

# 口コミに含まれる感情語を利用した観光地分類の検討

杉本 祐介<sup>1,a)</sup> 水野 忠則<sup>2</sup> 菱田 隆彰<sup>2</sup>

**概要**：近年、インターネット上では、Facebook や Twitter における日記や amazon.com や食ベログにおける口コミなど、ユーザからの投稿を利用したサービスが数多く普及している。これらのサービスに寄せられる投稿の中には、楽しい、きれいといった感情が含まれており、解析によってそれらを明らかにすることで、有用なデータを導き出すことが可能であると考えられる。そこで本研究では、口コミに含まれる感情に着目し、口コミに含まれる感情語を利用した観光地分類の検討を行う。

## Classification of tourist sites using the emotional words in user reviews

SUGIMOTO YUUSUKE<sup>1,a)</sup> MIZUNO TADANORI<sup>2</sup> HISHIDA TAKAAKI<sup>2</sup>

### 1. はじめに

近年、インターネット上では、Facebook や Twitter における日記や amazon.com や食ベログにおける口コミなど、ユーザからの投稿を利用したサービスが数多く普及している。これらのサービスが普及したことで、我々は、家族や友人、あるいは有名人などの近況や、売れ筋の商品や料理の美味しい飲食店などの情報を気軽に取得できるようになった。これらのサービスの利用者数は、携帯電話やスマートフォンなどの普及に伴い、増加の一途を辿っており、ユーザから寄せられる投稿の量も膨大なものになっている。ユーザから寄せられる投稿の中には、楽しい、きれいといった感情や、関連性の高いものとの結びつきを示す特徴語が含まれており、解析によって導き出した感情や特徴語を基に、データの分類や傾向の分析などを行う研究が数多く行われている。

我々はこれまでに、口コミから抽出した各情報の活用を目的として、商品や観光地に対して投稿された口コミからの感情語や特徴語の抽出を行い、それらの情報が有用であることを確認してきた。そこで今回は、観光地に対して投稿された口コミに含まれる情報のうち、感情に重点を置

き、感情語を利用した観光地分類の検討を行う。本論文では、第2節で関連研究について、第3節で我々が行った先行研究において開発したデータベースシステムについて述べる。第4節にて観光地に対する口コミについての調査とその結果を示し、第5節でまとめを述べる。

### 2. 関連研究

テキストからの感情の抽出については、数多くの研究が行われている。

徳久らの研究 [7] では、あらかじめ定義した感情表現を元に、Web テキストから感情生起要因（その感情表現を用いるのに至った要因）を収集し、収集した感情生起要因を用いてユーザの感情を推定している。徳久らは、構築した感情生起要因コーパスを用いて同コーパスから抽出したテストデータを評価し、高い精度の感情推定が実現されていることを確認した。

高村らの研究 [6] では、各単語を電子、各単語が持つ感情極性を電子のスピン向きとみなすことで、エネルギー関数による感情推定を行っている。高村らは、WordNet[1]、Penn TreeBank[2] の語釈文や表現を学習に利用し、General Inquirer[5] の語彙をテストデータとして評価を行い、高い精度の感情極性分類が行えることを確認した。

中山らの研究 [4] では、あらかじめ定義したシードを元に感情語やパターンを収集し、収集した感情語やパターンを新しいシードとすることで、より精度の高い辞書を構築

<sup>1</sup> 愛知工業大学大学院 経営情報科学研究科  
Graduate School of Business Administration and Computer Science, Aichi Institute of Technology

<sup>2</sup> 愛知工業大学 情報科学部  
Faculty of Information Science, Aichi Institute of Technology

a) yuusuke.sugimoto@gmail.com

する手法を提案している。

また、我々が行った先行研究としては、以下のようなものがある。

[8]では、価格比較 Web サイトに投稿された口コミを対象に感情語の抽出を行い、抽出した感情語を利用して口コミの要約などを行う手法を検討した。

[9]では、観光 Web サイトに投稿された口コミを対象に特徴語の抽出を行い、各観光地に対する口コミ数および抽出した特徴語を利用して各観光地の比較を行った。同時に、口コミを活用するための自律的データベースシステムとして、AHLE (Autonomous Human probes system with Local and Emotion functions)を開発した。AHLEについては、次節で詳しく述べる。

我々がこれまでに行った研究では、口コミから抽出した感情語およびその分類がどの程度正しいのかの確認をあまり行ってこなかったため、本研究では感情語を利用した観光地分類に先立ち、各口コミに対する感情語の分類がどの程度正しいのかの検証を行う。

### 3. AHLE について

データベースシステムにおいて感情や特徴語の情報を利用するには、感情や特徴語の取り扱いを考慮した設計を行う必要がある。また、投稿された口コミからの各情報の抽出や、抽出した情報を利用した検索、といった機能を実装する必要があり、一般のデータベースシステムだけでこれら全ての要求を満たすのは困難である。そのため我々は、感情や特徴語の取り扱いに対応したデータベースシステムを開発することにし、その第1段階として、感情の取り扱いに対応したデータベースシステムの開発を行うことにした。

口コミを活用するためには、口コミを収集、集約、解析、分析するといった作業が必要である。我々は先行研究において、これらをまとめて行えるようなデータベースシステムとして、感情・地方特化型自律システム AHLE (Autonomous Human probes system with Local and Emotion functions)を開発した。AHLEは、以下に示す4つの要素から構成されている。

- (1) 口コミの格納を行うライフログデータベース
- (2) データの解析を行う解析エンジン
- (3) ライフログデータベースとクライアント間の橋渡しを行うインターフェース
- (4) 口コミの取得や表示を行うクライアント

AHLEの概要図を図1に示す。ここで図中において、口コミではなくライフログという表記がされているが、これは、AHLEが口コミだけではなく、様々なライフログの集約および活用を想定しているためである。

AHLEでは現在、表1に示すようなデータモデルを利用している。このデータモデルは、中村らの論文[3]を参考

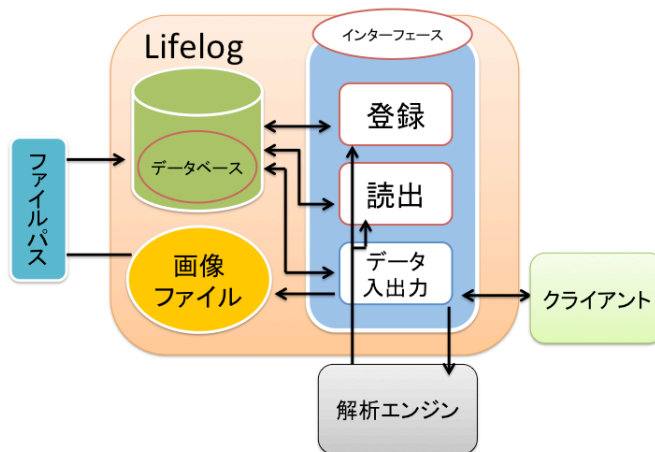


図 1 AHLE の概要図

表 1 データモデルの対比

観点	中村モデル	提案モデル	備考
WHEN	date	date	日付
	time	time	時刻
WHO	user	user	ユーザ
	party	party	同行したユーザ
	object	—	対象のユーザ
	—	target	対象
WHERE	location	—	場所
	—	fromLocation	投稿を行った場所
	—	toLocation	対象の場所
HOW	application device	application device	アプリケーション 利用したデバイス
WHAT	content	—	ログの内容
	ref.schema	—	外部スキーマ
	—	category	カテゴリ
	—	description	本文
	—	picture	画像
	—	evaluation emotion	評価値 感情値
WHY	—	—	(未採用)

に作成したものである。中村らは、ライフログに含まれる各データを5W1Hの観点から分類し、異なるライフログを1つのデータモデルに落とし込めるようにしている。

AHLEで利用しているデータモデルは、中村らの提案したデータモデルにいくつかの変更を加えてあり、また口コミに含まれる感情情報が有用であることが分かったため、他の値とは独立した、各感情値を格納するためのフィールドが用意されている。

このAHLEの実現には口コミから感情を抽出・分類するような仕組みが必要不可欠であり、先行研究では、AHLEにそのような仕組みを組み込む前段階として、口コミからの感情語の抽出・分類を行った。そこで本研究では次の段階として、抽出した感情語がどの程度正しく分類されているかの検証および、抽出した感情語を利用した観光地分類

の検討を行う。

## 4. 観光地に対する口コミからの感情語の抽出

### 4.1 調査対象

観光地の口コミに含まれる感情語の調査を行うため、観光 Web サイトであるトリップアドバイザー [12] に投稿された口コミを利用した。利用した口コミは、2014 年 4 月 22 日に登録されている全ての口コミを収集した。口コミの数は 47 都道府県で約 19 万件であった。今回は口コミに含まれる情報のうち、タイトルと本文から感情語の抽出を行った。

### 4.2 調査方法

今回の調査は、以下に示すような方法で行った。

- (1) ベースとなる感情語辞書を作成
  - (2) 感情語辞書に含まれる各単語について、日本語 WordNet[14] より類似した単語を取得、登録
  - (3) トリップアドバイザーより口コミを取得
  - (4) 取得した口コミのうち、調査対象となる部分に MeCab[13] を利用して形態素解析を行う
  - (5) 形態素解析によって得られた単語の原形と感情語辞書の各単語のマッチングを行う
- 1 および 2 については第 4.3 項、4 および 5 については第 4.4 項においてそれぞれ詳しく述べる。

### 4.3 感情語辞書の作成

感情語辞書を作成するにあたり、まず最初にベースとなる感情語辞書を手作業によって作成した。ベースとなる感情語の選定には [15] を参考にし、感情語の分類には Plutchik の感情の輪を参考にした。Plutchik の感情の輪は、感情を 8 つの基本感情（喜び、悲しみ、受容、嫌悪、恐れ、怒り、驚き、期待）に分類し、これら自身あるいはこれらの組み合わせによって感情を表現するモデルである。今回利用した感情語辞書ではそれに倣い、感情語を喜び、悲しみ、受容、嫌悪、恐れ、怒り、驚き、期待の 8 つのカテゴリに分類した。

ベースとなる感情語辞書を作成後、辞書に含まれる各単語について日本語 WordNet に問い合わせ、返ってきた単語を同じカテゴリに属する単語として追加することで、感情語の充実を図った。感情語の一例を表 2 に示す。

### 4.4 口コミの解析

口コミの解析は、MeCab を利用して形態素解析を行うことで文章を単語ごとに分割し、その結果求められた各単語の原形が感情語辞書に登録されている単語と一致した場合、対応するカテゴリの出現数を +1 する、という形で行った。ここで、形態素解析とは、文法や辞書などを元に、文を形態素（文字が意味を持つ最小の単位）に分解する技術

表 2 感情語の一例

カテゴリ	単語数	一例
喜び	411	喜ぶ、嬉しい、楽しい、良い
悲しみ	231	悲しい、痛い、がっかり、残念
受容	141	許容、許す
嫌悪	192	嫌い、苦手、後悔、辛い
恐れ	399	恐ろしい、心配、不安、問題
怒り	67	怒る、文句、クレーム
驚き	271	驚く、感心、びっくり
期待	265	期待、希望、欲しい

表 3 形態素解析の結果 (一部)

表層形	品詞	原形
期待	名詞	期待
通	名詞	通
、	記号	、
とても	副詞	とても
楽しかった	形容詞	楽しい
た	助動詞	た
です	助動詞	です
。	記号	。

のことであり、MeCab はそれに特化したオープンソース形態素解析エンジンである。

例文を使って解析時の動作の例を示す。例文には、“期待通り、とても楽しかったです。”という文を用いる。この文に対して形態素解析を行うと、表 3 に示すような結果を得ることができる。次に、ここで得た単語の原形が感情語辞書に登録されていないかを調べる。今回の例では、“期待”という単語が期待カテゴリに、“楽しい”という単語が喜びカテゴリにそれぞれ登録されているため、期待カテゴリと喜びカテゴリの出現数が 1 ずつプラスされることになる。

### 4.5 調査結果

先に述べたような調査を取得した口コミに対して行った結果を表 6 に示す。ここで、各感情の値は各カテゴリに属する感情語の出現数を示しており、表の下部には比較用として、先行研究において取得した、価格比較 Web サイトにおける調査結果を併記した。また、トリップアドバイザーにおける各感情値の出現率は、喜びが 57.3%、受容が 0.7%、恐れが 7.3%、驚きが 20.7%、悲しみが 4.5%、嫌悪が 3.1%、怒りが 0.3%、期待が 6.1%であった。価格.com では、喜びが 48.1%、受容が 3.5%、恐れが 12.3%、驚きが 8.8%、悲しみが 2.5%、嫌悪が 4.7%、怒りが 0.9%、期待が 19.2%であった。coneco.net では、喜びが 46.2%、受容が 4.4%、恐れが 12.6%、驚きが 7.3%、悲しみが 2.7%、嫌悪が 7.7%、怒りが 0.5%、期待が 18.5%であった。

この調査によって、各都道府県における各感情の出現率の傾向があまり変わらないことが分かった。また、喜びカテゴリに属する感情語の出現数が一番高いことは変わらな

かったが、他のカテゴリの出現率が価格比較 Web サイトと観光 Web サイトで違うことが明らかになった。一例としては、価格比較 Web サイトでは恐れカテゴリと期待カテゴリの出現率が比較的高かったのに対し、観光 Web サイトでは驚きカテゴリの出現率が比較的高いことなどが挙げられる。

#### 4.6 感情分類精度の検証

先に示したような調査によって得られた口コミから、ランダムに 100 件選出し、それらの口コミから感情語を抽出した結果を、手作業によって検証した。検証を行った結果を表 4 に示す。また、検証の際に見られた感情語の誤分類の例を表 5 に示す。この結果から、今回の手法を用いて分類した感情語は、一部誤りがあるものの、高い割合で正しく分類されていることが確認できた。この検証によって得られた感情語の誤分類のうち、特に多かったのが“結構”の誤分類である。これはおそらく、感情語辞書では“大いに結構”といった肯定のニュアンスの“結構”が登録されたが、実際には“そこそこ”というニュアンスの“結構”が多く使われているためであると考えられる。他には、“見上げる”が期待カテゴリ、“丸”が喜びカテゴリ、“取り組む”が嫌悪カテゴリ、“注文”が期待カテゴリといったように、そもそも感情語として適さないような単語が抽出されている場合もあった。このような単語を感情語辞書から外していくことで、精度の向上が実現できると考えられる。その単語単体では感情の分類として正しい場合でも、“～ない”というように否定語がつくことで逆の意味になってしまうケースも少なからず見受けられ、今後、感情語の分類精度を上げていく際には考慮が必要であると考えられる。

#### 5. まとめ

本研究では、観光地に対する口コミからの感情語の抽出およびその分類精度の検証を行い、自動的に分類した感情語が高い割合で正しく分類されていることを確認した。このように分類された感情語を観光地ごとにまとめることで、各観光地ごとに感情ベクトルを利用した分類や感情語を利用したカテゴリ化などを行うことができると考えられる。各観光地を感情ベクトルを利用して分類すると、似たような感情傾向を持つ観光地同士がグルーピングされることが期待され、ユーザの感情に合わせたレコメンドなどの実現に役立てることができると考えられる。

今後の課題としては、感情語辞書の充実や感情ベクトルを用いた観光地分類の実装、感情抽出モジュールの AHLE への組み込みなどが挙げられる。今回の検証では、抽出した感情語の分類が間違っているかどうかのチェックを行うことにより、間違った単語が感情語として抽出されることを防ぐことはできるが、原文には含まれているが抽出されていない感情語を減らすことはできていないため、その

チェックを行い、感情語辞書の充実を図る必要があると考えられる。

#### 参考文献

- [1] Fellbaum, C.: WordNet: An Electronic Lexical Database, MIT Press (1998).
- [2] Mitchell P. Marcus, Beatrice Santorini, M. A. M.: Building a Large Annotated Corpus of English: The Penn Treebank, Computational Linguistics, Vol. 19, No. 2, pp. 313-330 (1993).
- [3] 中村匡秀, 下條 彰, 井垣 宏: 異なるライフログを集約するための標準データモデルの考察, 電子情報通信学会技術研究報告, Vol. 109, No. 272, pp. 35-40 (2009).
- [4] 中山記男, 江口浩二, 神門典子: 感情表現の抽出手法に関する提案, 電子情報通信学会技術研究報告, Vol. 104, No. 416, pp. 13-18 (2004).
- [5] Stone, P. J.: The General Inquirer: A Computer Approach to Content Analysis, MIT Press (1966).
- [6] 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, Vol. 47, No. 2, pp. 627-637 (2006).
- [7] 徳久良子, 乾健太郎, 松本裕治: Web から獲得した感情生起要因コーパスに基づく感情推定, 情報処理学会論文誌, Vol. 50, No. 4, pp. 1365-1374 (2009).
- [8] 杉本祐介, 菱田隆彰, 水野忠則: インターネット上の書き込みに含まれる感情についての調査, 情報学ワークショップ 2013 (WiNF2013), pp. 195-199 (2013).
- [9] 杉本祐介, 土井千章, 中川智尋, 太田賢, 稲村浩, 水野忠則, 菱田隆彰: 口コミデータを活用するデータベースシステムの実現, 情報処理学会研究報告, モバイルコンピューティングとユビキタス通信 (MBL), Vol. 2014-MBL-70, No. 44, pp. 1-6 (2014).
- [10] オセニック株式会社: 価格比較サイト [coneco.net] コネコネット, オセニック株式会社 (オンライン), 入手先 (<<http://www.coneco.net/>>) (参照 2014-01-31) .
- [11] 株式会社カカクコム: 価格.com, 株式会社カカクコム (オンライン), 入手先 (<<http://kakaku.com/>>) (参照 2014-01-31) .
- [12] TripAdvisor LLC: ホテル・旅行の口コミ — トリップアドバイザー, TripAdvisor LLC (オンライン), 入手先 (<<http://www.tripadvisor.jp/>>) (参照 2014-01-31) .
- [13] 工藤 拓: MeCab: Yet Another Part-of-Speech and Morphological Analyzer, 京都大学 (online), available from (<<http://code.google.com/p/mecab/>>) (accessed 2014-01-31).
- [14] NICT 独立行政法人情報通信研究機構: 日本語 WordNet, NICT 独立行政法人情報通信研究機構 (オンライン), 入手先 (<<http://nlpwww.nict.go.jp/wn-ja/>>) (参照 2014-01-31) .
- [15] 日本語表現インフォ制作委員会: 感情表現インフォ: ビッタリの気持ちの描写が探せる言葉の辞書, 日本語表現インフォ制作委員会 (オンライン), 入手先 (<<http://hyogen.info/depa/kanjo>>) (参照 2014-01-31) .

表 4 検証結果

ロコミ番号	出現数	誤分類数	ロコミ番号	出現数	誤分類数
#1	6	1	#37	2	0
#2	4	0	#38	2	0
#3	4	1	#39	1	0
#4	4	0	#40	1	0
#5	4	1	#41	1	0
#6	4	0	#42	1	0
#7	4	2	#43	1	0
#8	4	0	#44	1	0
#9	4	0	#45	1	0
#10	3	1	#46	1	1
#11	3	0	#47	1	0
#12	3	1	#48	1	0
#13	3	0	#49	1	0
#14	3	0	#50	1	1
#15	3	0	#51	1	1
#16	3	0	#52	1	0
#17	2	0	#53	1	1
#18	2	0	#54	1	0
#19	2	1	#55	1	1
#20	2	0	#56	1	0
#21	2	0	#57	1	0
#22	2	0	#58	1	0
#23	2	1	#59	1	0
#24	2	0	#60	1	0
#25	2	0	#61	1	0
#26	2	0	#62	1	0
#27	2	0	#63	1	0
#28	2	0	#64	1	1
#29	2	0	#65	1	0
#30	2	1	#66	1	0
#31	2	1	#67	1	0
#32	2	1	#68	1	0
#33	2	0	#69	1	0
#34	2	0	#70	1	1
#35	2	0	#71-#100	0	0
#36	2	0	合計	135	19

表 5 感情語の誤分類の例

原文（一部抜粋）	感情語	分類
駐車場から目的地まで結構歩きましたが	結構	喜び
見上げる空を送電線が遮り	見上げる	期待
結構面白いです	結構	喜び
結構な高さのダムを上からも	結構	喜び
駐車スペースは結構混んでますが	結構	喜び
結構混んでいました	結構	喜び
飽きない程度に紅葉が楽しめます	飽きる	悲しみ
トマトやキュウリが丸のままおいてあったり	丸	喜び
注意したほうが良いと思います	いい	喜び
不自然に並んでいます	不自然	嫌悪
積極的に取り組まれている企業です	取り組む	嫌悪
天候が良ければ	良い	喜び
前回の満足度が高く、2度目の訪問です	満足	喜び
結構、自由に投げれます	結構	喜び
親子丼を注文したところ	注文	期待
良く仕事で通りますが	良い	喜び
それほど大きくはありませんので	大きい	驚く
静かな公園を訪ねるなら良いのでは	良い	喜び

表 6 調査結果

都道府県名	口コミ数	喜び	受容	恐れ	驚き	悲しみ	嫌悪	怒り	期待
愛知県	5805	5613	69	582	1780	432	264	27	463
愛媛県	1529	1599	22	195	522	107	108	9	142
茨城県	1502	1400	23	194	587	101	61	9	146
岡山県	1641	1628	8	163	573	149	62	8	177
沖縄県	15985	15086	194	2465	5598	1452	924	64	1667
岩手県	1317	1183	25	169	489	119	100	7	194
岐阜県	2278	2314	24	255	817	150	137	9	240
宮崎県	1685	1647	26	200	642	143	90	4	266
宮城県	1617	1583	18	183	524	163	84	6	145
京都府	7295	6882	74	808	3579	495	308	40	772
熊本県	2565	2528	46	261	947	182	117	14	261
群馬県	2287	2358	42	289	870	177	157	10	235
広島県	3375	2887	60	736	1507	299	240	17	530
香川県	1830	1766	27	302	739	152	187	10	189
高知県	1376	1249	13	219	659	241	100	1	180
佐賀県	793	745	14	103	297	82	45	3	114
埼玉県	2147	2059	21	220	594	125	83	10	159
三重県	4038	4485	37	481	1285	319	217	16	518
山形県	958	835	14	114	349	76	74	3	89
山口県	1551	1474	24	138	590	92	62	6	179
山梨県	2827	2821	31	347	1029	230	165	12	265
滋賀県	1537	1413	21	172	575	110	88	3	166
鹿児島県	4491	3540	60	673	1909	324	274	14	553
秋田県	933	777	10	109	327	86	39	3	89
新潟県	1991	2011	24	225	579	163	118	8	176
神奈川県	11603	12884	143	1360	3873	910	646	56	1053
青森県	2145	1938	21	314	837	166	119	6	211
静岡県	7004	7524	92	901	2189	543	401	30	707
石川県	2776	3059	34	285	1101	221	138	18	328
千葉県	5304	5708	90	671	1608	500	276	42	1049
大阪府	6051	6492	74	786	2037	465	298	24	576
大分県	2140	2293	36	319	824	173	109	6	254
長崎県	3289	3105	50	518	1306	283	203	11	476
長野県	5756	5986	86	697	2132	475	388	30	528
鳥取県	1185	1287	23	200	515	98	86	6	128
島根県	2307	2047	48	271	1016	171	114	12	344
東京都	28592	28845	331	3068	9564	1826	1285	117	2647
徳島県	1039	1013	12	279	540	77	61	3	138
栃木県	2992	2810	23	394	1164	268	181	15	306
奈良県	2989	2417	49	456	1581	193	162	18	335
富山県	1059	1032	10	149	474	96	84	6	107
福井県	1186	1096	13	209	521	107	76	9	130
福岡県	4298	4154	56	462	1239	279	180	17	341
福島県	1568	1453	27	225	515	180	118	8	162
兵庫県	5590	5749	72	586	1693	418	313	21	544
北海道	16448	16456	186	1969	5419	1315	776	72	1608
和歌山県	2517	2296	26	327	1012	215	150	10	257
全体	191191	189527	2429	24049	68527	14948	10268	850	20144
価格.com	4027	18109	1311	4634	3308	953	1759	323	7218
coneco.net	2000	11831	1136	3225	1857	699	1962	136	4745