# 飲食店レビューサイトにおける 再訪問・初訪問レビューの分類

# 白髮 宙海 $^1$ 村本 直樹 $^2$ 高橋 克郎 $^3$ 大島 裕明 $^4$

本研究では、飲食店レビューサイトにおいてユーザが投稿したレビュー文書に注目し、初訪問なのか再訪問なのかを分類する問題に取り組む、Web 上には商品やサービスを利用したユーザが情報を共有するレビューサイトがある。なかでも飲食店レビューサイトは広く一般に利用されており、味、価格、サービスなどについての飲食店のレビューが共有されている。テキストで記述されたレビュー文書を精査すると、初訪問のものと再訪問のものが混在していることが分かる。これらを自動的に分類することによって、飲食店の再訪問率の推定や、初めて訪れたユーザと、何度も訪れるユーザによる評価の違いの分析などが可能になると考えられる。

## 1 はじめに

本研究では、飲食店レビューサイトにおいてユーザが投稿したレビュー文書が、初訪問におけるものなのか再訪問におけるものなのかを分類する手法を提案する。Web 上には商品やサービスを利用したユーザや、利用を検討しているユーザが情報を共有するレビューサイトがある。そのなかでも飲食店のレビューが得られるサイトは複数存在し、多くの人々に広く利用されている。それぞれのサイトでは店名、食事のジャンル、営業時間、定休日、メニュー、住所(地図)などの情報が得られるほか、実際に飲食店を利用したユーザが投稿したレビューをみることができる。ユーザのレビューには、数値で表現される味・価格、接客・サービス、雰囲気、コストパフォーマンスなどの評価がある。さらに、飲食店を利用した感想などがテキストで記述されたレビュー文書が存在する。飲食店はこのような多数の項目で評価されている。

飲食店レビューには様々な観点からの評価がある一方,人々が飲食店を評価する観点はこれらに限定されない.たとえば,初めてでも入りやすい飲食店なのかどうかや,常連客に愛されている飲食店なのかどうかなどは,評価観点として興味深いと考えられる.そのような観点からの評価を行うためには,あるレビューが初訪問におけるものなのか再訪問におけるものなのかということを判別する必要がある.

たとえば、飲食店レビューサイトのひとつである食べ口グでは



図1 飲食店レビューサイトのページの例

#### https:

//tabelog.com/iwate/A0301/A030101/3010047/dtlrvwlst/B343076936/ 2019 年 1 月 5 日アクセス

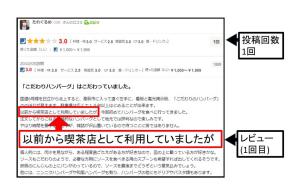


図2 投稿回数は1回だが再訪問であると推測できる例

https://tabelog.com/ibaraki/A0805/A080501/8001420/dtlrvwlst/ B116872188/ 2019 年 1 月 5 日アクセス

利用した飲食店に対して、複数回レビューを投稿することが可能である。図1の例では、投稿回数を見ることで、そのユーザがその飲食店を少なくとも二度利用したことが分かる。このように2回以上レビューが投稿されている場合には、2回目以降は再訪問の際のものであると判別できる。しかし、投稿回数が1回目のレビュー情報は、実際に読まなければ初訪問のものか再訪問のものかを判断することはできない。図2は投稿回数は1回であるが、そこでは「以前から喫茶店として利用していましたが、今回初めてハンバーグを食べに行ってきました」と記述されており、再訪問の際のものであると考えられる。

本研究ではレビュー文書から特徴抽出を行い、特徴ベクトルとして表現し、それを機械学習の分類器で学習することによって、初訪問におけるものなのか再訪問におけるものなのかの分類を行う。Bag-of-Words による特徴量、初訪問や再訪問のレビュー文書に現れやすい表現や語などに着目した特徴量、根拠文という初訪問か再訪問かの判断を決定づける文を利用した特徴量を用いて、どのような特徴ベクトルが、本研究での分類に貢献するかを明らかにした。



<sup>&</sup>lt;sup>1</sup> 非会員 兵庫県立大学 応用情報科学研究科 aa17j506@ai.u-hyogo.ac.jp

<sup>2</sup> 学生会員 兵庫県立大学 応用情報科学研究科

aa18c508@ai.u-hyogo.ac.jp

<sup>3</sup> 非会員 兵庫県立大学 応用情報科学研究科

ab18y501@ai.u-hyogo.ac.jp

<sup>&</sup>lt;sup>4</sup> 正会員 兵庫県立大学 応用情報科学研究科, 社会情報科学部 ohshima@ai.u-hyogo.ac.jp

# 2 関連研究

レビューに対して多様な観点から分析する研究は数多く行われ ている. Dave ら [2] は、Amazon などスコア付けされた商品レ ビューを使い, 商品への意見がネガティブ, ノーマル, ポジティブ のいずれであるかを推定する研究を行っている. Ding ら [4] は, Amazon の商品レビューで使われる文脈に注目し、そのレビュー が商品に対してポジティブな意見かネガティブな意見かを推定す る研究を行っている. Osman ら [11] は, ブログの文書から意見を 抽出する研究において、抽出する精度の高いコーパスを発見する 比較と検証の研究を行っている. Wijaya ら [13] は, Web 上の映 画のレビューを対象として, 使われている形容詞の規則性からポ ジティブ、ネガティブのスコア付けを行い、レビューから映画の ランキングを作る手法を提案した. Gilbert ら [7] は, Amazon の 商品レビューでは先に投稿したレビューを参照して投稿する追記 型のレビューが多いことに注目し、追記型のレビューを検出する 研究を行っている. Fayazi ら [5] は, Amazon の商品レビューで あからさまに高得点をつける欺瞞的なレビューを判別する研究を 行っている. Mullick ら [9] は, 新聞や Twitter や YouTube で投稿 される文書の内容が、意見か事実のどちらかであるかを分類する 研究を行っている. Cheng ら [1] は、Web 上の記事やスパムメー ルなどのリークされたデータセットを用いフェイクニュースを 発見する研究を行っている. Zagal ら [14] は、ゲームのレビュー サイトを利用し、どのようにゲームを楽しんだのかやユーザから の開発者への提案などのトピックを分類する研究を行っている. Mahony ら [10] は、TripAdvisor で書かれたレビューから、ホテ ルの選択に参考となったレビューだけを抽出する分類ベースの推 薦システムを研究している. Lappas ら [8] は,多くのレビュー が投稿されている際に、それらをすべて読むことが困難であると いう課題に対して, レビュー全体から選択と要約を行う研究を 行っている. Vuら [12] は、モバイルアプリのレビューにおける フレーズに注目しユーザの意見を抽出する研究を行っている.

# 3 再訪問レビューと初訪問レビューの分類問題

本研究で取り組む問題は,入力を飲食店レビューサイト上のレビュー文書として,出力を初訪問,再訪問,不明のいずれかのラベルとする3クラス分類問題である.

食べ口グでは、同一の飲食店に対し複数回のレビューの投稿が可能である。レビューは図1に示したように投稿した投稿回数と、それぞれのレビューに対して何回目に投稿されたものかが表示される。本研究では、投稿されたそれぞれのレビューにおいて、テキストで記載された部分をレビュー文書とよぶ。図1のレビューには、1回目に投稿されたレビュー文書と、2回目に投稿されたレビュー文書が存在することとなる。2回目以降に投稿されたレビュー文書はすべて再訪問レビューとなるため、本研究では1回目に投稿されたレビュー文書のみを対象とする。

以後,初訪問時に投稿されているレビュー文書を初訪問レビュー,再訪問時に投稿されているレビュー文書を再訪問レビューと呼ぶこととする.本研究で取り組む問題は,レビュー文書が初訪問レビューであるかの分類を行

うものである. ただし、レビュー文書の中には、人が読んでも初訪問時のものか、再訪問時のものか判別することができないものもある. そのようなレビュー文書については、不明というラベルを付ける. 手法の評価には、適合率、再現率、F値を用いる.

# 4 レビュー文書からの特徴ベクトルの作成

本節ではレビュー文書から特徴ベクトルを作成する手法を説明する. 我々は,以下の3種類の特徴量を組み合わせることで特徴ベクトルの作成を行った.

- 1. TF-IDF に基づく特徴量
- 2. 特有表現に基づく特徴量
- 3. 根拠文との類似性に基づく特徴量

1つ目の特徴量は一般的な文書検索で用いられるように、レビュー文書を Bag-of-Words として扱い、語に対応する次元の重みを TF-IDF によって重みづけするものである。2つ目の特徴量は初訪問や再訪問を特徴付ける語や言い回しを含むかどうかによるものである。3つ目の特徴量は再訪問レビューや初訪問レビューと判断できる文との類似性に基づくものである。

#### 4.1 TF-IDF に基づく特徴量

TF-IDF に基づく特徴量は、レビュー文書を Bag-of-Words と みなして、TF-IDF を重みづけとして特徴ベクトルを作成するも のである。まず、レビュー文書を形態素解析して、すべての語を 基本形にして分かち書きを行う。我々は、形態素解析器として MeCab を用いた。IDF を計算するためには、文書の全体集合の 範囲を決定する必要がある。今回は、学習データとして利用する レビュー文書集合を文書の全体集合とすることとした。その際、学習データのレビュー文書で1つのレビュー文書にしか現れない 語は無視することとした。

後述するデータセット全体において, 語の総種類数は 11,889 であった. このままでは次元数が大きいため, LSA [3] によって, 得られた特徴ベクトルを 300 次元に圧縮した. ただし, その際, 文書-単語行列のセンタリングを行うことで, LSA における第一特異ベクトルに相当する部分については利用しないこととした.

実験では、5分割交差検証を行った.そのため、交差検証の各回において、全レビュー文書集合のうちの80%が IDFを計算するための全体集合として用いられることとなる.そして、交差検証の各回において、TF-IDF 重みによる特徴ベクトルの次元数は多少変化することとなる.その後、LSA による圧縮が行われるため、得られる特徴ベクトルの次元数は300次元で一定である.

#### 4.2 特有表現に基づく特徴量

特有表現に基づく特徴量とは、あらかじめ人手でリストアップした、分類に役立つと考えられる特徴を含むかどうかということによる 16 次元の特徴量である.

人がレビュー文書を読んで、初訪問レビューであるか再訪問レビューであるかを判断する際には、その判断の根拠となる文が存在することが多い。そのように一文だけで初訪問か再訪問かを判断できる文を根拠文と呼ぶこととする。なお、本研究では、「.」「!」「改行」で区切られるものを文と扱う。根拠文には初訪問や再訪問を特徴付ける語や言い回しが含まれている可能性がある。



表 1 再訪問レビューに頻出する特有表現 9 種

特有表現(基本形)	例文
再訪(再訪問)	12 月再訪.
久しぶり	2015年5月久しぶりに行きました.
年ぶり	約1年ぶりとなるランチ訪問です.
回目	2回目の訪問です!
回数	訪問回数 3 回
毎回	ここでは毎回 750 円のミックスフライ
よく利用	私はよく利用しています.
前回は	前回は家族できました.
相変わらず	相変わらず何を食べても美味

表 2 初訪問レビューに頻出する特有表現 7 種

特有表現 (基本形)	例文
初訪 (初訪問)	今回が初訪でした.
ログ	ブログや食ベ <b>ログ</b> などを見て来店.
入る	入るきっかけがありませんでしたが…
近く	近くにあるこのお店に来ました.
見つける	また素敵なお店を見つけることが
思った	思った以上に具沢山♪
たまたま	たまたま物色中に発見しました.

以下は、レビュー文書の例である.

創業51年!!凄いです笑

昭和の映画に出てくるような定食屋さんでございます. 会社から車で数分なもんでたまに行かせていただいてます.\*<sup>1</sup>

この例では、「会社から車で数分なもんでたまに行かせていただいてます.」という一文だけから、すでに訪問したことがあることがわかる。すなわち、このレビュー文書が再訪問レビューであることがわかる。このように、根拠文を分析することによって、初訪問や再訪問を特徴付ける語や言い回しを発見することができると考えられる。本研究ではそのような語や言い回しを特有表現と呼ぶこととする。

根拠文を分析して特有表現を発見するために、後述する実験で使用するデータとは異なるレビュー文書を 618 件収集した. それらを直接観察することで、特有表現のリストアップを行った. 得られた特有表現について、再訪問レビューないしは初訪問レビューのいずれかの根拠文にのみよく現れる特有表現を取得した. 得られた特有表現のうち、再訪問レビューに頻出する特有表現 9 種を表 1 に、初訪問レビューに頻出する特有表現 7 種を表 2 にまとめる.

「初訪」「再訪」「ログ」といった特有表現は、文中にその語が含まれていることで本研究の分類問題に貢献する特徴となると考えられる.

たとえば、「初訪」は、「今回が初訪でした」や「ランチタイム に初訪問です」といった表現で用いられる。これらの表現は初め てその飲食店を利用した場合のものであり、「初訪」という特有 表現は初訪問レビューを特徴づける素性となると考えられる。

「ログ」という特有表現は、「ブログ」や「食べログ」の一部として現れる表現である。たとえば、「その時過去ブログを思い出した」「食べログの評価を見て決めただけに益々期待が高まります」「神戸のランチブログで、こちらの『冷やしきつねうどん』を知って、気になっていました」といったような記述で「ログ」が現れる。これらの文は、他のレビューを参照して飲食店を決定したことを示しており、そのようにして訪れるのはたいていの場合が初訪問であると考えられる。そのため、「ログ」という特有表現も初訪問レビューを特徴づける素性となると考えられる。

「再訪」は、初訪問レビューと再訪問レビューのいずれでも現れる語である。たとえば、「またぜひ再訪したいお店です」といったような記述は、初訪問レビューを示す特徴であると考えられる。一方で、「【再訪】」や「再訪.」のように、「再訪」という語が括弧でくくられていたり、体言止めで現れたりする場合には、再訪問のレビューであることが多い。そこで、「再訪」がレビュー文書中に現れる際に、その直後の文字が記号などである、すなわち、ひらがな、カタカナ、漢字、数字ではなかった場合に、そのような「再訪」を特有表現とすることとした。また、「再訪」ではなく「再訪問」の場合についても、同じ特有表現として扱うこととした。

「久しぶり」「相変わらず」という特有表現では、ひらがなでの表記である「ひさしぶり」「あいかわらず」も同じ特有表現として扱う。また、「年ぶり」「回目」「回数」という特有表現は、その直前に数字が表れることを条件とする。数字は漢字でもアラビア数字でもかまわないものとする。このように、いくらかの表記ゆれについては、同じ特有表現とみなすという処理を行った。

このように、16 種類の特有表現について、レビュー文書中に出現した場合には 1、出現しなかった場合には 0 とすることによって 16 次元の特徴ベクトルが作成される。それが特有表現に基づく特徴量である。

# 4.3 根拠文との類似性に基づく特徴量

根拠文との類似性に基づく特徴量は、レビュー文書中に、既知の根拠文と類似する文が含まれているかどうかによって作成する 30次元の特徴量である.

後述するが、実験データでは、レビュー文書に対する正解ラベルの付与を人手で行った. その際に、根拠文が存在する場合には根拠文を特定してもらった. 交差検証の各回において、訓練データに現れる根拠文のみを既知の根拠文として、特徴ベクトル作成を行った.

レビュー文書中に、既知の根拠文と類似する文が含まれているとすると、そのレビュー文書は、類似する根拠文が含まれるレビュー文書と同じクラスに分類することが適切である可能性が高い。そのような特徴を素性として用いた場合、既知の根拠文の数だけ素性が作られてしまう。実際には、いくつか類似する根拠文のタイプがあると考えられるため、まず根拠文のクラスタリングを行い、そのいずれかと類似する文を含むかどうかを素性とする



<sup>\*</sup> https://tabelog.com/hokkaido/A0101/A010204/1018517/ dtlrvwlst/B251685366/

表 3 根拠文におけるタイプ例

タイプ	例文
直接記入	12 月 初訪問
	再訪問です
紹介・発見	食べ口グを見て来たいと思っていました
	友人の紹介で知りました
常連	いつ食べても飽きません
	いつも美味しくて満足です

手法を考えた. 以下がこの特徴量作成の手順の概要である.

- 1. 根拠文を特徴ベクトル化
- 2. 根拠文を K-means 法を用いてクラスタリング
- 3. 入力のレビュー文書の各文を特徴ベクトル化
- 4. レビュー文書の各文と各クラスタの距離を計算
- 5. 距離が近いクラスタの有無により特徴ベクトル作成

まず、根拠文をすべて TF-IDF によって特徴ベクトル化する. 次に、根拠文のクラスタリングを行う. これは、先述したように、根拠文がある程度タイプ分けを行うことができるという仮説に基づくものである. 根拠文のタイプの例を表3に示す. このように、根拠文はいくつかの類似するグループにまとめることが可能であると考えられる.

クラスタリングは K-means 法を用いて行う. まず,根拠文を 初訪問レビューに現れたものと,再訪問レビューに現れたものに 分割した. そして,分割されたそれぞれの根拠文集合において, k=15 でクラスタリングを行った.初訪問レビューの根拠文と 再訪問レビューの根拠文からそれぞれ 15 個のクラスタ,合計で 30 個のクラスタが得られる.得られた各クラスタの重心を,各 クラスタの代表ベクトルとしする.

入力として未知のレビュー文書が与えられた際に、そのレビュー文書中の各文を、TF-IDFでベクトル化する。この際、IDFの計算には、先に根拠文を TF-IDFでベクトル化する際に計算された重みを用いる。

ここまでで、30個の根拠文のクラスタの代表ベクトルと、入力のレビュー文書の各文の特徴ベクトルが得られたこととなる。それらのすべての組み合わせについて距離を求める。そして、それぞれのクラスタの代表ベクトルにおいて、文の特徴ベクトルとの距離の最小値が、しきい値以下だった場合に 1、しきい値よりも大きかった場合に 0 とすることで根拠文との類似性に基づく特徴量を作成した。

以下で、この手法の実装について述べる。概要は、図 3 のようになっている。まず、根拠文クラスタの代表ベクトル集合を  $E = \{e_1, e_2, \dots, e_{30}\}$  と表すこととする。 $e_i$  は i 番目の根拠文クラスタの代表ベクトルを表す。

入力として与えられた未知のレビュー文書は文集合とみなす。それぞれの文は TF-IDF によって特徴ベクトルとして表現される。そこで,レビュー文書の特徴ベクトル集合を  $R=\{r_1,r_2,\cdots,r_n\}$  と表すこととする。ここで,n はレビュー文書 R

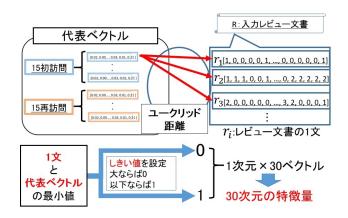


図3 根拠文との類似性に基づく特徴量作成の概要

における文の数である.  $r_j$  は j 番目の文の特徴ベクトルを表す. 次に、あらゆる  $i(1 \le i \le 30)$  と  $j(1 \le j \le n)$  の組み合わせについて距離を求める. 今回は、距離としてユークリッド距離を用いた. ベクトル  $e_i$  とベクトル  $r_j$  のユークリッド距離  $d(e_i,r_j)$  は、以下のように定義される.

$$d(e_i, r_j) = \sqrt{(e_i - r_j) \cdot (e_i - r_j)} \tag{1}$$

 $e_i$  に対して、最も類似していた、すなわち、最も距離が小さかった文の距離が、あるしきい値  $\theta$  よりも小さいかどうかを求めて、それが  $\theta$  よりも小さいかどうかによって、根拠文との類似性に基づく特徴量における  $i(1 \le i \le 30)$  番目の重み  $v_i$  を以下のように定義する.

$$v_i = \begin{cases} 1 & (\min_j(d(e_i, r_j)) \le \theta) \\ 0 & (\text{otherwise}) \end{cases}$$
 (2)

#### 5 評価実験

本節では、評価実験について述べる. まず、実験で用いるデータセットについて説明し、その後、実験の詳細を説明する.

#### 5.1 データセット

本実験で使用するデータは食べログから収集したレビュー文書である.このレビュー文書に対しラベル付けを行い,データセットを作成した.

収集したレビュー文書は、2,538 件の食べログのレビュー文書である。そのうち、113 件のレビューを除いた 2,425 件のレビュー文書を最終的に利用することとなった。まず、そのデータの収集の方法と、113 件の除外したレビュー文書について説明する。

先述した通り、本研究では、食べ口グで各飲食店に対してあるユーザが投稿したレビュー文書のうち、1回目に投稿されたレビュー文書を対象とする。すなわち、同一ユーザが同一店舗に2回以上投稿している場合は、2回目以降のレビュー文書は収集しない

収集するために、地域と飲食店のジャンルを選定した.地域は47都道府県として、食べログでクラス分けされている6つのジャンルを選択する.選択した地域と飲食店のジャンルは表4の通りである.

表 4 レビュー文書を収集する地域とジャンル

地域	47 都道府県		
ジャンル	フレンチ	寿司	ステーキ・ハンバーグ
	うどん	定食・食堂	パン・サンドウィッチ



図4 評価欄のみでレビュー文書が存在しない例

表 5 除外するレビュー文書の概要

	該当数
追記型	103
画像のみ	4
評価欄のみ	6

47 都道府県において 6 つのジャンルの飲食店を対象とするため、地域とジャンルの組み合わせは全部で 282 通りとなる. それぞれの組み合わせにおいて、飲食店を 3 店舗ずつ、計 846 店舗を選択した. ただし、レビューが 3 件以上投稿されている飲食店のみを対象とした. 次に、選択した 846 店舗の飲食店ごとに、レビュー文書を投稿したユーザから、ランダムに 3 名のユーザを選択し、各ユーザが 1 回目に投稿したレビュー文書を収集した. これにより、846 店舗に対しそれぞれ 3 件のレビュー文書の総計 2,538 件のレビュー文書が収集された.

このように収集されたレビュー文書には、本研究に適さないものがいくつか含まれている。まず、1つのレビュー文書中に複数回にわたって追記が行われているものである。食べログでは、ユーザが投稿したレビュー文書を後日編集することが可能であり、その機能を利用して、複数回のレビューを一つのレビュー文書に記載していることがある。そのようなものは、今回は人手で除外した。このように、追記が行われていたレビュー文書は収集したレビュー文書中に 103 件存在した。

また、レビュー文書にテキストが書かれていない場合もあり、そのような空のレビュー文書も除外した. たとえば、画像だけが投稿されている場合や、図 4 のようにスコアは投稿されているが、レビュー文書は存在していない場合があった. そのような、レビュー文書が存在しないもは合計 10 件あった.

これらの除外されたレビュー文書の総数は表 5 に示すとおり 113 件となり、これらのレビュー文書を除くとデータセットに おけるレビュー文書の総数は 2,425 件となった.次に、これら 2,425 件のレビュー文書に対してラベル付けを行った.

表 6 評価者ごとのラベル付けの結果

ラベル	1	2	3	4	5
評価者 X	1,250	343	278	56	498
評価者 Y	1,220	311	444	54	396
評価者 Z	1,288	381	197	67	492

ラベル付けは、3名の評価者で行った。それぞれの評価者には、 実際にレビュー文書を読んでもらい、初訪問におけるものか、再 訪問におけるものかを5段階で評価して行ってもらった。5段階 はそれぞれ以下のとおりである。

- 1. 確実に初訪問
- 2. 初訪問だろうと思われる
- 3 不明
- 4. 再訪問だろうと思われる
- 5. 確実に再訪問

評価者ごとのラベル付けの結果を表 6 に示す。これらの一致度を計るために、Fleiss [6] の  $\kappa$  係数を導出した。  $\kappa$  係数の値は0.58 となり、評価者3名のラベル付けの結果は、中程度の一致であった。

ラベル付けと同時に、根拠文の収集を行った。ここで収集する 根拠文は評価者が、ラベル「1」または「5」を選択したときに、 レビュー文書内において出現する文のうち、そのラベルを付与す るにあたって決め手となった文のうち一番最初に出現したものを 選択してもらうというものである。

収集された根拠文の例を表7に示す.

実験で3つのクラス初訪問,再訪問,不明が必要である。そこで,3名の評価者によるラベル付けの結果を以下のルールを用いて集約することで,これらの3つのクラスを作成した。

**初訪問 (-1)** 2名以上の評価者がラベル「1」を付与したレビュー文書

再訪問 (+1) 2名以上の評価者がラベル「1」を付与したレビュー 文書

不明 (0) それ以外

この結果, データセットにおけるレビュー文書 2,435 件において, 3 クラスそれぞれのレビュー文書数は表 8 のようになった.

### 5.2 分類器の選定のための実験

本研究の目的は、与えられたレビュー文書を初訪問、再訪問、不明の3つのクラスに分類することであり、それにあたって、どのような特徴ベクトル作成方法を用いるべきか、また、どのような機械学習の分類器を用いるべきかを明らかにすることが評価実験の目的である。

そこで、まず、レビュー文書の分類に適した機械学習の分類器 を選ぶための実験を行った。本実験の結果をうけて、適切である と考えられる分類器を用いて特徴ベクトルの作成方法についての 実験を行う。

実験で用いた分類器は、以下の7種類である.



## 表 7 根拠文の例

初訪問 たまたま通りがかりに見つけて寄ってみました.

こちらのお店がオープンしてからずっと行きたいなぁと思っていました. どこか行きたくて色々検索してたら、良さそうな所で空きがあったので.

~~ 実店舗初訪問 (2015. 3)~~

再訪問 小さい頃から、祖父母に連れられてよく来ていたお店.

学生時代の一時期,道後に滞在したことがあり,よく行っていました.

久しぶりに, ここ弥太郎寿司に行ってきました.

9月に2回目の訪問.

表 8 3 クラスそれぞれのレビュー文書数

初訪問 (-1)	不明 (0)	再訪問 (+1)	
1,249	739	437	

- Naïve Bayes
- 決定木
- ロジスティック回帰 (LogisticRegression)
- SVM
- kNN
- AdaBoost
- RandomForest

特徴ベクトルとしては、提案するすべての特徴量を用いることとした。 すなわち、4.1 節の 300 次元、4.2 節の 16 次元、4.3 節の 30 次元をすべて連結した 346 次元のベクトルを用いた。

なお、4.3 節の (2) 式におけるユークリッド距離のしきい値  $\theta$  については、予備実験を行い、 $\theta=1.8$  と決定した.

実験は5分割交差検証で行った.提案する特徴ベクトルの作成 手法は,データの全体集合に依存する部分があるため,交差検証 の各回において,訓練データとテストデータを分割し,毎回ベク トルの作成方法から実行し直す方法をとった.

いくつかの分類器においては、ハイパーパラメータについてグリッドサーチを行い、最適なパラメータを用いることとした。グリッドサーチは、訓練データにおいて5分割交差検証を行い、その結果によって最適なパラメータを決定するという方法で行った。最適なパラメータを決定するための評価軸としては、正解率(accuracy)を用いた。

試したハイパーパラメータは以下のとおりである.

SVM-kernel = {linear, rbf}

SVM- $\gamma = \{0.1, 0.01, 0.001\}$ SVM-C =  $\{1, 10, 100\}$ 

 $LogisticRegression-C = \{0.01, 0.1, 1, 10, 100, 1000\}$ 

 $kNN-k = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ 

SVM では、カーネルとして線形カーネルと RBF カーネルを用いた。それぞれのカーネルにおいて、機械学習における正則化の強さを表す定数 C を変化させた。また、RBF カーネルのハイパーパラメータである  $\gamma$  の値も変化させた。ロジスティック回帰における定数 C も変化させた。kNN における k は 1 から 10 まで

表 9 7つの分類器を用いた実験結果

US	適合率	再現率	F1
なし	0.6690	0.3513	0.2380
あり	0.6160	0.3558	0.2472
なし	0.2472	0.4659	0.4551
あり	0.4423	0.3950	0.4061
なし	0.5965	0.5987	0.5697
あり	0.5639	0.5270	0.5332
なし	0.5791	0.5785	0.5387
あり	0.5572	0.5253	0.5321
なし	0.4696	0.5129	0.4538
あり	0.4433	0.4247	0.4270
なし	0.5288	0.5439	0.5028
あり	0.4956	0.4622	0.4716
なし	0.4962	0.5183	0.4103
あり	0.5288	0.5439	0.5028
	なあなあなあなあなあなあな	なし 0.6690 あり 0.6160 なし 0.2472 あり 0.4423 なし 0.5965 あり 0.5639 なし 0.5791 あり 0.5572 なし 0.4696 あり 0.4433 なし 0.5288 あり 0.4956 なし 0.4962	なし 0.6690 0.3513 あり 0.6160 0.3558 なし 0.2472 0.4659 あり 0.4423 0.3950 なし 0.5965 0.5987 あり 0.5639 0.5270 なし 0.5791 0.5785 あり 0.5572 0.5253 なし 0.4696 0.5129 あり 0.4433 0.4247 なし 0.5288 0.5439 あり 0.4956 0.4622 なし 0.4962 0.5183

変化させた.

表8をみると、レビュー文書数がそれぞれのクラスに偏っていることがわかる。このような場合には、分類器によっては多くのレビュー文書が存在するクラスを予測することが不当に多くなってしまう場合がある。そのような時に用いられる手法として、アンダーサンプリングという手法がある。これは、学習時において、多くのアイテムが存在するクラスから学習データを取り除き、すべてのクラスのアイテム数を等しくすることによって、バランスさせたデータで学習を行うという手法である。今回は、アンダーサンプリングを行わなかった場合と、行った場合の両方について実験を行った。

7つの分類器を用いて、実験を行った結果を表9に示す。表の2列目のUSという項目は、アンダーサンプリングのあり、なしを表している。ここで示されているのは適合率と再現率とF1であるが、これらの値は、3つのクラスにおけるそれぞれの評価値の加重平均を、5分割交差検証のそれぞれにおいて求めた平均となっている。7つの分類器の中で、平均的に高いF値が得られたのはSVMとロジスティック回帰であった。

そこで,以後は,SVM とロジスティック回帰を用いて,特徴ベクトルの作成において,3種類の特徴量がどのように貢献しているかの検証を行う.



SVM						
	アンダーサンプリングなし			アンサーサンプリングあり		
本研究の素性の組合せ	適合率	再現率	F1	適合率	再現率	F1
(1): TF-IDF に基づく特徴量	0.5039	0.5562	0.4882	0.5337	0.4989	0.5076
(2):特有表現に基づく特徴量	0.3844	0.5414	0.4217	0.5430	0.4499	0.4390
(3):根拠文との類似性に基づく特徴量	0.2755	0.5121	0.3506	0.4101	0.3694	0.3472
(1)+(2)	0.5769	0.5752	0.5267	0.5610	0.5294	0.5365
(1)+(3)	0.5014	0.5538	0.4934	0.5263	0.4960	0.5046
(2)+(3)	0.3851	0.5414	0.4217	0.5236	0.4503	0.4460
(1)+(2)+(3)	0.5791	0.5785	0.5387	0.5572	0.5253	0.5321

表 10 SVM における様々な特徴量の組み合わせにおける実験結果

表 11 ロジスティック回帰における様々な特徴量の組み合わせにおける実験結果

ロジスティック回帰						
	アンダーサンプリングなし			アンサーサンプリングあり		
本研究の素性の組合せ	適合率	再現率	F1	適合率	再現率	F1
(1): TF-IDF に基づく特徴量	0.5299	0.5628	0.4977	0.5385	0.4952	0.5050
(2):特有表現に基づく特徴量	0.3913	0.5463	0.4251	0.5396	0.4507	0.4438
(3):根拠文との類似性に基づく特徴量	0.2966	0.5142	0.3520	0.4175	0.3868	0.3845
(1)+(2)	0.5901	0.5909	0.5504	0.5767	0.5414	0.5487
(1)+(3)	0.5326	0.5653	0.5056	0.5371	0.4940	0.5035
(2)+(3)	0.4917	0.5360	0.4365	0.5141	0.4515	0.4497
(1)+(2)+(3)	0.5965	0.5987	0.5697	0.5639	0.5270	0.5332

#### 5.3 3つの特徴量の組合せについての実験

4節では、以下の3つの特徴量について説明した。

- 1. TF-IDF に基づく特徴量
- 2. 特有表現に基づく特徴量
- 3. 根拠文との類似性に基づく特徴量

これらが、レビュー文書の分類においてどのように貢献しているかを明らかにするのが本実験の目的である。そこで、これらの3つの特徴量から、あらゆる組み合わせで特徴ベクトルを作成して、それらの特徴ベクトルを用いた際の分類結果を明らかにする。

先述した実験より、分類器としては、SVM とロジスティック 回帰を用いることとした。それぞれのハイパーパラメータについては、先述したものを用いたグリッドサーチを行った。

SVM を分類器として用いた場合の結果を表 10 に示す. また, ロジスティック回帰を分類器として用いた場合の結果を表 11 に示す.

実験の結果、SVM とロジスティック回帰のいずれの分類器を用いた場合であっても、すべての特徴量を組み合わせた場合が、最も良い結果が得られたといえる。それぞれにおいて、F値の平均は、SVM を用いた場合には 0.5387、ロジスティック回帰を用いた場合には 0.5697 となった。分類の精度としては決して高いとは言えない結果であった。ロジスティック回帰では、適合率は

表 12 SVM の場合の混同行列

	予測ラベル				
		初訪問	不明	再訪問	
正解ラベル	初訪問	1067	144	38	
	不明	472	245	22	
	再訪問	279	67	91	

表 13 ロジスティック回帰の場合の混同行列

		予測ラベル			
		初訪問	不明	再訪問	
	初訪問	1056	157	36	
正解ラベル	不明	434	272	33	
	再訪問	235	78	124	

0.5965 であり、飲食店における再訪問率の多寡を報告するために 多少利用することが可能であると考えられる.

混同行列による実験結果の分析を行った。SVM を用いた際の混同行列を表 12 に示す。ロジスティック回帰を用いた際の混同行列を表 13 に示す。それぞれ,5 分割交差検証における結果の混同行列をすべて足し合わせたものになっており,混同行列中のレビュー文書数は 2,425 となっている。



この表より、SVM とロジスティック回帰ともに、再訪問と不明のレビュー文書が誤って初訪問と予測されることが多いことが確認できる.

ランダムで分類を行った場合,ある一つのレビュー文書が再訪問のレビュー文書である確率は 437/2425 = 18.02% となる.ロジスティック回帰を用いた分類器を利用して,再訪問のレビュー文書を収集しようとしたとすると,分類器が再訪問と判断した件数は 193 件,予測と正解が一致したものは 124 件であったことから,124/193 = 64.2% の割合で再訪問のレビュー文書を得られることとなる.再現率は高いとは言い難いが,ランダムで収集するよりも高い精度で再訪問のレビュー文書を収集することができる可能性が示唆されているといえる.

#### 6 まとめと今後の課題

本研究では飲食店のレビュー文書が、初訪問におけるレビュー 文書であるか、再訪問におけるレビュー文書であるか、そのいず れか判断できないかを自動的に分類する問題に取り組んだ.

データセットとして食べログからレビュー文書を 2,425 件収集し、それらのレビュー文書に対して、3 名評価者によって初訪問レビューか再訪問レビューか不明レビューであるかラベル付けを行った。ラベル付けの結果では、人の評価者であればレビュー文書を読むことによって、ある程度は一致してこのような分類を行うことが可能であることが示された。また、ラベル付けの結果では、ある飲食店に対する1回目のレビュー文書であるにも関わらず、再訪問レビューであったと 2 名以上の評価者が判断されたものが、全レビュー文書 2,425 件中に 437 件あり、本研究で取り組んだ課題が存在することが確認された。

それらのレビュー文書を分類するため、本研究では3種類の特 徴量によって特徴ベクトルを作成する手法を提案した.

- 1. TF-IDF に基づく特徴量
- 2. 特有表現に基づく特徴量
- 3. 根拠文との類似性に基づく特徴量

1つ目の特徴量は、レビュー文書を Bag-of-Words として扱い、TF-IDF 重みづけによって作成するものである. LSA による次元 圧縮を行うため、300次元の特徴量であった. 2つ目の特徴量は、特有表現という初訪問や再訪問のレビューにおいて特有に出現する語や言い回しによる特徴量であった. 特有表現として 16 種類を取り上げたため、16次元の特徴量であった. 3つ目の特徴量は人が初訪問レビューや再訪問レビューと判断する決め手となった根拠文を利用したものである. 人がレビュー文書を判断するのは、レビュー文書全体からではなく、一部の根拠となる文によると考えられる. そのような根拠文と類似する文がレビュー文書に存在するかどうかによる特徴量であった. 30 個の根拠文のクラスタを構成したため、30次元の特徴量であった. そして、これら3つの特徴量を用いて特徴ベクトルを作成し、機械学習の分類器をもちいて分類を行った.

7つの分類器を試した結果, SVM とロジスティック回帰を用いることが良いということが示唆された. そこで, SVM とロジスティック回帰を用いて, 3種類の特徴量のどの組み合わせを用い

ることが適切かを評価する実験を行った. その結果, 提案した 3 種類の特徴量をすべて用いることが良いという結果が得られた.

それぞれの F 値の平均は,SVM を用いた場合には 0.53652,ロジスティック回帰を用いた場合には 0.5697 となった.実験の結果よりロジスティック回帰を用いた場合の適合率は 0.5965 であり,たとえば,飲食店の再訪問率を得ることに利用することが可能であるということが示唆された.

## 謝辞

本研究の一部は JSPS 科学研究費助成事業 JP16H02906, JP17H00762, JP18H03243, JP18H03244, JP18H03494 による助成を受けたものです. ここに記して謝意を表します.

#### 参考文献

- [1] Li-Chen Cheng, Judy C. R. Tseng, and Tsai-Yu Chung. Case study of fake web reviews. In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, pages 706–709, 2017.
- [2] Kushal Dave, Steve Lawrence, and David M. Pennock. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th International Conference on World Wide Web*, pages 519–528, 2003.
- [3] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R Harshman. Indexing by latent semantic analysis. *Journal of the American society for information science*, 41(6):391–407, 2013.
- [4] Xiaowen Ding and Bing Liu. The utility of linguistic rules in opinion mining. In Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 811–812, 2007.
- [5] Amir Fayazi, Kyumin Lee, James Caverlee, and Anna Squicciarini. Uncovering crowdsourced manipulation of online reviews. In Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 233–242, 2015.
- [6] Joseph L Fleiss. Measuring nominal scale agreement among many raters. Psychological bulletin, 76(5):378, 1971.
- [7] Eric Gilbert and Karrie Karahalios. Understanding deja reviewers. In Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work, pages 225–228, 2010.
- [8] Theodoros Lappas, Mark Crovella, and Evimaria Terzi. Selecting a characteristic set of reviews. In Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 832–840, 2012.
- [9] Ankan Mullick, Surjodoy Ghosh D, Shivam Maheswari, Srotaswini Sahoo, Suman Kalyan Maity, Soumya C, and Pawan Goyal. Identifying opinion and fact subcategories from the social web. In *Proceedings of* the 2018 ACM Conference on Supporting Groupwork, pages 145–149, 2018.
- [10] Michael P. O'Mahony and Barry Smyth. Learning to recommend helpful hotel reviews. In *Proceedings of the Third ACM Conference* on *Recommender Systems*, RecSys '09, pages 305–308, 2009.
- [11] Deanna Osman, John Yearwood, and Peter Vamplew. Using corpus analysis to inform research into opinion detection in blogs. In *Proceedings of the Sixth Australasian Conference on Data Mining and Analytics Volume 70*, pages 65–75, 2007.
  [12] Phong Minh Vu, Hung Viet Pham, Tam The Nguyen, and Tung Thanh
- [12] Phong Minh Vu, Hung Viet Pham, Tam The Nguyen, and Tung Thanh Nguyen. Phrase-based extraction of user opinions in mobile app reviews. In *Proceedings of the 31st IEEE/ACM International Conference* on Automated Software Engineering, pages 726–731, 2016.
- [13] Derry Tanti Wijaya and Stéphane Bressan. A random walk on the red carpet: Rating movies with user reviews and pagerank. In Proceedings of the 17th ACM Conference on Information and Knowledge Management, pages 951–960, 2008.
- [14] José P. Zagal, Amanda Ladd, and Terris Johnson. Characterizing and understanding game reviews. In *Proceedings of the 4th International* Conference on Foundations of Digital Games, pages 215–222, 2009.

