

マイクロブログにおける“おいしさ”に言及したシズルツイートの自動抽出

森田 真季[†] 荒牧 英治^{††} 灘本 明代^{†††} 宮部 真衣[†]

[†]和歌山大学システム工学部

^{††}奈良先端科学技術大学院大学研究推進機構

^{†††}甲南大学知能情報学部

1 はじめに

飲食店を検索する際、多くのユーザがインターネットサイトを利用しており、飲食店情報の検索を主としたサイトが普及している¹。しかしマイボイスコム²の調査によると、このような検索サイトに関して、「欲しい情報に辿り着かない」「掲載情報が古い」等の不満が挙げられており、使いやすさの上で問題点を抱えている^{2,3}。そのため、“ユーザが必要としている情報”を“素早く”提示する仕組みが必要である。

「もちもち」や「サクサク」等、食品のおいしさを表す表現としてシズルワードがある。シズルワードと飲食店を結びつけることで、飲食店に対してより詳しいおいしさの情報を与えることができる可能性がある。また、近年、Twitterなどのマイクロブログで様々な情報発信がなされており、飲食店についての最新の情報も含まれていることが想定される。マイクロブログはリアルタイム性が高く⁴、マイクロブログを用いることで、「掲載情報が古い」という、従来の課題を解決できる可能性がある。

そこで本研究では、シズルワードを含み、おいしさについて言及しているツイート（シズルツイート）を用いた飲食店提案システムの構築を目指し、Twitterからのシズルツイートの自動抽出を試みる。

2 関連研究

手塚ら [1] は、地理位置情報、時間情報、ツイートの内容でグルーピングを行い、マッピングを行うシステムを提案している。マッピングの精度向上の為、ツイート本文から特徴語を抽出し、グルーピングを行っ

ている。ツイートから必要なツイートを分類するという点で本研究と関係している。

渡辺ら [2] は、味覚や触感を表すオノマトペを用いた料理レシピ検索システムを提案している。このシステムではオノマトペを用いることで、ユーザの持つ曖昧なニュアンスを検索に反映することができる。本研究とは、使用する単語や、推薦対象が異なる。

3 シズルツイートの自動抽出

3.1 シズルワードとシズルツイート

前述したように、シズルワードとは、「もちもち」や「サクサク」等、食品のおいしさを表す表現である [3]。本研究ではシズルワードを含み、おいしさについて言及しているツイートを「シズルツイート」と呼ぶ。シズルツイートを元に、飲食店の情報と、言及されているおいしさを結びつけることができれば、飲食店に対してより詳しいおいしさの情報を与えることができる。

しかし、シズルワードを含むツイートがシズルツイートであるとは限らない。例えば、「ふわふわ」というシズルワードを含むツイートには「パンケーキ」のような食品についての言及もあれば、「タオル」のような食品に関係しないツイートも存在する。つまり、単純にシズルワードをキーワードとして検索するだけでは、不要なツイートも蓄積されることになるため、おいしさに言及しているツイートのみを抽出する必要がある。そこで本研究ではシズルツイートの自動分類器を構築する。

3.2 コーパス

本手法で使用するシズルワードは、大橋 [3] らの定義する「おいしさを感じる言葉」を参考に選定した。シズルワードの3領域（食感系・味覚系・情報系）のうち、今回は「ふわふわ」のように、食品以外にも共起しやすい可能性のある食感系を対象とした。さらに、食感系のシズルワードの上位30件が含まれるツイート数を、位置情報付きのツイートデータ⁵を対象として調査し、取得されたツイート数の上位6件から「ジューシー」を省いた、「ふわふわ」「もちもち」「サクサク」「しっとり」「とろける」の5語を調査対象とすること

Automatic Extraction of Sizzle Tweets referring to the "Taste" in Microblog Posts

Maki MORITA[†] Eiji ARAMAKI^{††} Akiyo NADAMOTO^{†††} Mai MIYABE[†]

[†]Faculty of Systems Engineering, Wakayama University

^{††}Nara Institute of Science and Technology

^{†††}Faculty of Intelligence information, Konan University

¹飲食店情報の検索に関するアンケート調査（第6回）：http://myel.myvoice.jp/products/detail.php?product_id=19313

²「飲食店の検索」についての調査：https://www.smartsurvey.jp/board/press_view/129/60

³飲食店情報の検索に関するアンケート調査（第4回）：http://myel.myvoice.jp/products/detail.php?product_id=16909

⁴「テレビの見方が変わる」～ツイッターの利用動向に関する調査～：https://www.nhk.or.jp/bunken/summary/research/report/2010_08/100806.pdf

⁵2011年7月15日から約1年間収集された、24,817,908件のデータである。

表1: 「ふわふわ」に関するコーパス例

正例/負例	ツイート例
+1	茶房古蓮で宇治しぐれ。氷がふわふわでうまうま。 http://twitpic.com/5suser
-1	めちゃ気持ちよかったー。ふわふわの緑の絨毯！ http://t.co/knW8evK

表2: 各コーパスのツイート数と正例数

シズルワード	ツイート数	正例数
ふわふわ	998	409
もちもち	369	336
サクサク	466	383
しっとり	1169	330
とろける	844	679

とした¹⁾。位置情報付きツイートデータから、ツイート本文に指定シズルワードを含むツイートを取得し、それらがシズルツイートであるかどうかを3~4人で判定してコーパスを作成した。表1は「ふわふわ」というシズルワードに関する作成したコーパスの一部である。正例、負例については評価者の多数決結果を採用した。各コーパスのツイート数と正例数は表2の通りである。

3.3 シズルツイート分類器の構築

3.2節で述べたコーパスを用いて、シズルツイートであるかを判定する分類器を構築した。今回は、素性として以下の項目を用いた。

- シズルワードの周辺文脈（形態素）
- 食材，食品，料理名の有無

形態素解析にはJuman[4]を用いた。また、食材，食品，料理名の有無は、代表的な食材，食品，料理が含まれる辞書（以下、本稿では食品辞書と記す）を用いて判断している。これらの素性を用いて、TinySVM²⁾により分類器を構築した。素性は形態素（1-gram）を用いて、SVMによる学習には多項カーネル（d=2）を用い、パラメータはデフォルト値を用いた。

4 実験

構築した5つのシズルワードに関する分類器の精度を検証する為に、5分割交差検定を行った。表3は各シズルワードにおける全形態素を用いた場合の5分割交差検定の結果である。「しっとり」以外は適合率、再現率、F値のいずれも0.7を越えており、高精度に分類できることが分かった。「しっとり」についてはF値が0.6を越えたものの、再現率は0.6を僅かに下回った。

周辺文脈のウィンドウサイズを1~10まで変化させた場合の、F値の変化を表4に示す。どの単語において

¹⁾ 「ジュシー」はシズルツイートの割合が非常に高く、分類の必要性が低いと判断し、除外した。

²⁾ TinySVM: <http://chasen.org/~taku/software/TinySVM/>

表3: 全形態素を用いた場合の5分割交差検定結果

シズルワード	適合率	再現率	F 値
ふわふわ	0.842	0.704	0.766
もちもち	0.934	0.958	0.945
サクサク	0.880	0.982	0.927
しっとり	0.823	0.587	0.685
とろける	0.927	0.907	0.916

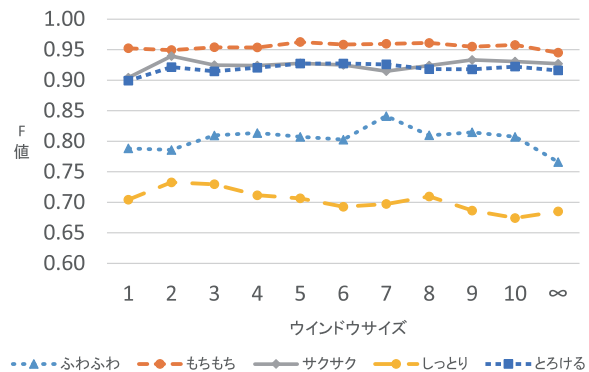


図1: ウィンドウサイズ1~10でのF値の変化

もウィンドウサイズによる大幅なF値の変化は見られなかったが、相対的にF値が低い「ふわふわ」「しっとり」においては、単純に全形態素を用いるよりも、周辺文脈のみを用いることで精度が向上する可能性がある。

5 おわりに

本研究では、シズルツイートを用いた飲食店提案システムの構築に向けて、シズルツイートの自動抽出が可能か調査した。「ふわふわ」「もちもち」「サクサク」「しっとり」「とろける」の5語について、それぞれ分類器を構築し、5分割交差検定により判定精度を検証した。「しっとり」以外は適合率、再現率、F値のいずれも0.7を越えており、高精度に分類できることを確認した。相対的にF値が低い「ふわふわ」「しっとり」においては、周辺文脈のみを素性とすることで精度が向上する可能性がある。

今後は、今回用いていないシズルワードを対象とした分類、精度を検証する。

参考文献

- [1] 渡辺知恵美, 中村聡史: オノマトペロリ: 味覚や食感を表すオノマトペによる料理レシピのランキング, 人工知能学会論文誌, Vol.30, No.1, SP2-N, pp.340-352 (2015).
- [2] 手塚悠太, 服部哲, 速水治夫: 時間軸と特徴語によるツイートマッピング, マルチメディア, 分散協調とモバイルシンポジウム 2011 論文集, pp.408-413 (2011).
- [3] 大橋正房, シズル研究会: 「おいしい」感覚と言葉 食感の世代, 株式会社 B・M・FT 出版部 (2010).
- [4] Daisuke, K. and Kurohashi, S.: A Fully-Lexicalized Probabilistic Model for Japanese Syntactic and Case Structure Analysis, In Proceedings of HLT-NAACL2006, pp.176-183 (2006).