いて 新聞記事の印象値を算出する手法(算出法)を開発し た. そして, この算出法が新聞記事から算出する印象 値とその新聞記事を人々が読んだときに感じる印象値 との関係を回帰分析により定式化することで, 算出法 が算出した印象値を回帰式により補正するという手法 を開発した. その結果, 新聞記事に対しては高い精度 で印象値を算出することができたが, ツイートに対す る有効性は検証されていない. 今後の課題と言える.

3.6 キーフレーズ抽出モジュール

ノイズを除去したツイート群からキーフレーズの 抽出を行う. 本システムでは、Yahoo!JAPAN が提供し ている Yahoo!キーフレーズ抽出 API を用いる.

ノイズを除去したツイート群を 100 ツイートずつに 分割し、これをパラメータとして分割数分 API リクエ ストを行う. 1 回のリクエストで 20 個のキーフレーズ とそれぞれのスコアを得ることができる. 但し、複数 回リクエストした場合、得られるキーフレーズに重複 が生じることがある. そのため、重複しているキーフ レーズはスコアの平均値を求め 1 つにまとめる. 最後 に、スコア上位の 20 個のキーフレーズを、スコアをも とにタグクラウド形式に変換し、利用者に提示する. 3.7 グラフ生成モジュール

ユーザの印象選好やキーワードの印象文脈を可視化するため、ツイートの印象値をもとにグラフを生成する. グラフの生成には、Google 社が提供しているGoogleChartTools を用いる.

ユーザ分析の場合は、各ツイートの印象値に基づいて散布図を 3 つ生成する. 「x 軸:楽しい⇔悲しい $\times y$ 軸:うれしい⇔怒り」、「x 軸:うれしい⇔怒り $\times y$ 軸:のどか⇔緊迫」、「x 軸:のどか⇔緊迫 $\times y$ 軸:楽しい⇔悲しい」の 3 つであり、それぞれの散布図において、ユーザ TL を赤色の丸印で示し、ホーム TL を青白の丸印で示す.

キーワード分析の場合は、各ツイートの印象値に応じてそれぞれのツイートを5段階に再分類する. 例えば、「楽しい⇔悲しい」の印象軸では、印象値が1以上2.5未満のツイートを「楽しい」に、2.5以上3.5未満のツイートを「やや楽しい」に、3.5以上4.5未満のツイートを「普通」に、4.5以上5.5未満のツイートを「きしい」に、5.5以上7以下のツイートを「悲しい」

に再分類する. 本システムは、印象軸ごとに、それぞれのツイート数をカウントし、GoogleChartTools を用いて円グラフを生成する.

4. システムの実行例

本システムは、表 2 に示した開発環境において Web アプリケーションシステムとして実装されており、動作が確認されている.本章では、本システムの実行例を示す.

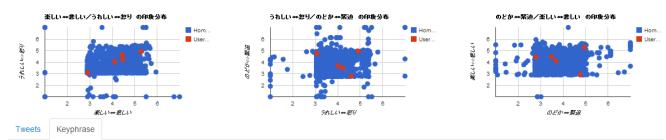
表 2 システムの開発環境

0S	CentOS6.5 64bit
Web サーバ	Apache2.4
データベース	MySQL5.0.10
言語	PHP5.4.7, JavaScript HTML
使用 API	Twitter REST API v1.1, Yahoo!キーフレーズ抽出 API, GoogleChartTools

Twitter印象分析システム キーワード分析 ユーザ分析

user▼

ユーザ「MorinoKumazo」の分析結果



UserTimelineのキーフレーズ

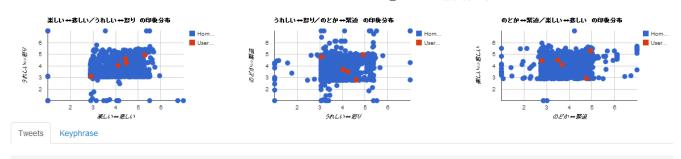
購買 後略, 居住地 写真 千葉工大 未来 谷津1丁目っぽい 返 モリシア 聖地巡礼 第3回WI2研究会 新番組 アニメ 学食 主人公 学生参加報告 バディ・コンプレックス RT 事務 明細9日

HomeTimelineのキーフレーズ

PaymoneyTo my Pain This will be a year of action カプツ 須藤真澄 淡路恵子 You Don't Know What Love Is Continued congressional inaction is leaving more 〈まもと Watch President Obama's weekly address on why Japanisches Fachmagazin provoziert mit Manga-Cover ハッピーバースデー 年質はがき mammograms and contraception cost 松紅 Read the stories of Americans who have ユニュード モン AKIBA PC Hotline Dropbox Briefly Goes Down Sexismus-Vorwurf

図 2 ユーザ分析の実行例 1 (ユーザ TL とホーム TL のキーフレーズを表示させる場合)

ユーザ「MorinoKumazo」の分析結果



楽しい/悲しい/うれしい/怒リ/のどか/緊迫

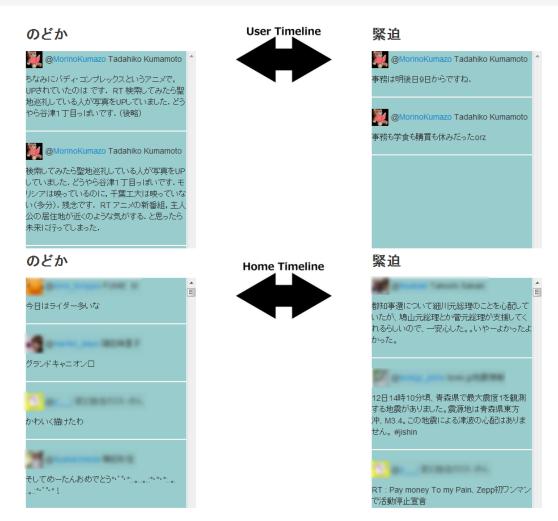


図 3 ユーザ分析の実行例 2 (「のどか⇔緊迫」を選択し,該当するツイートを表示させる場合)

4.1 ユーザ分析の場合

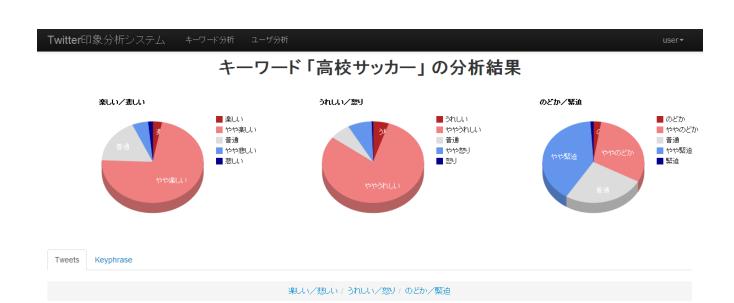
まず、ユーザ分析の実行例を図 2 に示す.分析対象となるユーザとして著者の一人である @Morino Kumazo を指定しており、画面上部には「x 軸:楽しい \leftrightarrow 悲しい $\times y$ 軸:うれしい \leftrightarrow 怒り」、「x 軸:うれしい \leftrightarrow

怒り \times y軸:のどか \leftrightarrow 緊迫 \setminus y軸:楽しい \leftrightarrow 悲しい \setminus のるつの散布図が表示されている。散布図の下には、ユーザ TL とホーム TL のそれぞれにおいて抽出されたキーフレーズがタグクラウド形式に変換され、表示されている。

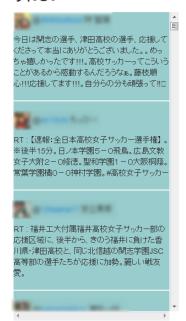
また、本システムの利用者が選択した印象に基づいて、当該印象に分類されたユーザ TL とホーム TL を表示させることもできる.表示させる印象として「のどか⇔緊迫」を選択した場合の実行例を図 3 に示す.この操作により、それぞれの印象に該当するツイートの内容を読むことができる.

4.2 キーワード分析の場合

次に、キーワード分析の実行例を図 4 に示す. 分析 対象となるキーワードとして「高校サッカー」を入力 しており、画面上部には「楽しい⇔悲しい」、「うれしい⇔怒り」、「のどか⇔緊迫」の3つの印象軸に対応した円グラフが表示されている。この円グラフを見ることで、直感的にどのような印象文脈でのツイートが多いかを知ることができる。また、円グラフの下には、利用者が選択した印象(この例では「うれしい⇔怒り」)に分類されたツイートが表示されている。もちろん、図2同様、取得したツイートから抽出されたキーフレーズをタグクラウド形式で表示させることもできる。



うれしい



怒り



図 4 キーワード分析の実行例 (キーワードとして「高校サッカー」を入力した場合) (「うれしい⇔怒り」を選択し,該当するツイートを表示させる場合)

5. むすび

本論文では、システム利用者が指定した条件(Twitter ユーザ名もしくはキーワード)に基づいて該当するツイートを自動的に収集・分析し、任意のユーザの印象選好(そのユーザが普段、どのような話題の、どのような印象のツイートを見たり投稿したりしているか)や任意のキーワードの印象文脈(そのキーワードがどのような印象のツイートにおいて用いられているか)を可視化する手法を提案するとともに、Webアプリケーションシステムとして実装した.

ユーザ分析では、利用者が指定した Twitter ユーザのユーザタイムライン(ユーザ TL)とホームタイムライン(ホーム TL)を収集・分析し、そのユーザの印象選好を可視化することで、そのユーザをフォローしても良いか、しない方が良いかの判断を支援する機能を提供している。キーワード分析では、利用者が入力したキーワードを含むツイートを収集・分析し、そのキーワードの印象文脈を可視化することで、キーワードの使われ方を調べることができる.

今後の課題として,以下のようなものが挙げられる. まず,本システムで用いている印象マイニングシステ ム[3]は、新聞記事の分析用途で設計・開発されたもの であり, ツイートの印象マイニングに対する有効性は 検証されていない. ツイートには非文法的なものや極 端に短い文からなるものも多く, 現在の辞書ベースの 方法では十分な精度が得られないかもしれない. した がって,ツイートに対する現印象マイニングシステム の精度を検証しつつ, ツイートに対応可能な印象マイ ニング手法を開発していく必要がある.一方,ツイー トには Twitter 特有の顔文字やネットスラング, 言い回 しが含まれる場合があり、ツイートの印象にも影響を 与えているものと考えられる[23]. そのような場合に も正確に印象値を算出できるように印象マイニングシ ステムを改良する必要がある.また,本システムを発 展させることで, ユーザの印象選好やキーワードの印 象文脈を可視化するだけでなく、実際にフォローすべ きユーザを推薦したり、利用者の興味のあるキーワー ドを提示したりすることのできるシステムを開発して いきたい.

謝辞

本研究は, JSPS 科研費 24500134 ならびに福田将治 奨学寄付金による助成の成果であり, ここに記して謝 意を表すものとする.

参考文献

[1] SOCIALPLATFORMS GWI.8 UPDATE: Decline of Local Social Media Platforms, https://www.globalwebindex.net/social-platforms-gwi-8-update-d

- ecline-of-local-social-media-platforms/
- [2] Documentation | Twitter Developers, https://dev. twitter.com/docs
- [3] 熊本忠彦,河合由起子,田中克己:新聞記事を対象とするテキスト印象マイニング手法の設計と評価,信学論(D), Vol.J94-D, No.3, pp.540-548, 2011.
- [4] テキスト解析:キーフレーズ抽出 API Yahoo!デベロッパーネットワーク, http://developer.yahoo.co.jp/webapi/jlp/keyphrase/v1/extract.html
- [5] Google Charts Google Developers, https://developers.google.com/chart/
- [6] Michael Mathioudakis, and Nick Koudas, Twitter-Monitor: Trend Detection over the Twitter Stream, Proc. of the ACM SIGMOD International Conference on Management of Data, pp.1155-1158, Indianapolis, USA, 2010.
- [7] Fabian Abel, Qi Gao, Geert-Jan Houben, Ke Tao, Twitter-Based User Modeling for News Recommendations, Proc. of the 23rd International Joint Conference on Artificial Intelligence, pp. 2962-2966, Beijing, China, 2013.
- [8] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo, Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors, Proc. of the 19th International Conference on World Wide Web, pp. 851-860, Raleigh, North Carolina, USA, 2010.
- [9] Jianshu Weng, Ee-Peng Lim, Jing Jiang, and Qi He, TwitterRank: Finding Topic-Sensitive Influential Twitterers, Proc. of the Third ACM International Conference on Web Search and Data Mining, pp. 261-270, New York, USA, 2010.
- [10] Adam Sadilek, Henry Kautz, and Jeffrey P. Bigham, Finding Your Friends and Following Them to Where You Are, Proc. of the Fifth ACM International Conference on Web Search and Data Mining, pp. 723-732, Seattle, USA, 2010.
- [11] Marco Pennacchiotti, and Siva Gurumurthy, Investigating Topic Models for Social Media User Recommendation, Proc. of the 20th International Conference Companion on World Wide Web, pp. 101-102, Hyderabad, India, 2011.
- [12] Peter D. Turney: Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. Proc. of the 40th Annual Meeting on Association for Computational Linguistics, pp.417-424, Philadelphia, USA, 2002.
- [13] Bo Pang and Lillian Lee: Seeing Stars: Exploiting Class Relationships for Sentiment Categorization with Respect to Rating Scales. Proc. of the Annual Meeting on Association for Computational Linguistics, pp.115-124, Morristown, NJ, USA, 2005.
- [14] Kevin Hsin-Yih Lin, Changhua Yang, and Hsin-Hsi Chen: Emotion Classification of Online News Articles from the Reader's Perspective. Proc. of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, Vol.1, pp.220-226, 2008.
- [15] 矢野絵美, 北野有亮, 末吉恵美, 篠原勲, ピンヤポン シニーナット, 加藤俊一, 消費者の感性モデルを利用したレコメンデーションシステムの構築, 情報処理学会論文誌: データベース, Vol.44, No.SIG 8 (TOD 18), pp.46-54, 2003.
- [16] 石塚賢吉,鬼沢武久,加藤茂:物語のシーンの印象に基づいた声楽曲の生成,日本感性工学会論文誌,Vol.10,No.4,pp.523-534,2011.

- [17]宮川祥子,清木康:特定分野ドキュメントを対象とした意味的連想検索のためのメタデータ空間生成方式,情報処理学会論文誌:データベース, Vol.40, No.SIG 5 (TOD 2), pp.15-28, 1999.
- [18]清水浩平, 萩原将文:名詞と動詞の組み合わせに 対する印象推定法,日本感性工学会論文誌, Vol.10, No.4, pp.505-514, 2011.
- [19]山本祐輔, 手塚太郎, アダムヤトフト, 田中克己, WebAlert: Web 情報の印象集約を利用した閲覧ページ内容に対する反対意見提示, 日本データベース学会論文誌, Vol.7, No.1, pp.251-256, 2008.
- [20] Tadahiko Kumamoto and Katsumi Tanaka: Web OpinionPoll: Extracting People's View by Impression Mining from the Web, Proc. of the 14th ACM Conference on Information and Knowledge Management, pp.265-266, Bremen, Germany, 2005.
- [21] 張建偉,河合由起子,熊本忠彦,田中克己:地域性に基づく発信者の観点差異を可視化するセンチメントマップシステムの提案,情報処理学会論文誌データベース, Vol.3, No.1 (TOD 45), pp.38-48, 2010.
- [22]河合由起子,熊本忠彦,田中克己:印象と興味に基づくユーザ選好のモデル化手法の提案とニュースサイトへの応用,知能と情報,Vol.18, No.2,pp.173-183,2006.
- [23] 若井祐樹,田中美羽,熊本忠彦,灘本明代:顔文字を考慮したニュースに対するツイートの感情抽出手法の提案,第5回データ工学と情報マネジメントに関するフォーラム(DEIM 2013), No.D9-5, 2013.