5L-04

Twitterからの土産に関するレビュー情報の抽出

真辺 諒† 長尾 哲志† 安藤 一秋‡
†香川大学大学院工学研究科 ‡香川大学工学部

1. はじめに

旅行に行った際，多くの人が土産を購入するが，土産の選定は悩ましい問題である．土産の選定時に，現地でしか購入できない商品や定番商品，注目商品およびそれらの口コミ情報などがわかれば，この状況を改善できると考える．

そのような中，長尾ら[1]は，現地でしか購入できない土産に関する情報を QA サイトや口コミサイト，ブログから自動収集して，ユーザに提示するシステムを提案した．しかし，長尾らが対象とする土産情報は，比較的静的な情報であるため，リアルタイム性や話題性に関する情報が不足する．

そこで本研究では，リアルタイム性や話題性など，土産に関する動的な情報を補うための情報源として，Twitter に注目する．Twitter は即時性や簡便性の高さから多くの人が利用し，日々大量の情報を発信しており，発信される情報には，土産に関する情報も存在している可能性がある．また，美味しさ感覚を表す言葉である「シズルワード」にも注目する．シズルワードを利用することで，Twitter から土産に関するレビュー情報を取得できる可能性があると考える．

本稿では，土産に関するレビュー情報の抽出におけるシズルワードの有用性について調査する．具体的には，SVM（Support Vector Machine）を用いて土産に関するレビューツイートを抽出することで，シズルワードの素性としての有用性を確認する．

2. シズルワードを含むツイートの分析

2.1 分析対象とするシズルワード

大橋ら[2]は，「おいしそう」，「食べたい」を感じる言葉や，飲食の欲求を喚起する言葉を「シズルワード」と定義し，さらに「うまみのある」「香ばしい」などの味覚系表現，「もちもち」「サクサク」などの食感系表現，「新鮮な」「季節限定」などの情報系表現の 3 分野に分類した．本稿では，味覚系表現 70 語，食感系表現 78 語，情報系表現 74 語のシズルワードを対象にして，ツイート分析する．

2.2 土産に関するツイートの収集

Twitter Search API を利用して，キーワード「土産」または「みやげ」を含むツイートを収集する．2016 年 7 月 1 日から 2017 年 7 月 31 日の 1 年間に渡って収集した結果，土産に関する 1,360,548 件のツイートが収集できた．以降，このツイート集合を利用する．

2.3 シズルワードを含むツイートの抽出

収集したツイート 1,360,548 件から，シズルワードを含むツイートを抽出して分析する．抽出したツイートを人手で調査した結果，1,360,548 件中 147,149 件のツイートにシズルワードが含まれていた．また，以下のシズルワードは，他のシズルワードと比べて，ツイート本文中の出現数が多いことがわかった．

- 味覚系シズルワード：美味（55,727 件），甘い（5,868 件），辛い（3,239 件）
- 食感系シズルワード：しっとり（1,299 件），サクサク（1,016 件），ふわふわ（717 件）
- 情報系シズルワード: こだわり（18,284 件），グルメ（5,671 件），話題（2,727 件）

加藤ら[3]は，シズルワードと食品の関連性の分析において，Twitter 上の食に関するツイートの多くが，ユーザが実際に食した時の感想であることを確認している．したがって，これらを含むツイートには，土産に関するレビュー情報が含まれる可能性が高いと考えた．以降，これらのシズルワードを含むツイートを分析する．

2.4 シズルワードを含むツイートの分析

2.3 で示したシズルワードを含むツイートをそれぞれ 100 件ずつランダムに抽出し，レビューツイートが含まる割合を調査する．

表 1 にシズルワードを含むツイートに含まれるレビューツイートの割合を示す．表 1 に示すように，味覚系シズルワードである「美味」を含むツイートには，6 割近くのレビューツイートが含まれていることがわかる．また，「甘い」，「辛い」を含むツイートには，それぞれ 100 件中 25 件のレビューツイートが存在していた．食感系シズルワードを見ると，2 割から 4 割り程度の割合でレビューツイートが含まれていた．情報系シズルワードは，レビューツイートの件数が他の分野のシズルワードに比べて極端に少ないことがわかった．

これらの結果から，味覚系・食感系シズルワードを含むツイートは土産の感想を表現するレビュー情報に含まれる可能性があり，シズルワードは Twitter から土産に関するレビュー情報を抽出する手がかりとして有用であると考える．また，情報系シズルワードは，食品を宣伝するような言葉が多く，商品紹介サイトや購入サイトに誘導するツイートに多く含まれていたことから，土産に関するレビュー情報の抽出には利用しない．

表 1 . シズルワードを含むツイート 100 件に含まれるレビューツイート数

味覚系シズルワード	レビューツイート数	食感系シズルワード	レビューツイート数
美味	59/100	しっとり	27/100
甘い	25/100	サクサク	21/100
辛い	25/100	ふわふわ	38/100
情報系シズルワード	レビューツイート数		
こだわり	1/100		
グルメ	4/100		
話題	5/100		

Extraction of Reviews of Souvenirs from Twitter
Manabe Ryo† Nagao Noriyuki† Ando Kazuaki‡
†Graduate School of Engineering, Kagawa University
‡School of Engineering, Kagawa University

3. ノイズツイートのフィルタリング

シズルワードを含むがレビュー情報を含まないノイズツイートを分析した結果，bot による自動ツイートと外部サイトへの誘導を行う宣伝ツイートの 2 つが多く見られた．これらのノイズツイートに対し，フィルタリングする方法について述べる．

3.1 bot ツイートのフィルタリング

bot ツイートのフィルタリングでは，投稿数が多い「twittbot.net」，「Botbird tweets」，「rakubo2」の 3 つのクライアントからツイートされているものを bot とみなしてフィルタリングする．また，ユーザ名またはプロフィール欄に bot の文字が含まれているユーザの内，ユーザが手動でツイートしたと考えられる「Twitter for iPhone」，「Twitter for Android」，「Twitter Web Client」の 3 つのクライアントを除くクライアントからツイートされた場合，bot としてフィルタリングする．

100 件のツイートデータを対象に bot ツイートのフィルタリングを行った結果，F 値が 0.96 となり，効果的にフィルタリングできていることを確認した．

3.2 宣伝ツイートのフィルタリング

宣伝ツイートは，Web メディアの公式アカウントや商品宣伝を行うアカウントの場合が多く，これらは恒常的に URL 付きのツイートを投稿していることがわかった．そこで，URL 付きツイートを投稿しているユーザの過去 20 件のツイートを取得し，そのうち 15 件以上のツイートが URL 付きツイートの場合，フィルタリングする．

100 件の URL 付きツイートデータを対象に，宣伝ツイートのフィルタリング行った結果，F 値が 0.92 となり，効果的にフィルタリングできていることを確認した．

4. SVM によるレビューツイートの抽出

4.1 レビューツイートの分類手法

SVM を用いてレビューツイートの抽出を行う．「土産」または「みやげ」を含む 1,360,548 件のツイートデータに対して，3.のフィルタリングを行った結果，928,715 件のツイートが得られた．以降，これらのツイートを実験データとして用いる．SVM の素性には，Bag of Words（BoW）を用いる．ただし，品詞が，名詞，動詞，形容詞以外の単語は除外する．素性値には，TF-IDF を利用する．この手法をベースラインとして，ベースライン手法の素性にシズルワードを加えた以下の 2 手法と比較することで，シズルワードの有効性を確認する．

- 手法①：ベースライン手法にシズルワードの有無を素性として追加
- 手法②：ベースライン手法にシズルワードを素性として追加

追加する素性には，味覚系 70 種と食感系 78 種のシズルワードを利用する．また，味覚系のみ，食感系のみ，味覚系と食感系の両方の 3 通りで評価する．適合率，再現率，F 値を評価尺度とし，10 分割交差検証で評価する．学習データは，928,715 件のツイートに対し，人手でラベリングした土産に関するレビューツイート 300 件，その他のツイート 300 件の合計 600 件のツイートを用いる．

4.2 実験結果

分類結果を表 2 に示す．なお，表中の適合率，再現率，F 値は Positive／Negative クラスの平均を示す．表 2 を見ると，手法①で味覚系/食感系シズルワードを利用した場合が最も正答率が高く，ベースライン手法より 3.2 ポイント向上している．手法②で味覚系/食感系シズルワードを利用した場合，ベースライン手法より正答率は向上したが，0.8 ポイント差であった．

味覚系と食感系シズルワードを比較すると，食感系シズルワードのみを利用した場合，手法②ではベースライン手法より正答率が下がった．一方，味覚系シズルワードのみを利用した場合，両手法でベースライン手法より正答率が向上した．これは，食感系シズルワードの出現数が少なく，味覚系シズルワードの方がレビューツイートと共起する可能性が高いことが原因と考えられる．

表 2. 十分割交差検定による分類器の評価

	適合率	再現率	F 値	正答率
baseline	0.741	0.740	0.740	74.0%
手法①（味覚系）	0.771	0.768	0.768	76.8%
手法①（食感系）	0.741	0.740	0.740	74.0%
手法①（味覚系+食感系）	0.774	0.772	0.771	77.2%
手法②（味覚系）	0.754	0.750	0.749	75.0%
手法②（食感系）	0.739	0.738	0.738	73.8%
手法②（味覚系+食感系）	0.752	0.748	0.747	74.8%

5. おわりに

本稿では，Twitter から土産に関するレビュー情報を取得することを目的に，シズルワードを含むツイートについて調査・分析を行った．その結果，味覚系・食感系のシズルワードを含むツイート中に土産に関するレビューが多く含まれることがわかった．そして，SVM によるレビューツイートの分類器を構築し，ベースライン手法と比較することで，シズルワードの素性としての有効性を確認した．シズルワードを素性に利用することで，正答率が向上することを確認した．

今後は，テキスト内の単語のシズルワードとの共起関係を調査し，土産に関するレビューツイートの抽出性能の向上を目指す．

参考文献

[1] 長尾哲志，安藤一秋：オンラインショップで購入できない土産を提示するシステムの構築，第 14 回情報科学技術フォーラム講演論文集，pp.71-72，2015.

[2] 大橋正房，武藤彩加，山本眞人，爲国正子，汲田亜紀子，渋澤文明，小川裕子：「おいしい」感覚と言葉 食感の世代，BMFT 出版部，2010

[3] 加藤大介，宮部真衣，荒牧英治，灘本明代：シズルワードに着目した Twitter 上のおいしさの表現の分析，第 12 回日本データベース学会年次大会，B6-6，2014

[4] 川島崇秀，佐藤哲司，神門典子：Twitter からの消費者ニーズの抽出手法に関する提案，第 14 回日本データベース学会年次大会，B5-1，2016