

感情語辞書と位置情報付きツイート分析に基づいた  
アプリケーション「EmoNavi」の観光利用の検討

齋藤 一 横川 祥司

北海道情報大学

A Study of Application "EmoNavi" for Tourism based on  
Analysis of the Tweets with Location Information and  
Emotion Words Dictionary

Hajime SAITO and Shyoji YOKOKAWA

Hokkaido Information University

平成28年11月

北海道情報大学紀要 第28巻 第1号別刷

## 〈研究ノート〉

## 感情語辞書と位置情報付きツイート分析に基づいた アプリケーション「EmoNavi」の観光利用の検討

斎藤 一\* 横川 祥司†

### A Study of Application “EmoNavi” for Tourism based on Analysis of the Tweets with Location Information and Emotion Words Dictionary

Hajime Saito\* Shyoji Yokokawa†

#### 要旨

EmoNavi は位置情報を含んだツイートを収集し、感情語辞書を用いて感情を特定する分析システムと、感情語が含まれた複数のツイートを地図上に表示させるシステムで構成される。分析システムは、既存の観光情報分析手法と崩れた語の処理に強い形態素解析ツール、新規の名詞の獲得ができる専門用語自動抽出システムを組み合わせることで実現した。抽出したツイートを地図上に表示するシステムは、Web アプリケーションとして開発することで、PC の他、スマートフォンやタブレット端末からも利用できる。本稿では EmoNavi の観光利用の可能性について検討する。

#### Abstract

We have developed the application “EmoNavi” to visualize the feeling of people in a place of sightseeing based on analysis of the tweets with location information and emotion words dictionary. EmoNavi consists of three parts; gathering tweets module, analyzing tweets module, and viewer displaying emotion icons on map. Analyzing module was developed by some methods, those are morphological analysis for informal words, and extracting technical terms. Displaying module was developed as a web application for PCs and smart devices. In this paper we discuss the possibility of EmoNavi on tourism.

#### キーワード

Twitter, 感情語辞書 (Emotion Dictionary), 位置情報 (Location Information)

\* 北海道情報大学情報メディア学部情報メディア学科准教授, Associate Professor, Department of Information Media, HIU

† 北海道情報大学大学院経営情報学研究科 (平成 25 年 3 月修了), Graduate School of HIU

## 1. はじめに

近年、観光振興のために、SNS（Social Networking Service）の一つである、Twitter の位置情報付きのデータを分析する研究が行われている。Twitter はツイートと呼ばれる140字以内のコメントの投稿を主な機能としている。文字数が少ないことで、気軽に今感じたことをツイートしやすいため、リアルタイム性が強く、最新の話題や情報を取得しやすいという特徴を持っている。

遠藤らは、日本国内で発信される位置情報付きツイートを対象に、桜や紅葉の見頃をTwitter から推定する手法を提案している[1]。佐伯らは、対面での調査を行わずに訪日外国人の実際の訪問先を得るため、ツイートに付与された位置情報を用いて、外国人ユーザの使用言語と東京での滞在先の分布を月別に分析している[2]。

一方で、ツイートに含まれる感情語に着目した研究や取り組みがある。ここで、感情とは、『喜んだり悲しんだりする、心の動き。気持ち。気分[3]』を意味し、本研究では、感情を表現する単語を感情語と呼ぶ。内藤らは、イベントについてのツイートを提示した場合に、ツイートに感情語が含まれる場合と含まれない場合における読者の感情の変化について調査を行った。調査・分析の結果、提示するツイートは感情語を含んでいる方が共感を促進することが分かった[4]。『feel on Twitter』は、ツイートに含まれる感情語を用いたスマートフォン向けアプリケーションである[5]。このアプリケーションは、ツイートに含まれる感情語に合わせた漫画を表示する機能を持っている。その際、ツイートはその漫画の吹き出し内に表示されるため、アプリケーションのユーザは、ツイートの台詞で構成された漫画を読むように楽しむことができる。また、大量のツイートを分析することで、肯定的な意見が多いのか、否定的な意見が多いのかを判別す

る Web サービス『Twitter 感情分析所』も運用されている[6]。

本研究では、札幌駅から半径 20 km 以内の位置情報付きツイートを収集し、それらのツイートにどのような感情語が含まれるのかに基づいて、「喜び」「悲しみ」「怒り」「驚き」の4つに分類した。また、これらのツイートの感情を表すアイコンを地図上に表示するアプリケーション『EmoNavi』を開発した。EmoNaviにより、札幌で「今」どのような感情の人が多いのか、また、指定した日付において、どのような感情の人が存在したのかを可視化することができる。本稿では、EmoNavi のツイートの処理と可視化の方法を示す。また、EmoNavi の観光利用について検討する。

## 2. EmoNavi の概要

### 2-1 先行研究

徳久らは、観光開発のヒントを得るために、ブログから旅行者の評判情報を抽出する手法を提案した[7]。この手法は、1) 観光地のブログ記事と一般のブログ記事を比較することで、観光地ブログ記事に出現しやすいキーワード（特徴語）を抽出すること、2) 感情の原因・状態・表出を表す表現を手がかりとしてブログ記事から感情を推定すること、という2点の要素技術を用いて評判情報を抽出している。

本研究では、アプリケーションにより即時性を持たせるため、徳久らの手法を参考に、ブログではなく、Twitter を利用する。

### 2-2 EmoNavi の構成

本研究では、位置情報付きツイートを「地域ツイート」、それ以外のツイートを「一般ツイート」と呼ぶ。EmoNavi は以下の4つのステップで処理を行う。

1. ツイートの収集：地域ツイートおよび一般ツイートを収集しデータベースに保存する。

2. 感情推定：感情語辞書を参照し、ツイート内の感情を推定する。
3. 特徴語抽出：地域ツイートに含まれる名詞に対して、次章で説明する特徴度を計算し、その値の高い名詞を特徴語として抽出する。ツイート、特徴語、感情推定の結果をデータベースに保存する。
4. 結果出力：保存されたツイートのデータを感情推定の結果のアイコンを添えて地図上に表示する。

EmoNavi の開発には、PHP5.3.1[8]とMySQL5.1.66[9]、TwitterAPI1.1[10]を用いた。EmoNavi は 2012 年 12 月から運用している。尚、TwitterAPI の仕様変更に伴い、ツイート収集を行うプログラムを、2015 年 12 月に新しい仕様（JSON 形式対応版）に変更している。EmoNavi は、日本国内のどの地域でも適用可能である。本稿では、ツイートの量が多く、本学から近い『札幌』を対象とした処理について示す。次の章では各ステップの処理の詳細を、実装方法と共に示す。

### 3. EmoNavi の実装

#### 3-1 ツイートの収集

##### 3-1-1 地域ツイートの収集

札幌駅より 20km 圏内に含まれる地域ツイートを TwitterAPI の一種である TwitterSearchAPI により収集した。TwitterSearchAPI は 1 時間あたり 150 回の実行制限、1 回あたり 100 件のツイートの取得制限があるため、取得間隔を 1 分間（1 時間あたり 60 回）とし、取得件数を 1 回あたり 100 件に設定した[11]。地域ツイートはデータベースに保存する。テーブルには、ツイートの ID、ツイート自身が持つ URL、そのツイートを行ったユーザ名、ツイート本文、ツイートの位置情報、ツイートされた時間を保存した。

##### 3-1-2 一般ツイートの収集

一般ツイートは、Twitter API の仕様変

更以前は、TwitterStreamingAPI を用いてパブリックタイムラインからランダムにツイートを収集していた。しかし、API の仕様変更によりパブリックタイムラインが廃止されたため、2015 年 12 月以後は、EmoNavi 専用の Twitter アカウント (@EmoNavi) のホームタイムラインからツイートを収集している。このアカウントでは、日本語の一般的なツイートを取得するため、朝日新聞、毎日新聞、日本経済新聞、読売新聞、産経新聞、および、NHK から配信されているツイートのフォローのみを行っている。一般ツイートは 30 分毎に、最新のツイートを 10 件取得している。尚、一般ツイートは、3-3 で示す特徴語抽出のために使用する。

### 3-2 感情の推定

#### 3-2-1 感情の種類と感情語辞書

徳久ら[7]は感情の種類を「喜び」、「嫌」、「悲しみ」、「驚き」、「感情なし」の 5 種類に設定している。本研究では、「嫌」の感情を、「怒り」とし、「喜び」、「怒り」、「悲しみ」、「驚き」の 4 種類とする。また、感情語辞書[12]を基にし、表 1 のような感情と感情語のセットを作成した。このセットを用いて、ツイートに感情語が含まれるのかを検索し、含まれた感情語の種類によってツイートの感情を決めることを、本研究では、感情の推定と呼ぶ。また、ツイート内に複数の感情語が見つかった場合、現在に一番近い感情を採用するという考えから、ツイートの最後に現れた感情語を用いて感情の推定を行う。感情と感情語のセットは、感情表現辞典に加え、喜びと悲しみの 2 極の感情で分ける単語積極性辞書[13]のデータも用いた。そのため、驚きと怒りと比べて、喜びと悲しみの感情語が多い。その結果、喜びが 691 語、怒りが 228 語、悲しみが 497 語、驚きが 133 語となった。

#### 3-2-2 顔文字の活用

ツイートには顔文字が数多く使われている。顔文字は人の感情や動作を表しているものが

多く、感情推定に有用である[14]。位置情報付きのツイートは主に iPhone などのスマートフォンから行われる場合が多いと考えた。そこで、iPhone の文字変換機能の辞書内にデフォルトで入っている 202 の顔文字を感情語辞書内に挿入した (表 2)。顔文字の感情分けは、顔文字を感情別でまとめている Web サイト、matsucon[15]を参考にした。顔文字 202 個中、160 個の顔文字に感情の種類を決めることができた。顔文字の感情の種類の内訳は、喜びが 60 個、怒りが 29 個、悲しみが 40 個、驚きが 31 個となり、それらを感情語辞書に追加した。

表 1 感情語の例

喜	喜び, 好ましい, 期待している, 美味しい, にっこり, すてき, たのしい, やったー
悲	悲しみ, 苦しい, 恐れ, 怖い, 残念, 痛ましい, 偽る, 危ない, つらい, 寒い
怒	怒り, 嫌, 好かない, 憤慨, 酷い, 気に入らない, 悔しい, おこる, やめろ, むかつく
驚	驚く, 驚嘆, 度肝を抜かれる, 目を見張る, 放心, マジ?, ショック, ぎよつ, びくっ

表 2 顔文字と感情の対応例

喜	(^_^), (=^▽^ )人(^▽^=), ♪(^▽^ ) (^_^)a, (^_^)v, (^3^)-☆, (・ω・)ノ, (▽▽)
悲	(;_;)/~::~, (-_-), (-.-;), (._.), (/ _;), (>_<), (@_ρ_@), (p_-), (u_u),
怒	(*`ω´), (*`へ´), (;^_ ^A, (#`^´)
驚	( ; `Д` ), ((((( ; `Д` ))))))), ( `Д` ), (・・?), (° O °)\(-.-; , Σ(¯ . ¯)ノ, Σ (・ □ ・ ;), Σ(° Д °)

### 3-3 特徴語の抽出

本研究では、地域ツイートを形態素解析し、

その結果から抽出した名詞に対し、スコア関数 Okapi-BM25[7]により、地域の特徴を表しているかを示す値、特徴度を計算する。特徴度の高い名詞を特徴語と呼ぶ。特徴語は、地域ツイートの頻出し、かつ、地域ツイートおよび一般ツイートを通して網羅的に現れない名詞とする。

#### 3-3-1 名詞の抽出

地域ツイートから名詞を抽出するために形態素解析を行う。しかし、ツイートの中には「ど一思う？」や「ビミョー」といった、小文字(捨て仮名)や長音符を用いた語が含まれる場合が多い。また、「Google で検索する」を意味する「ぐる」に代表されるネット上での俗語が用いられる場合もある。このような表現は形態素解析ツールの辞書にない場合も多い[16]。こういった問題を解決するために形態素解析ツール JUMAN7.0[17]を用いた。JUMAN7.0 は一定間隔で内部の語が更新される Web 自動辞書を持ち、小文字、長音符の自動認識のできる形態素解析ツールである。

また、Twitter では、新製品や新しい建物の名称、新たに開催されるイベント名等の新語が扱われることがある。辞書に登録されていない新語を取り出せるようにするため専門用語自動抽出システム TermExtract を用いた。TermExtract は辞書に登録されていない複合名詞や専門用語を見つけ出すことができる[18]。

#### 3-3-2 特徴度の判定

本研究では、3-2-1 で、地域ツイートから抽出した名詞  $q$  が地域の特徴を表しているかを表す特徴度  $s$  を算出するため、Okapi-BM25を用いる。Okapi-BM25 は、主に文書検索に使用される関数で、文章の長さの結果が左右されづらい特徴を持っており、本研究では、式(1)で表す。収集した最新の地域ツイート 500 件を一つにまとめた文書を  $D_t$  とする。また、一般ツイートを最新のツイートから順に、500 件ずつまとめた文書を 3 つ( $D_{g1}$ ,  $D_{g2}$ ,  $D_{g3}$ )用意する。本来は、ツイートをまとめた文書ではなく、一つ一つのツイートについて処理をするべきではあるが、サーバへの負荷や計算時間

を考慮した結果、この方法を採用することにした。Okapi-BM25の算出で用いる文書集合  $D$  を  $\{D_t, D_{g1}, D_{g2}, D_{g3}\}$  とする。式(1)を用いて  $q$  の特徴度  $s(D_t, q)$  を計算する。式(1)の  $f(q, D_t)$  は  $D_t$  における  $q$  の出現頻度、また、 $|D_t|$  は  $D_t$  の長さ、 $avgdl$  は文書集合  $D$  における文書の平均長(平均文字数)である。 $k$  と  $b$  は定数であり、文献[7]と同様にそれぞれ 2.0 と 0.75 とした。式(2)の IDF (Inverse Document Frequency) は、逆文書頻度を表し、 $q$  が特定の文書にのみ出現するとき大きな値となる。また、式(2)において、 $N$  は文書集合  $D$  の大きさを示す。 $n(q)$  は文書集合  $D$  の中で、 $q$  を含む文書の数を示す。

$$s(D_t, q) = IDF(q) \cdot \frac{f(q, D_t) \cdot (k + 1)}{f(q, D_t) + k \cdot (1 - b + b \cdot \frac{|D_t|}{avgdl})} \quad \dots (1)$$

$$IDF(q) = \log \frac{N - n(q) + 0.5}{n(q) + 0.5} \quad \dots (2)$$

名詞  $q$  が、地域ツイートで頻出し、かつ特定のツイートでのみ多く出現するのであれば特徴度は大きくなる。逆に、地域ツイートおよび一般ツイートを含め網羅的に出現するのであれば、特徴度は小さくなる。特徴度の大きい(ここでは3以上の)  $q$  を特徴語として採用する。この計算は1日毎に行い、特徴語を更新していく。表3はそのようにして得られた特徴語の例である。

表3 抽出した特徴語の例

特徴語	特徴度
サッポロビール	3.343323883
コンサドーレ	3.200254329
日高	3.200254329
コーチャンフォー	3.111821442
ノースサファリサッポロ	3.090642111
定山溪	3.089629108
狸小路	3.079842724

### 3-4 抽出結果の出力

EmoNaviは、感情推定および特徴語抽出を行った結果を地図上に表示する。スマートフォンやタブレット、モバイルPC等の携帯性に優れたインターネットにアクセスできる機器での利用を想定しているため、WebアプリケーションとしてPHPで開発した。地図はGoogleMapsAPI[19]を用いた。

ブラウザでEmoNaviのURL[20]にアクセスするとトップページが表示される。トップページには「日付の設定」と「感情アイコンの表示」のプルダウンメニューと、「キーワード検索」のためのテキストボックスおよび「検索」ボタンがある。感情アイコンには、図1の4種類がある。検索ボタンをタップすると、指定した感情のアイコンが地図上にプロットされる(図2)。表示される最大ツイート数は5,000とした。また、感情アイコンをタップすると、ツイートの本文が表示される。また、ツイートを表示エリアないのTimelineボタンをクリックすると、収集したツイートの一覧を表示することができる。(図3)。設計当初、特徴語および感情語が含まれる地域ツイートのみを地図上に表示する予定であった。しかし、特徴語と感情語が同時に含まれるツイートが少ないため、感情語のみが含まれている場合も地図上に表示することにした。また、特徴語のみを含む地域ツイートも、GoogleMapsのデフォルトのアイコンで表示される。特徴語は、タイムライン表示でも確認することができる(図3)。




感情アイコン	感情の種類
	喜び、好ましい
	悲しみ、恐れ
	怒り、嫌
	驚き

図1 感情アイコンの種類



図 2 地図上の表示例



図 3 タイムラインの表示例



図 4 EmoNavi の利用画面

尚、特徴語抽出にサーバの負荷が掛かりすぎて不安定となるため、現在は、地域ツイートの収集間隔を1分間隔から3分間隔に減らしている。また、収集間隔は変更せず、特徴語抽出の機能をオミットしたプログラムも別サイト[21]で公開している。

## 4. EmoNavi の活用

### 4-1 利用シーンの検討

図 5 は、札幌雪祭り期間中の EmoNavi の画面である。本研究では、EmoNavi のツイートアイコンの数から混雑具合を予想し、感情アイコン付きのツイート（図 5 では驚きの感

情アイコン付きのツイート）で穴場スポットを見つける、といった利用の仕方を想定している。つまり、ある程度観光目的やプランが決まっている中で、混雑回避や想定外のイベントや名所の発見等を狙っている。

### 4-2 利用実験

4-1 で示したような EmoNavi の活用を検討するために、実際の観光中での利用実験を行った。実験では、図 6 のように出発地を札幌駅、目的地をすすきの駅として設定した。被験者は札幌駅を出発後、EmoNavi から得られる情報をもとに 90 分間自由に観光をしてもらいつつ、徒歩で目的地であるすすきの駅を目指してもらった。

実験は、2013 年 1 月 24 日に被験者 15 名によるテストを行った。被験者は 20 歳から 40 歳までの男女、15 名であり、一人で EmoNavi を使う 5 人と、複数人で一つのスマートフォンから EmoNavi を使う 5 つのグループで行った。



図 5 被験者の行動履歴の例

### 4-3 実験結果

実験後のアンケートにより、観光利用できたかどうかを調査した。その結果、EmoNavi が観光の手掛かりになったと答えた被験者は 69% となった。逆に手掛かりにならなかったと答えた被験者は 31% となった。それぞれの理由について、被験者の自由記述内容を以下に示す。

<手掛かりになった理由>

- ・ ただ、ぶらつくより、目的地の参考になるものがあつた方が行動しやすかつた。
- ・ 少しではあるがイベント情報のツイートがあつたので手がかりになつた。

- ・多くのツイートから目的の場所を見つけることができた。
  - ・ツイートを見て次の行先を決めることができた。しかし、スポットに関するツイートが少なく見つけるのに時間がかかってしまった。
  - ・ラーメンを食べたく思い、現在地から近い店を探す時に役立った。
  - ・ラーメン屋を探した時に役に立った。
  - ・飲食店を探す分には問題なかった。
- <手掛かりにならなかった理由>
- ・良い情報がなかった
  - ・関係ないツイートが多くて探すのが大変だった。
  - ・検索が引っかけられないことが多かった。感情の良し悪しを注目することが無かった。
  - ・知っているスポットの周辺に、新たな情報や意外性のある情報が提示されていると理想的だと思う。

## 5. おわりに

EmoNavi は、地域ツイートの収集とその感情推定を行い、感情アイコン付きツイートを地図上に表示させる。地域ツイートを地図上に表示するプログラムはWebアプリケーションとして開発することで、PCの他、スマートフォンやタブレット端末からも利用できる。

札幌市内でEmoNaviの利用実験を実施し、観光行動中にリアルタイムで、位置情報と感情が特定されたツイートの抽出と地図上への表示が可能であることを確認した。しかしながら、現時点では、EmoNaviの地域ツイートを感情アイコン別でフィルタリングして地図上に表示する機能のみの評価に留まっており、特徴語が有効に活用できていない。今後、一般ツイート集の収集方法や特徴度の計算方法の再検討と改良等を行っていく予定である。

### 参考文献

- [1] 遠藤雅樹他(2016),「ツイートをを用いた生物季節観測の見頃推定手法による情報提供の検討」,『観光と情報』, Vol.12, No.1, pp.47-60.
- [2] 佐伯圭介他(2015),「Twitterデータを利

- 用した訪日外国人の訪問先の言語別分析」,『観光と情報』, Vol.11, No.1, pp.45-56.
- [3] 大辞林 第二版, 三省堂。
- [4] 内藤和宏, 榎堀優, 梶田将司, 間瀬健二(2012),「Twitter コメントに含まれる感情語がイベント印象に与える影響の評価」,『情報処理学会インタラクシオン 2012』, pp.871-876.
- [5] 「Feel on!」, <http://www.feel-on.com/> (2015年8月31日サービス終了) (2013年1月9日アクセス)。
- [6] 「Twitter感情分析所」 <http://mueki.net/twana/> (2013年1月9日アクセス)。
- [7] 徳久雅人, 奥村秀人, 村田真樹(2011),「観光開発支援のためのブログ記事からの評判分析」,『観光と情報』, Vol.7, No.1, pp.85-97.
- [8] 「PHP」, <http://php.net/manual/> (2013年1月9日アクセス)。
- [9] 「Mysql」, <http://www.jp.mysql.com/> (2013年1月9日アクセス)。
- [10] 「TwitterAPI」, <https://dev.twitter.com/> (2016年1月6日アクセス)。
- [11] 辻村博(2010),「TwitterAPIプログラミング」,株式会社ワークスコーポレーション
- [12] 中村明(1993),「感情表現辞典」,東京堂出版。
- [13] 高村大也, 乾孝司, 奥村学(2006),「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』, Vol.47, No.2, pp.627-637.
- [14] 中村明(1993),「感情表現辞典」,東京堂出版。
- [15] matsucon, <http://matsucon.net/material/dic/> (2013年8月6日アクセス)。
- [16] 笹野遼平, 鍛冶伸裕(2012),「新しい語・崩れた表記の処理」,『情報処理』, Vol.53, No.3, pp.211-216.
- [17] 日本語形態素解析システム JUMAN7.0: <http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN>, (2013年, 1月9日アクセス)。
- [18] 小島浩之, 前田朗(2003),「キーワード(専門用語)自動抽出システムの構想とその展開」,『第51回日本図書館情報学会研究発表要綱』, pp.17-20.



- [19] Googlemap : <https://maps.google.co.jp/>  
(2013年1月9日アクセス)。
- [20] EmoNavi,  
[http://sherry.do-johodai.ac.jp/  
FeelTwitter/](http://sherry.do-johodai.ac.jp/FeelTwitter/)  
(2016年8月31日アクセス)。
- [21] EmoNavi (感情推定機能のみ版),  
<http://s123.do-johodai.ac.jp/FeelTwitter/>  
(2016年8月31日アクセス)。