

飲食情報多角化のためのマイクロブログにおける「おいしくない」情報の抽出 Extraction of “Not Tasty” Information in Microblog for Diversification of Food and Drink Information

下津 拓未¹ 平林 (宮部) 真衣² 吉野 孝³
Takumi Shimotsu Mai Miyabe Hirabayashi Takashi Yoshino

1. はじめに

インターネットを通じた様々な情報が飛び交う現代社会において、情報の提供や発信、共有などが盛んに行われている。SNS 上で投稿・拡散される様々な情報の中には、消費者にとって有益な情報も多数含まれる。この点に関しては、飲食に関しての情報提供媒体であるグルメサイトも例外ではなく、飲食店や飲食に関する口コミなどは消費者にとって有益な情報源となり得る。実際に、飲食店の検索において約 8 割近くの人がグルメサイトを閲覧するといった調査結果が挙げられている¹。これらの口コミにおいて、大橋らが定義したおいしさを想起させる「シズルワード」が利用されており、消費者にとって有益な情報源となっている [1]。我々はこれまでに、シズルワードと対をなす、飲食に関する否定的感情表現として「アンチシズルワード」を定義し、その特徴分析を行った [2]。結果として、飲食に関する直接的否定表現である「不味い」や「おいしくない」といった語句を利用することでアンチシズルワードの抽出が可能であることが判明した。しかし、「不味い」「おいしくない」といった端的な表現は、飲食に対する不満以外の場面で使用される可能性がある。例えば、「この条件はおいしくない」や「不味い状況」といった記述は、飲食ではない場面で用いられると考えられる。したがって、「おいしくない」や「まずい」といった端的な表現が飲食に関連して用いられているかどうかを分類する必要性がある。

そこで本研究では、アンチシズルワード抽出のために、飲食との関連性の有無に関する調査を目的とし、Twitter²の投稿内容に対するテキスト分類を行う。これらの分類を行うことで、飲食に関する不満や不快感などといった否定的感情表現を含む情報の抽出についての検討を行う。また、本研究では、抽出の際に利用する検索クエリに着目し、アンチシズルワードの抽出において最適な検索クエリの検討を行う。

2. 関連研究

本章では、飲食物に関する感情表現の研究、Twitter に着目した飲食に関する研究について述べ、本研究の位置づけを明らかにする。

¹ 和歌山大学大学院 システム工学研究科, Graduate School of Systems Engineering, Wakayama University

² 東京大学大学院 医学系研究科, Graduate School of Medicine and Faculty of Medicine, The University of Tokyo

³ 和歌山大学 システム工学部, Faculty of Systems Engineering, Wakayama University

¹ グルメサイト離れが加速? テーブルチェックが飲食店と消費者に意識調査 (2020 年 1 月 10 日): <https://www.inshokuten.com/foodist/article/5622/>

² <https://twitter.com/>

2.1 飲食物に関する感情表現の研究

食品の口コミにおいて、おいしさ表現の一つであるシズルワード [1] は重要な役割を持っている。シズルワードとは、大橋らによって定義された人においしさを喚起させる言葉である。適切なシズルワードを用いることで、人々に味を想起させやすくすることが出来るため、これまでにシズルワードを用いた様々な研究が行われている。

加藤らは、Twitter, オンラインショッピングサイトなどを含む Web サイト, レシピサイトで使用されているシズルワードを含むテキスト同士の比較分析を行い、Web サイトでは、Twitter に比べて情報系のシズルワードの使用率が高いことを明らかにした [3]。森田らは、シズルワードを飲食検索サービスにおいて利用することで、消費者の嗜好を考慮した飲食店情報の提示を行った [4]。

このように飲食物の味の好みに着目した様々な研究は行われているが、「おいしくない」ことを想起させる表現についての分析を行っている研究はない。本研究では、飲食に対する「おいしくない」ことを想起する表現を含む文章を対象に分析を行う。

2.2 Twitter に着目した飲食に関する研究

Twitter では、最新の多種多様な情報が飛び交っているため、重要な情報が多く存在している。このことから、Twitter に着目した分析がこれまでに行われている。

下室らは、Twitter 上の食品名に関するツイートに着目し、ユーザの食事嗜好に考慮した食事メニュー推薦システムを提案した [5]。Drajad らは、ツイートデータからレストランの口コミを抽出し、価格や味などといった 4 つの側面に分類することで、レストランにおける課題発見を試みた [6]。

このように、Twitter 上から飲食に関する内容に着目した研究は行われているが、否定的表現を含む飲食物に関するツイートに対する分析は行われていない。本研究では飲食に関する否定的感情に着目し、ツイートの分類を行う。

3. 手動によるテキスト分類

本研究では、分析対象データとして代表的なマイクロブログの一つである Twitter を利用する。Twitter は、飲食と消費者の行動に関する有益なデータソースになる可能性が高く [7]、他の SNS に比べ、ネガティブな投稿が多いことが調査結果として挙げられている [8]。これらの点から、Twitter は、飲食における消費者の客観的かつ否定的な感情を含む意見が多く、様々な表現や単語を収集が可能であると考えた。

3.1 アンチシズルワード

飲食に関する「おいしさ」を想起させる表現であるシズルワードの対をなすものとして、我々は「おいしくない」ことを意図して用いられる表現を「アンチシズルワード」と定

義した。これまでに我々は、飲食に関して「不味い」や「おいしくない」といった感情を抱いた原因や要素をアンチシズルワードと見なし、Twitter のテキストの中からアンチシズルワードの抽出・系統分類を行った。抽出した一部のアンチシズルワードを表 1 に示す。この調査では、「不味い」や「微妙」などの飲食に関する直接的否定表現を検索し、その中からアンチシズルワードの抽出を試みており、Twitter 上からアンチシズルワードの抽出が可能であることを明らかにした。一方で、これらの直接的否定表現は、飲食以外の場面でも用いられることがあるため、飲食との関連性が低いものも抽出されてしまう可能性がある。そこで本章では、「まずい」や「おいしくない」といった直接的な否定表現を含むテキストを分析し、その傾向について考察する。

3.2 分析の概要

飲食における直接的な否定表現として、まず 21 種類の表現を検索クエリとして Twitter を検索した結果、10,000 件以上のツイートを取得できた計 9 種類の単語 (表 2) を含むツイートを分析対象として用いることとする。

今回取得したデータは、2017 年 5 月～2022 年 5 月の期間にツイートされたものである。なお、リツイートやリプライのテキストデータは対象から省いている³。以下の手順で分析を行った。

- (1) 検索クエリを含むツイートを収集
9 種類の検索クエリごとに最大 100,000 件のツイートデータを収集した。
- (2) ツイートデータに対する前処理
重複した内容のテキストや Twitter 特有の文字である URL やハッシュタグ、絵文字などは、分類において不必要であるため、削除した。
- (3) テキストの形態素解析
形態素解析器 MeCab⁴を使用し、テキストを原形にし、表記ゆれをなくした。
- (4) ランダムサンプリング
抽出したデータから、目視による分類が可能な範囲のデータ数として検索クエリごとに 1,000 件をランダムサンプリングした。
- (5) テキストの 2 値分類
ランダムサンプリングしたデータから飲食に関連する否定的内容が記載されているものを目視で判断し、2 値分類を行った⁵。今回は、「飲食に関する否定的な感情表現が含まれている内容」を正例、「それ以外の内容」を負例として分類を行った。

3.3 分類結果と考察

抽出したデータ数と分類結果を表 3 に示す。表 3 から、分類においてある程度均等性が見られたクエリと偏りが見られたクエリに 2 分化された。本節では、それぞれについて考察する。

³ リツイートやリプライでは、データの重複が多数確認されたため

⁴ <http://taku910.github.io/mecab/>

⁵ 今回は、主観による評価のみを行ったが、今後は客観性を考慮するために第三者による分類を行う

- (1) 均等性が見られた検索クエリにおける傾向
均等性が見られたクエリとしては、「不味」「おいしくない」「美味しくな」「無味」が挙げられる。「不味」や「美味しくな」「無味」といった漢字表記の「味」が含まれているものは、比較的飲食に関連する内容が存在していることが判明した。これらの点から、今後使用する検索クエリとして、「味」が付く否定的な表現を利用することで、アンチシズルワードを抽出できる可能性が高くなると考えられる。
- (2) 偏りが見られた検索クエリにおける傾向
偏りが見られた検索クエリとしては、「まず」「びみょう」「ビミョウ」「微妙」が挙げられる。「まず」に関しては、「まず～しよう」といった副詞として使用されていたツイートがほとんどであった。これらの点から、今回は「まず」による抽出を行ったが、今後は「まずい」や「まずく」といった検索クエリの拡張による抽出の可能性が考えられる。これにより、副詞として利用される内容を選べることができ、飲食に関する否定表現を多く抽出できる可能性があると考えられる。また、「びみょう」「ビミョウ」「微妙」に関しては、「微妙な反応」といった飲食に関連しないものや「びみょうな違い」といったそもそも否定的な表現として使用されていないものが多数確認された。これらの 3 つの検索クエリに関しては、「まず」から「まずい」といったように検索クエリとして拡張が困難であるため、今後検索クエリとして利用するためには、「びみょう」のみを含むテキストを抽出するのではなく、他の飲食に関連する単語と「びみょう」の両方を含むデータの収集が考えられる。例えば、「びみょう」と「食べ物の名前」の両方を含むツイートを収集することで、「びみょう」のみの抽出に比べ、飲食に関する否定的情報である可能性が高くなると考えられる。

4. 機械学習によるテキスト分類

本章では、3. 章で利用した検索クエリごとに分類したテキストデータをもとに機械学習による 2 値分類を行う。

4.1 素性

3.2 節の手順で作成した形態素解析済みのデータから、必要な品詞の選定を行った。今回は、文章の構成として重要であると考えられる「名詞」「動詞」「形容詞」に加え、否定語である「ない」を含む「助動詞」を含めた計 4 種類の品詞のみのデータとし、それ以外の品詞に関しては削除する。また、今回使用する素性としては、以下の 3 種類である。

- (1) One-hot
文章に単語が含まれているか否かの「0」「1」のみで構成されたバイナリ型の学習データ。
- (2) Bag of Words
文章中に多く使用された単語は、その文章の重要なキーワードとなっていると仮定し、使用された単語の回数の特徴量として考慮したカウント型の学習データ。

表 1: 事前分析で抽出したアンチシズルワードの一部

分野	アンチシズルワード
味覚系	ムラがある, 脂っこい, はっきりしない, クリーミーすぎ, あっさりしすぎ, 独特の風味, こってり
食感系	パサパサ, 口に残る, シャキシヤキ, ゴムみたい, ネバネバ, ドロドロ, ネチャネチャ, ほぐれにくい
情報系	物足りない, 値段相応, 口に合わない, コンビニ飯, 生, 染み込んでいない, ゼロ, 存在感がない

表 2: 検索クエリ

まず, マズ, 不味, びみょう, ビミョウ, 微妙, おいしくな, 美味しくな, 無味

(3) TF-IDF

ある単語の出現頻度である TF とその単語が出現する文書数の逆数である IDF を乗算した値を特徴量として考慮した学習データ。

4.2 SVM による分類結果と考察

3.2 節の手順で作成された訓練データを用いて, SVM による学習を行い, 5 分割交差検証による分類精度の評価を行った。表 3 の結果から, 「マズ」や「微妙」などのラベルに偏りが見られたデータに関してはそのままのデータを使用してしまうと, うまく学習されない可能性が考えられる。そのため, 学習の際に「マズ」「びみょう」「ビミョウ」「微妙」の 4 種類の検索クエリに関しては, 少ない正例クラスを正しく分類させるために, 重みづけを行うこととする。本稿では, SVM のモデルとして, 線形型モデルを用いた。精度評価の指標として, 「適合率」「再現率」「F 値」を用い, 交差検証で出力された 5 回分の評価値の平均値を算出する。4.1 節の 3 つの素性ごとに「適合率」「再現率」「F 値」の算出結果を表 4 に示す。結果に対して, 「検索クエリ」と「素性」の 2 点に着目し, 考察を行う。

(1) 検索クエリ

結果として, 「不味」「おいしくな」「美味しくな」「無味」といった分類が均等にされていた検索クエリにおいては 0.7~0.8 と高い精度であった。一方で, 偏りがあった検索クエリとして重みづけを行った「マズ」や「びみょう」「ビミョウ」などにおいては, 「適合率」が高く, 「再現率」が低い値として算出された。本研究では, テキスト中から多様なアンチシズルワードを抽出することを目指しているため, 飲食を対象とした否定的な内容の文章の検出漏れを防ぐ必要がある, 「適合率」よりも「再現率」を重視している。よって, 「マズ」や「びみょう」「ビミョウ」などにおいては, 「適合率」と「再現率」のトレードオフの関係性から, 「再現率」をより考慮することで今後検索クエリとして利用が可能であると考えられる。また, 「びみょう」「ビミョウ」「微妙」では, 同じ読み方の単語で, 正例と負例の割合が似ていたが, 精度においては「微妙」のみが低い精度であった。このように, 似たような単語においても表記によって精度に差がある可能性があるため, 精度の低い検索クエリにお

いても表記を変化させ, 新たな検索クエリとして利用することで, 分類精度が高いクエリとしての利用可能性があると考えられる。

(2) 素性

素性に関する大きな差は見られなかった。この結果は, Twitter の特性である文字数制限による情報量の少なさが原因であると考えられる。また, 検索クエリにおいて 3.3 節で均等性が見られた「不味」「おいしくな」「美味しくな」といったクエリでは, F 値においては TF-IDF 型がより高い精度であることが判明した。

5. おわりに

本研究では, 飲食に関する否定的感情表現を含む情報の抽出において使用する検索クエリに対する検討を行った。飲食に関する端的な否定を表す表現として「おいしくな」や「微妙」を検索クエリとしたツイートに対して, 飲食に関する否定的表現を含む内容の有無について機械学習を用いた分類を行った。分類の結果, 以下の点を明らかにした。

手動による分類

- (1) 「不味」「おいしくな」「美味しくな」「無味」などにおいては, 飲食に関する否定的感情表現を含む内容が多く抽出された。
- (2) 「まず」や「微妙」などに関しては, 飲食との関連性のない内容を含むツイートが多く抽出されたことから, 検索クエリの文字数を増やすといった拡張や同時に使用する検索クエリの増加などが必要であると考えられる。

機械学習による分類

- (1) 手動による分類結果として均等性が見られた検索クエリにおいては, SVM による分類評価から, 精度の高い自動分類が可能であると考えられる。
- (2) 手動による分類結果として偏りが見られた検索クエリにおいては, SVM による分類評価として十分な精度が見られなかった。精度向上のために今後は, データサンプリングによるデータの偏りを均等にす手法を利用する。

今後は, 検索クエリから抽出されたツイートデータを対象に, 機械学習を用いたアンチシズルワードの自動抽出を行う。

表 3: 抽出した検索クエリごとの手動による分類結果

検索クエリ	まず	マズ	不味	びみょう	ビミョウ	微妙	おいしくな	美味しくな	無味
抽出数	80,325 件	71,422 件	78,782 件	34,067 件	15,329 件	90,976 件	24,564 件	83,444 件	68,145 件
正例	15 件	216 件	646 件	81 件	49 件	45 件	437 件	654 件	317 件
負例	985 件	784 件	354 件	919 件	951 件	955 件	563 件	346 件	683 件

表 4: 検索クエリごとの SVM による精度評価の結果

	適合率			再現率			F 値		
	One-hot	BoW	TF-IDF	One-hot	BoW	TF-IDF	One-hot	BoW	TF-IDF
マズ	0.71	0.72	0.78	0.46	0.45	0.48	0.55	0.55	0.59
不味	0.74	0.76	0.71	0.78	0.79	0.91	0.76	0.77	0.80
びみょう	0.58	0.51	0.57	0.34	0.39	0.32	0.39	0.41	0.35
ビミョウ	0.60	0.57	0.67	0.16	0.16	0.11	0.25	0.25	0.19
微妙	0.38	0.35	0.30	0.12	0.11	0.05	0.18	0.17	0.08
おいしくな	0.74	0.74	0.75	0.59	0.61	0.73	0.66	0.67	0.74
美味しくな	0.79	0.80	0.75	0.79	0.76	0.95	0.79	0.78	0.84
無味	0.75	0.76	0.86	0.66	0.67	0.58	0.70	0.71	0.69

※ 「まず」に関しては、正例が極端に少なかったため省く

謝辞

本研究の一部は、JSPS 科研費 18K11555 および 21K12138 の助成による。

参考文献

- [1] 大橋正房, 光岡裕子, 須藤正彦, 北村奈津子, 米山真知子, 猿渡一兵: SIZZLE WORD シズルワードの現在「おいしいを感じる言葉」調査報告, 株式会社 B・M・FT 出版部 (2018).
- [2] 下津拓未, 平林(宮部)真衣, 吉野孝: アンチシズルワード: 個人に最適化した飲食情報提供のための飲食物に対する負の感情表現の分析, マルチメディア, 分散, 協調とモバイル (DICOMO2022) シンポジウム, 3F-1, pp.557-562 (2022).
- [3] 加藤大介, 宮部真衣, 荒牧英治, 灘本明代: インターネット上のメディア毎の「おいしさ」表現比較分析, DEIM Forum, F5-5, pp.1-8 (2015).
- [4] 森田真季, 宮部真衣, 荒牧英治, 灘本明代, 吉野孝: 嗜好とリアルタイム性を考慮した飲食店検索システムの構築, 2018 年度 情報処理学会関西支部 支部大会 講演論文集, B-0, pp.1-3 (2018).
- [5] 下室孝平, 灘本明代, 荒牧英治, 宮部真衣: 食事メニュー推薦のためのマイクロブログを用いたユーザの食事嗜好・履歴推定手法の検討, 情報処理学会第 79 回全国大会, No.4, pp.369-370 (2017).
- [6] Dradjad Bima Ajipangestu, Riyanarto Sarno: Sentiment Analysis Based on The Aspect of Culinary and Restaurant Review Using Latent Dirichlet Allocation and Support Vector Machine to Improve the Profitability of Culinary Business and Restaurant in Surabaya, Proceedings of the 3rd International Conference on Business and Management of Technology, Advances in Economics Business and Management Research, Vol.202, pp.80-86(2021).
- [7] Luis G.Moreno-Sandoval, Carolina Sanchez-Barriga, Katherine Espindola Buitrago, Alexandra PomaresQuimbaya, Juan Carlos Garcia: Spanish Twitter Data Used as a Source of Information About Consumer Food Choice, CD-MAKE 2018, LNCS 11015, pp.134-146(2018).
- [8] 村益有那, 寶雪: ソーシャルネットワークサイト上における若者の自己開示と感情表現に関する研究—Twitter と Facebook の内容分析及び大学生へのインタビュー調査から—, 第 31 回情報通信学会大会予稿集, A-1, pp.1-32 (2014).