

ぐるなび API を用いた否定語が含まれる口コミの調査

2017SC051 西村真瑠人 2017SC083 常賀翔一朗

指導教員：河野浩之

1 はじめに

近年では、自分で飲食店を探す人々が増え、食への関心が高まっている傾向にある。ぐるなびは1996年に誕生したグルメサイトである。グルメサイトの会員数に着目してみるとぐるなびが約1885万人とグルメサイトを利用している人は少なくない。

リサーチ会社・マイボイスコムは、8月1日から5日にかけて「飲食店情報の検索に関するアンケート調査」を実施した。計4,575件のデータを用いて「テキストマイニング」を活用した分析を実施しており、4大グルメサイトの性質が明らかとなっている [1]。

このような飲食店検索サイトに関して、「欲しい情報が得られない」、「お店に対する信頼度がない」等の不満が挙げられており、サイトに対する問題点が上がっている [2]。そのため、口コミやマイクロブログから情報を収集して必要な情報を調査する研究は数多くされてきた。吉野らは、リアルタイム性の高いマイクロブログと美味しさを表すシズルワードに着目した飲食店情報推薦システムの開発と実験を行っている [3]。

しかし、飲食店に対して否定的な言葉が口コミに含まれているか調査する研究はなされていなかった。私達は普段生活する上で否定的な言葉に対して抵抗がある。例えば、飲食店の口コミに「まずい」という言葉があった場合おそらく良い印象を抱く人はいない。「まずい」のような飲食店に対する否定的な言葉が口コミにあったら私達はそのお店には行かないはずだ。

そこで本研究の目的として否定語を設定し、その言葉が口コミに含まれているか調査を行う。そのため、本研究の流れとしてまずグルメサイトからテキスト（レビューや口コミなど）を抽出する。抽出する手段としてぐるなび API を使用し、テキストを集める。その収集したテキストを MeCab を用いて調査する否定語があるか確認する。そして指定した地点から口コミに否定的な言葉が含まれる飲食店はどれくらいあるのか調査を行う。

本研究論文は全6章で構成される。2章では先行研究について、3章では提案するシステムの構築や使用するアプリの比較を述べる。4章ではアルゴリズムを用いてシステムの概要を説明する。5章で実験の評価、6章で本研究をまとめる。

2 口コミ収集に関する関連研究

この章では本研究の関連研究について紹介する。美味しさを表す言葉を抽出することに関する先行研究とテキストマイニングに関する先行研究を述べる。2.1節、2.2節では、推薦システムを用いた先行研究の比較を行う。2.3節では、

グルメサイトから否定語を抽出する理由を述べる。

2.1 ふんわりサーチ：個人の嗜好を考慮した「おいしさ」情報提示による飲食店推薦システム [3]

吉野らは、外食時の店選びで、「料理の味」が最も重視されると考えた。そこで飲食店情報の提供サービスにおいて、「おいしさ」の情報として、「もちもち」や「サクサク」等の食品のおいしさを表す語句である「シズルワード」に着目した。また、マイクロブログの一つである Twitter から、シズルワードを含むツイートを抽出することにより、おいしさに関する情報を取得店選びの参考になる情報として提供できる可能性がある、と考えた。ユーザに合った「シズルワード」を設定し、Twitter から抽出することによって、ユーザの「行きたい」と考える飲食店を推薦した。シズルワードにより、既存システムでは注目しなかった飲食店に対して、興味を持つきっかけを与えられる。

今後の課題として、検索できる店舗数の増加、検索速度等のユーザビリティの向上、既存の飲食店検索システムとの組み合わせによる有用性の向上であった。

2.2 テキストマイニングの手法の活用による観光者の経験に着目した観光地の理解構造の変化に関する研究 [4]

泉澤らは、地域経済の活性化を観光振興によって進めようと計画する都市が少なくない、という点の背景として観光が広域産業の経済活性化のプラットフォームになりうるものがあげられる。観光地に対する認知は主にメディアから発信される観光情報によって観光者の中にあらかじめかたちづくられる。ある特定の観光地における経験を経た観光者には、自身の中にある特定の観光地に対する新たな理解の構造が構成されていると考えられ、そうであるならば、今度新たに構成された理解の構造をもとに観光経験が組み立てられると言える。

2.3 API を用いた否定語が含まれる口コミの調査する理由

最近では、インターネットの普及と共に、インターネット上に様々なウェブサイトが存在する。様々なウェブサイトの中でも、大きく発達しているのが飲食を扱っているグルメサイトである。例を述べると、「食べログ」や「ぐるなび」などの飲食店情報の検索サイトである。人間の生活において「食」は非常に重要なパーツであることが言えることから、近年では、若者から高齢者まで9割以上がグルメサイトを使用している。しかし、グルメサイト使用者の意見としては、「グルメサイトの数が増加しすぎてしまったことでどのグルメサイト使用すればいいかわからない」という意

見が増えてきている [1]. そこで、否定語が含まれる口コミを調べることで、飲食店の情報を入手することができる考えた。

3 ぐるなび API を用いた否定語が含まれる口コミの調査で扱うツール

第 3 章では本研究で使用するツールを紹介する。3.1 節では本研究で使用するグルメサイト、3.2 節では、本研究で使用する API、3.3 節では本研究で使用する形態素解析を紹介する。

3.1 ぐるなびの特徴

ぐるなびは昔から親しまれているグルメサイトで、4 大グルメサイトの中では最も老舗のグルメサイトである。口コミを投稿する際に写真を添えて投稿することができる機能や、店から口コミに対して返信することができる機能があることで客の来店ハードルを下げ、実際に店への集客につながりやすい機能を実装している点が大きな魅力である。グルメサイトの中でも昔から愛用しているユーザが多いことから多くの信頼性があるサイトでもある。

3.2 ぐるなび API の特徴

ぐるなび API の主な機能を以下に紹介する。

1. レストラン検索 API は、ぐるなびに掲載されている飲食店の基本情報（住所など）や緯度経度、設備情報、クーポン URL、画像 URL など取得することができる。
2. 多言語版レストラン検索 API は、ぐるなび外国語版に掲載されている飲食店の基本情報（住所など）や緯度経度取得することができ、5 言語対応（日本語、英語、中国語：繁体字、中国語：簡体字、ハングル）している。
3. 応援口コミ API は、ぐるなび店舗ページの応援口コミタブに掲載されているぐるなび会員の皆様の投稿画像 URL や口コミ情報などを取得することができる。

それぞれの機能を使うときには対応したアクセスキーと URL を python 上で記入することで、得たい情報を得ることができる。本研究では、応援口コミ API を使用して実験を行った。他のグルメサイトには、API 機能がないものや口コミを収集する機能がなかったためである。なお、無料のぐるなび API サービスは 2021 年 6 月 30 日に終了となる。

3.3 推薦の手法

文書分析を行うツールとして、「MeCab」、「Unidic」、「Janome」の 3 つが挙がった。

「MeCab」は C++ によって書かれた形態素解析ツールである。未知語に対しては定義を変更可能なため、ノイズの削除が可能である。また判別精度の高さ、実行速度も早さも特徴である。

「Unidic」は国語研が現代日本語書き言葉均衡コーパスにおいて用例収集がしやすくなるように定義した「短単位」

を元に構築した辞書である。検索システムにおいては、単語分割の一貫性が高い。自然言語処理で構文解析や意味解析などをする上では分割が短過ぎて使いにくい。

「JUMAN++」は形態素解析器としては MeCab の方が多機能だが、依存構造などを解析できる KNP が使えるというのは MeCab に無い利点である。RNN 言語モデルを使うことで非常に高精度な形態素解析を実現している。

本研究では、MeCab を使用する。理由として比較した他の形態素解析よりも反応速度が速く、汎用性が高いためである。また今回使用する python との相性も考え使いやすいため選んだ。

4 ぐるなび API を用いた否定語が含まれる口コミの調査のアルゴリズム

ここでは、アルゴリズムを用いて本研究の流れを説明する。4.1 節では先行研究と本研究の使用するツールの比較を行った。4.2 節ではアルゴリズムを用いて本システムの流れ、4.3 節では本研究の否定語についてを説明する。

4.1 先行研究と本研究の比較

先行研究と本研究のデータ収集方法、形態素解析、口コミを収集する際に使用する API の比較を表 1 に表す。

吉野らは、シズルワードを用いて口コミを収集したのを利

表 1 先行研究 [2] と本研究の比較

	吉野ら	本研究
データ収集	Twitter よりシズルワードを含むツイートを収集	ぐるなびの口コミを収集
形態素解析	なし	Mecab
API	Twitter API	ぐるなび API

用して、本研究では、シズルワードの代わりに否定語を用いて口コミを収集することにした。

4.2 ぐるなび API を用いた否定語が含まれる口コミの調査のアルゴリズム

今回作成するシステムのアルゴリズムを図 1 に表す。まず、調べたい地点の緯度経度を python プログラムに入力する。本研究では栄と八事で調査を行う。

次に範囲を設定をする。範囲は、半径 300m、500m、1000m、2000m、3000m と指定ができ、毎回範囲を変えて繰り返し行って口コミを収集する範囲を広げていく。

API はぐるなび API を使用して飲食店に対する口コミを抽出する。ぐるなび API のアクセスキーを取得した後に、python を用いてプログラムを作成して、口コミデータを取り込めるものを作成する。

その後、口コミをデータ化して MeCab で分かち書きにする。分かち書きにすることで、探したい否定語を捜しやすくするためである。その分かち書きにした後、コメントの中

に否定語があるか確認する。

結果を踏まえてその周辺のお店のコメントに否定語が含まれるかを確認し、どれくらいの範囲で飲食店の口コミに否定語が含まれているか検証する。初めに半径 300m から調査していき、範囲を広げて半径 3000m の栄と八事の飲食店の口コミを収集して、否定語の数を調べたら終了となる。

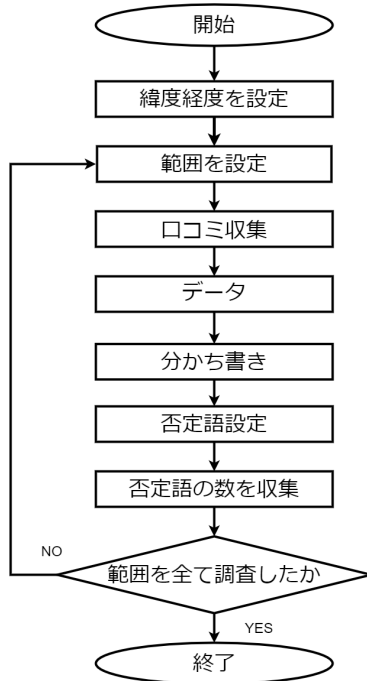


図 1 本研究のアルゴリズム

4.3 飲食店に対する否定語

本研究では、先行研究で使用していた「シズルワード」を参考に、否定語を設定した。シズルワードとは、「おいそう」、「食べたい」、「飲みたい」と感じ、飲食欲求を喚起する言葉を表す。他にも「もちもち」、「ジューシー」、「とろける」、「濃厚」、「うまみのある」、「贅沢な」、「揚げたて」、「焼きたて」などさまざまな言葉がある。否定語とは、「良くない」、「まずい」、「悪い」、「嫌い」など否定を表現する言葉である。

そこで本研究では「良くない」と「まずい」という店や食事に対する否定語を選び、その言葉がコメントに含まれるか検証する。なお、この要旨には「良くない」という否定語が含まれる口コミの調査のプログラムを記述した。

5 ぐるなび API を用いた否定語が含まれる口コミの調査の実験結果

ここでは、ぐるなび API を用いた否定語が含まれる口コミの調査の実行結果を記す。5.1 節では python プログラムと実行結果、5.2 で考察を行う。

5.1 ぐるなび API の python プログラム

本研究のプログラムを Listing 1 のぐるなびの python プログラムで説明する。

一行目の keyid にぐるなび API のアクセスキーを記した。2 行目の hit-per-page は取得する最大件数を指定するもので、1 回のリクエストで得る最大投稿件数のことを意味する。上限は 50 件であり、本研究では多くのデータを取得したいため、50 件取得するようにした。

4 行目の latitude は緯度、5 行目の longitude は経度を表したもので、本研究では栄にあるオアシス 21 と南山大学付近の駅で八事駅の場所を指定した。緯度、経度を調べる際に google map を使用して取得した。下記のもの、オアシス 21 の緯度経度である。7 行目の range は GPS 検索対象半径を設定するもので、単位は m で表し、緯度と経度からの検索範囲 (半径) を示し、以下は半径 300m で指定した。

次に、11 行目の def review-search から口コミを検索する関数を作成した。response でそれぞれの検索の答えを返し、15 行目の total-hit-count でヒットした飲食店の口コミをカウントする。

Listing 1 python プログラムの一部

```
1 keyid = "*****"
2 hit_per_page = 50
3
4 latitude = 35.1705798
5 longitude = 136.9094715
6
7 search_range = 1
8
9 query = {"keyid": keyid, "latitude":
          latitude, "longitude":longitude, "range":
          search_range, "hit_per_page":
          hit_per_page}
10
11 def review_search(query):
12     rev_list = []
13     res = json.loads(requests.get("https://
          api.gnavi.co.jp/PhotoSearchAPI/v3/",
          params=query).text)
14     if "gnavi" not in res:
15         total_hit_count = res["response"]["
          total_hit_count"]
16 }
```

次に、Listing 1 で出た実験結果を MeCab を用いて分かち書きにするためにデータ化する作業を行う。実験結果で出た口コミをテキストファイルに保存して、分かち書きになるようにコマンドプロンプトで実行した。その時の分かち書きされた結果の一部を Listing 2 に表す。4 行目の comment の部分が飲食店の口コミコメントである。

Listing 2 分かち書きされた実行結果の一部

```

1 ' areaname _ m ': ' 栄周辺',
2 ' areaname _ s ': ' 錦3 丁目',
3 ' url _ 320 ': ' https :// mr . gnavi . co
  . jp / cont / menu _ image / 4 b / 15
4 ' comment ': ' コーヒーの抽出方法が掲示して
  あった が 見え ずらく 良くない と
  思った。 ',

```

分かち書きにした後, python 上で指定した単語が口コミの中にあるか確認する. Listing 3 は否定語が含まれる口コミの有無を確かめたプログラムを表した.

本研究では使用する否定語は「良くない」に指定し, UTF8N に変換したものを読み込み, lines というリストに実行結果より口コミが含まれたテキストファイルの各行を読み込んだ.

3 行目の for でリストから要素を 1 行の文字列を取り出して, in でその文字列に検索対象の文字が存在するか調べる. もし検索文字が含まれていたら, index を調べて表示する. 実行すると, 否定語の「良くない」が含まれていなかったら何も返ってこない状態となり, 含まれていた場合はその言葉が含まれている列番号を返すプログラムとなっている.

Listing 3 の 6 行目は実行結果であり, 54 という結果が出力された. Mecab によって分かち書きされた実行結果の 54 行目に否定語が 1 つあったと解釈する. よってぐるなびの口コミから 1 つだけ否定語を含むコメントが出てきたこととなる.

Listing 3 実行結果

```

1 with open('aaa.txt',mode='r',newline'',
  encoding='utf-8')as f_in:
2     line = [line for line in f_in]
3 for i in lines:
4     if '良くない' in i:
5         print(lines.index(i))
6     54

```

本研究では, 栄と八事の緯度経度から範囲を 300m, 500m, 1000m, 2000m, 3000m の中の店の口コミを収集し, 否定語が含まれているかどうかを行った. 実験結果を表 2 に表す.

表 2 は, トータルでヒットした口コミ数から否定語が含まれる口コミの割合を表示した. 計算は小数点第 6 位を四捨五入して, 単位は%で表した. 栄からは, 半径 300m の範囲で 135 件中 0 件, 半径 500m で 426 件 0 件, 半径 1000m で 1154 件中 1 件, 半径 2000m で 1924 件中 3 件, 半径 3000m で 2874 件中 4 件の口コミから否定語を抽出することができた. 八事からは, 半径 300m で 17 件中 0 件, 半径 500m で 21 件中 0 件, 半径 1000m で 60 件中 0 件, 半径 2000m で 119 件 1 件, 半径 3000m で 229 件中 2 件の口コミから否定語を抽出することができた.

表 2 実験結果

	否定語が含まれる 口コミの割合 (%)	否定語が含まれる 口コミの割合 (%)
範囲	栄	八事
300m	0	0
500m	0	0
1000m	0.087	0
2000m	0.156	0.840
3000m	0.140	0.873

6 考察

今回の実験を通して, 栄と八事の周辺の飲食店に対して否定語が含まれる口コミの収集を確認することができた. また, トータルでヒットした飲食店の口コミと否定語が含まれる口コミの数の合計は, 栄が 6513 件中 8 件, 八事が 446 件中 3 件という結果になった. 割合にしてみると, 栄が 0.123 %, 八事が 0.673 % となり, 八事が栄より 0.55 % 否定語が含まれる口コミの割合が高いことが判明した.

7 結び

本研究では, ぐるなび API を用いて, 設定した範囲の中から飲食店に対する口コミを収集し, 得られた結果を MeCab で分かち書きして, 否定語が含まれている口コミの調査を行った.

また, 飲食店の情報として, 栄と八事周辺の飲食店は, 否定語が含まれる口コミの数の割合が栄で 0.123 %, 八事で 0.673 % とそれぞれ数値が低いことから, ユーザが否定的な印象を抱く飲食店が集まっていないことがわかった.

参考文献

- [1] 口コミラボ編集者:【徹底比較】飲食店が導入すべき「4 大グルメサイト」それぞれの強みは? 食べログ・ぐるなび・ホットペッパーグルメ・一休. com, <https://lab.kutikomi.com/news/2020/09/15/gourmetreservationsite/>, 参照 Jan 7, 2021.
- [2] 【飲食店情報の検索】に関するアンケート調査 (第 12 回), https://myel.myvoice.jp/products/detail.php?product_id=26508, 参照 Jan 7, 2021.
- [3] 吉野 孝, 森田 真季, 宮部 真衣: “ふんわりサーチ: 個人の嗜好を考慮した「おいしさ」情報提示による飲食店推薦システム,” 研究報告セキュリティ心理学とトラスト (SPT), 2019-SPT-33, pp. 1-8 (2019.5).
- [4] 泉澤 圭亮, 中鉢 令兒: “テキストマイニングの手法の活用による観光者の経験に着目した観光地の理解構造の変化に関する研究,” 日本観光研究学会機関誌, 2020-31-1, pp. 45-56 (2019.9).