

マイクロブログを用いた飲食店の“おいしさ”情報収集手法の構築

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

飲食店検索において、インターネットサイトを利用するユーザが増加しており、飲食店情報の検索を主とした Web サイトも普及している¹⁾。しかし、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、使いやすさの上で課題点を抱えている²⁾。

「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 [1] がある。シズルワードと飲食店を結びつけることで、飲食店で提供されている料理の食感や匂いなどのおいしさ情報をわかりやすく表現し、ユーザが求める情報として提供できる可能性がある。また、Twitter などのマイクロブログで発信される様々な情報には、飲食店についての情報も含まれていると考えられる。マイクロブログはリアルタイム性が高く³⁾、マイクロブログから情報を収集・利用することにより、「掲載情報が古い」という課題を解決できる可能性がある。

これまでに我々は、シズルワードとマイクロブログに着目した飲食店検索システムを提案してきた [2]。提案システムは、シズルワードに基づいて飲食店に関するおいしさ情報を蓄積する機能と、蓄積した情報をユーザが検索できるようにする機能から成る。提案システムの実現には、各店のおいしさに関する情報の収集と蓄積が必要である。そこで本稿では、おいしさ情報を収集・蓄積する手法について述べる。

2 関連研究

鈴木ら [3] は、オノマトペを用いたグミ推薦システムを提案している。このシステムは、グミと関連の高

いオノマトペを収集し、同義語をまとめた上で、グミと関連付けて推薦を行う。オノマトペを用いることで、グミの微妙な食感の違いを表現している。

また、渡辺ら [4] 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。

これらの研究は、飲食に関する情報検索において、味覚などを表すことば（オノマトペ）を利用している点で本研究と類似するが、本研究ではオノマトペに限定しない、より広い概念であるシズルワードを利用した飲食店検索システムの実現を目指す。

3 飲食店に関するおいしさ情報の収集手法

本研究では、飲食店に関するおいしさ情報の収集先として、マイクロブログの一つである Twitter を利用する。本研究ではシズルワード [1] 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」と定義する。飲食店に関するおいしさ情報は、シズルツイートと、そのツイートで言及される飲食店を結びつけることにより収集・蓄積する。なお、本研究で収集する「おいしさ情報」は、シズルワードおよびツイート内容とする。情報蓄積の3つのステップを以下に示す。

ステップ1: Twitter からツイートを収集する。

ステップ2: 収集したツイートからシズルツイートを抽出する。

ステップ3: 抽出したシズルツイートをもとに、おいしさ情報を蓄積する。

ステップ1では、Twitter からシズルワードを含むツイートを取得する。検索するシズルワードは、大橋らの定義する「おいしさを感じる言葉」[1] を参考に選定した。シズルワードはおいしさの感じ方によって3領域（食感系・味覚系・情報系）に分けられており、各領域の「おいしさを感じる言葉ランキング」の上位から、食感系30語、味覚系29語、情報系31語⁴⁾の計90語を用いた。

ステップ2では、ステップ1で取得したツイートから、SVMにより構築した分類器（シズルツイート分類

⁴⁾味覚系は有用ではなかった「美味」を除外した。また、情報系は同率順位のシズルワードが含まれているため、31語となっている。

Development of a Method for Collecting Information on “Palatability” in Restaurants using Microblogs

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

[†]Faculty of Systems Engineering, Wakayama University

^{††}Institute for Research Initiatives, 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

³⁾“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

表 1: 「サクサク」における提案手法の情報収集結果

ツイート内容	ツイート数 (件)
ステップ 1: 「サクサク」を含む (A 群)	29,908
ステップ 2: シズルツイートと判定された (B 群)	9,938
ステップ 3: 飲食店名が取得された (C 群)	2,896
シズルツイート内から取得	888
位置情報をもとに取得	16
シズルツイート以外の 発信者のツイートより取得	1,992

器) によって、シズルツイートを抽出する。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「このパンケーキふわふわしてる」のような食品についての言及もあれば、「新品のタオルはふわふわだ」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、おいしさに関係しない、不要なツイートも蓄積されてしまうため、シズルツイートのみを抽出する。領域単位のシズルツイート分類器を構築した結果 [2]、いずれの分類器も、比較的高精度 (F 値 > 0.6) に分類可能であったため、この分類器を用いる。

ステップ 3 では、ステップ 2 で抽出したシズルツイートを解析し、おいしさ情報をデータベースに蓄積する。蓄積される情報は以下の 4 つである。

- A: おいしさ情報 (シズルワード, ツイート内容)
- B: 飲食店名
- C: 位置情報
- D: ツイート情報 (発信者 ID, 投稿日時)

飲食店名については、まず、CRF⁵ を用いて、シズルツイート内から抽出を試みる。抽出されなかった場合、ツイートを付与されている位置情報をもとに、Google プレイス検索⁶ を用いて位置情報付近の飲食店を検索し、飲食店名を取得する。ツイートを位置情報が付与されていない、または Google プレイス検索で取得できない場合、発信者のツイートのうち、シズルツイートの発信前後 2 時間の範囲で投稿されたものを収集し、ツイート内に含まれる飲食店名を取得する。

4 実験

3 章で述べたおいしさ情報の収集手法により、情報の蓄積がどの程度可能かを調査した。対象ツイートは 2016 年 12 月 20 日～2016 年 12 月 26 日に収集された、本研究で対象とするシズルワード (90 語) のいずれかを含む 1,540,362 件のツイートである。

今回は、食感系シズルワードの一つである「サクサ

ク」を対象に、提案手法の各ステップを実行し、蓄積された情報について検証を行った。「サクサク」を用いた情報収集の結果を表 1 に示す。

表 1 より、「サクサク」を含むツイート 29,908 件に対し、飲食店名が取得されたツイートは 2,896 件であった。ここで、各ステップで取得されたツイートを、ステップ 1 から順に A 群、B 群、C 群とする。

A 群のうち、無作為抽出した 200 件のツイートについて、シズルツイートであるかどうかを人手で評価し、シズルツイート分類器の自動判定の精度を検証した。人手評価の結果、200 件中 63 件が正例、137 件が負例であった。判定精度の検証結果は、適合率 0.788、再現率 0.825、F 値 0.806 となり、ステップ 2 ではシズルツイートが高精度に分類できていることがわかった。

また、B 群から無作為抽出したツイート 200 件について、ツイート内に飲食店名が含まれるかどうか調査したところ、25 件 (12.5%) のツイートの飲食店名が含まれた。しかし、B 群全体で飲食店名が取得されたのは 9,938 件中 888 件 (8.9%) であり、無作為抽出時の結果と比較すると、十分に取得できていない可能性がある。ただし、200 件のツイートの確認結果を見ると、そもそも飲食店名が含まれるツイートの割合が低いため、CRF や位置情報を利用する以外の飲食店名取得手法を検討する必要がある。

5 おわりに

本稿では、シズルワードとマイクロブログを用いた飲食店検索システム実現のための、おいしさ情報の収集手法について述べた。提案手法による情報抽出実験の結果、以下の 2 点の知見を得た。

1. 「サクサク」というシズルワードを含むツイートについては、高精度 (F 値 0.806) にシズルツイートかどうかを判定可能である。
2. 飲食店名や位置情報が含まれるツイートの割合は低いため、飲食店名取得方法を再検討する必要がある。

今後は、「サクサク」以外のシズルワードを対象とした精度検証を行う。

参考文献

- [1] 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) .
- [2] 森田真季, 荒牧英治, 灘本明代, 宮部真衣:“おいしい”情報自動抽出におけるシズルワードの比較分析, 信学技報, Vol.116, No.78, IEICE-NLC2016-11, pp.53-58 (2016) .
- [3] 鈴木彰真, 野々村翔, 村田嘉利:オノマトベを用いた客観性のあるグミ推薦サービスの有用性評価, マルチメディア, 分散協調とモバイルシンポジウム 2014 論文集, pp100-107(2014).
- [4] 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ベ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) .

⁵CRF++(<https://taku910.github.io/crfpp/>) を用いた。

⁶<https://developers.google.com/places/web-service/search?hl=ja>