

美容室推薦システムのレビュー抽出方法の検討

○宗野みゆき 李皓（静岡大学）

Consideration of Review Extraction Method on Beauty Parlors Recommendation System

*M.Sono and H.Lee(Shizuoka University)

レビューデータから価値のあるレビューを抽出する方法として、本研究では TFIDF の高い特徴語と、極性辞書に掲載された評価表現語、そしてそれらが係り受け関係にある文を含むレビューを抽出する手法を提案した。被験者による妥当性の検証結果から、特徴語が含まれているレビュー文は理解度が高い結果を得た。しかし、特徴語と評価表現語が係り受け関係にある文を含むレビューについては完全な優位性があるわけではなく、部分的に支持される結果となった。今後の課題は美容室レビューデータに合わせた辞書の作成、類義語による特徴語の抽出である。さらにレビューのカテゴリ分類やレビュー抽出条件の見直し、教師あり学習による新たなレビュー抽出方法などを検討し、美容室推薦システム構築におけるもっとも適切なレビューの抽出方法を開発していく。

キーワード: レビュー推薦, 自然言語処理, TFIDF, 感情分析

1 研究背景

新型コロナウイルスの感染拡大から人との接触を避ける生活スタイルに変化し、購買行動については店頭での購入からインターネットショッピングの利用が急増している。場所や時間を問わず手軽に利用できる一方で、店舗のように専門家から商品などの説明を受けることが出来ないため、人々は口コミサイトのレビューを閲覧して情報を収集している。

口コミに関するアンケート結果¹⁾では、今まで利用したことがない商品・サービス購入時において口コミを確認する、確認することの方が多くと回答したのは全体の84.4%であった。また評判がよくない場合は全体の76.4%が購入を取りやめる、取りやめることが多いという結果が報告されていることから、口コミの一種であるレビューの影響力や信頼度が大きいと言える。

レビューには購入者やサービスの利用者による実体験に基づいた情報が含まれているが、購入者や利用者が増えると同時に投稿数は多くなり、レビュー数が増えた分だけでユーザーにとって必要な情報が掲載されているレビューの探索が困難になっていく。多くのレビューの中から内容に具体性や説得力があるレビューを目視で確認するのは非常に手間がかかる。そのため、欲しい情報をもっと便利に見つかる機能の実現を口コミサイトに求めている人が全世代を通して最も多いというアンケート結果²⁾も報告されている。ユーザーの検索を支援する機能として推薦システムが用いられることがある。膨大な量の商品やサービスの中からユーザーの好みやニーズにマッチすると思われるものを提示する機能³⁾はAmazon⁴⁾や楽天市場⁵⁾などのショッピングサイトで主に活用されており、購入履歴や閲覧履歴から類似した商品をシステムが提示する機能が該当する。

2 先行研究

辻井・津田⁶⁾は宿泊レビューデータを分析して、ホテルの特徴を表す属性が評価語と係り受け関係になっている情報、すなわち宿泊検討者にとって有意義な情報を抽出する手法を検討した。その結果、宿泊エリアやユーザーの評価基準などによって同じ数値評価でも全く異なる指標や視点で評価・採点されており、レビューの数値評価のみの利用や購入の検討は不適切であることを示した。

川又ら⁷⁾は料理サイトに料理別で掲載されている既存のレビューをコーパスとして飲食店に対するレビューから料理に対する評価文を抽出した。料理のレビューと他のレビューと異なり、低頻度の特徴語が多いことから頻度情報を抑えた改良版のTFIDFにて料理に関する特徴語を網羅的に抽出することに成功した。

尾内・林⁸⁾は単語の重要度を表すTFIDFの高い単語が肯定的・否定的な評価表現語と係り受け関係にある文を含む映画のレビューを読んだとき、特徴語のみを含み、評価情報を含まないレビューを読んだ時よりもユーザーはその映画に興味を喚起させられることを示した。

先行研究ではレビューデータから特徴語を抽出することに加えて、評価表現語との係り受け関係が確認できる文を提示することでユーザーのレビュー閲覧時の負担を減らす役割を果たすことが示された。しかし、宿泊予約や飲食店、映画といったレビューデータは多くの研究で用いられることが多い。そのため、既存研究で用いられていない新たなジャンルのレビューデータでも、いずれの研究と同様の結果を得ることが出来るか検証するために、本研究では美容室のレビューデータ

を使用して分析・実験を行った。

3 研究目的

我々は従来のユーザーの嗜好や希望に合った商品やサービスの推薦に加えて、膨大な量のレビューの中からユーザーにとって有用な情報を含んだ「参考価値のある」レビューも推薦システムを導入する「レビュー提示型推薦システム」を考案し、構築を研究の最終目標とする。レビュー提示型推薦システムの構築に先立ち、本研究では「価値のある」レビューがどのような特徴をもっているか被験者実験を通じて明らかにすることを目的に、抽出手法の検討を行った。

本研究では汎用性の高い尾内・林⁸⁾の手法を参考に特徴的な単語と、特徴的な単語の良し悪しが明確に判断できる情報の抽出をおこなう。美容室のレビューデータを適用して先行研究のような結果が得られることを示すために次の3つの仮説を立て、分析とレビューの抽出、そして被験者実験による仮説の検証を行った。

- 仮説 1: 商品やサービスなどの特徴語を含むレビューはユーザーの理解度が高い。
仮説 2: 特徴語と評価表現語が係り受け関係にある文章を含むレビューは特徴の良し悪しに対するユーザーの理解度をより高められる
仮説 3: 特徴語と評価表現語が係り受け関係にある文章を含むレビューは参考になる

4 抽出手法

本研究で使用したデータセットは2012年1月から2014年1月にかけて株式会社リクルート⁹⁾のHot Pepper Beauty¹⁰⁾に投稿されたレビューデータである。国立情報学研究所¹¹⁾が提供するオープンデータであり、2021年9月の時点で店舗データが存在する23万5975件のレビューデータを分析・抽出対象として採用した。

レビューデータにはレビューIDとレビュー本文以外にも投稿者による評価値も含まれている。雰囲気(以下、ムード)、接客・サービス(以下、サービス)、技術・仕上がり(以下、テクニック)、メニュー・料金(以下、メニュー)の4項目とその平均を四捨五入して整数にした値の総合ポイントの計5つの評価値である。また、投稿者のニックネームや性別、世代などの属性の情報も含まれているが本研究では使用していない。

なお、レビュー本文の分析では、MacOS環境のJupyter Notebook上で、レビューデータのレビューID、レビュー文、ムードポイント、サービスポイント、テクニックポイント、メニューポイント、総合ポイントを1つのデータフレームとして読み込んで分析を行った。

分析用のプログラム言語はPythonを使用した。

文章から品詞や評価の情報を抽出するために形態素解析を行う。形態素解析は文章を文節ごとに区切る解析のことであり、本研究ではオープンソース形態素解析エンジンであるMeCabを使用した。MeCabのシステム辞書は固有表現に強いmecab-ipadic-Neologdを採用した。

次に特徴語の抽出について述べる。レビュー文の形態素解析を行った結果、重複を除いて39758単語が抽出された。このうち、全体の5%以上のレビューに含まれたTFIDFの高い40の名詞を特徴語とし、レビューデータに付随した4つの評価項目に関連する単語のカテゴリ分類を行った。店内の飾りつけや清潔感などの雰囲気に関連する「雰囲気」カテゴリ、店員の役職や接客、サービスに関連する「接客・サービス」カテゴリ、美容師をはじめとしたスタイリストの技術力や施術の仕上がりの度合いに関連する「技術・仕上がり」カテゴリ、美容室の施術メニューやメニュー料金に関連する「メニュー・料金」カテゴリと定義して、大学生3人の協力によって行った。2人以上が同一カテゴリに分類した単語をそのカテゴリの特徴語とし、どのカテゴリにも当てはまらず分類ができない場合や3人の回答が異なる場合は特徴語から除外した。なお、カテゴリの重複はなく、1つの特徴語は1つのカテゴリにのみ属するとして分類した結果はTable 1の通りである。

Table 1 特徴語のカテゴリ分類

カテゴリ	特徴語
雰囲気	雰囲気、居心地、感じ、イメージ、店内
接客・サービス	お願い、担当、スタッフ、スタイリスト、対応、アドバイス、説明、相談、接客、カウンセリング、お世話、提案
技術・仕上がり	仕上がり、髪型、セット、スタイル、スタイリング、技術
メニュー・料金	カット、カラー、パーマ、トリートメント、シャンプー、値段

4-1 特徴語を含むレビューの抽出

本研究ではレビュー内の単語の重要度を測る指標にScikit-learnパッケージのTFVectorizerによって求められるTFIDFを用いた。

TFIDFは特定の単語の出現頻度を表すTF値とレビュー数を特定の単語が出現するレビュー数で割り、文書数の規模に応じた変動影響を緩和するため対数logをとったIDF値を掛け合わせた数値である。以下にTFIDF値を求める式を示す。

$$TFIDF_{wi,dj} = TF_{wi,dj} \cdot IDF_{wi} \quad (1)$$

$$TF_{wi,dj} = \text{レビュー}dj \text{ 中の単語}wi \text{ の出現頻度} \quad (2)$$

$$IDFwi = \frac{\log 1 + \text{全レビュー数}(J)}{\text{単語}wi\text{が登場するレビュー数}} \quad (3)$$

Table1 のカテゴリごとに式(1)で求めた特徴語の TFIDF の平均値を計算し、すべてのレビューで4つのカテゴリの平均値「word-score」を求めた。

レビューに1つ以上の特徴語が含まれていると word-score が0よりも大きい。逆に word-score が0の場合は Table1 の特徴語を含まないレビューである。本研究では被験者実験を用いて特徴語の有無によってユーザーのレビューに対する理解度に違いがあるかどうかを調べることで仮説1の検証を行う。

4-2 評価表現フレーズを含むレビューの抽出

特徴語の評価情報を抽出するために形態素解析に加えて係り受け解析と感情分析を行った。

係り受け解析とは文節間の修飾関係を明らかにする解析手法であり、本研究ではサポートベクターマシンに基づく日本語係り受け解析器である CaboCha を採用した。また、係り受け解析の曖昧性を軽減させるためにレビュー文を句読点や記号（?や!など）を区切り位置の対象として解析を行った。

そして、感情分析は係り受け解析からレビュー文の係り受け関係を明らかにし、単語に係る評価情報を分析するルールベースでの感情分析を行った。

評価情報を表す単語は極性辞書に掲載されている単語であり、本研究では極性辞書に小林ら¹²⁾の日本語極性評価辞書（用言編）と東山ら¹³⁾の日本語評価極性辞書（名詞編）の2種類を採用した。それぞれの辞書は単語IDと単語、そして肯定的な単語には positive の頭文字 p、否定的な単語には negative の頭文字 n といった極性情報も付与されている。

2つの分析結果から、特徴語と係り受け関係にある単語が肯定的あるいは否定的な評価表現語のフレーズ「評価表現フレーズ」を抽出した。

係り受け解析と感情分析の際に、特徴語が肯定的な評価表現語と係り受け関係にある場合は $p=1$ として式(4)、否定的な評価表現語と係り受け関係にある場合は $n=1$ として式(5)、特徴語と評価表現の語が係り受け関係でない場合は $x=0$ として式(6)を適用して改めて評価表現を加味した TFIDF を求める。

$$pTFIDFwi, dj = TFwi, dj \cdot IDFwi \cdot p \quad (4)$$

$$nTFIDFwi, dj = TFwi, dj \cdot IDFwi \cdot n \quad (5)$$

$$xTFIDFwi, dj = TFwi, dj \cdot IDFwi \cdot x \quad (6)$$

上記の式(4)～(6)で特徴語の評価表現を加味した TFIDF から Table1 のカテゴリごとに平均値を計算し、4つのカテゴリの平均値「phrase-score」をレビューごとに求めた。

phrase-score が正の場合は肯定的な評価表現フレーズ、負の場合は否定的な評価表現フレーズ、そして0の場合は評価表現フレーズが含まれていないと言える。よって、特徴語の評価情報を含むレビューと含まないレビューの分類は phrase-score の値が0以外と0で判定することが出来る。本研究では特徴語の評価情報の有無による理解の違いについて、仮説2の検証実験を行う。

5 レビューの分類

4章で求められた word-score と phrase-score からレビューを分類し、分類数を Fig. 2 に示す。

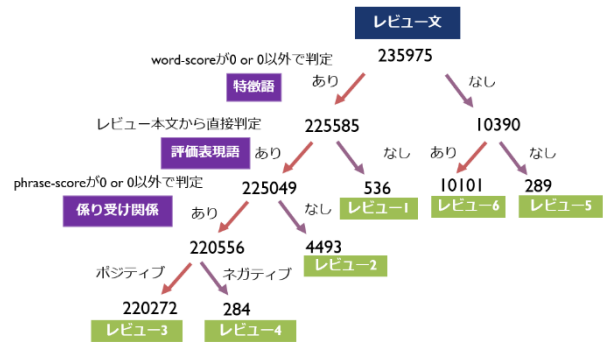


Fig. 2 レビューの分類結果

Fig. 2 ではレビュー1はレビュー中に特徴語を含むが、極性辞書に記載された評価表現語を含まないレビューである。レビュー2はレビュー中に特徴語と評価表現語のどちらも含まれているが、係り受け関係がないレビューである。レビュー3と4の分類は phrase-score の値で分類し、phrase-score > 0 のレビューを肯定的なレビュー、phrase-score < 0 を否定的なレビューと定義した。よって、レビュー3は肯定的な評価表現フレーズを含んだレビュー、レビュー4は反対に否定的な評価表現フレーズを含んだレビューである。本研究で採用したデータセットでは評価表現を含むレビューの多くは肯定的であることがわかる。レビュー5とレビュー6はどちらも特徴語を含まないレビューである。その内、レビュー5は評価表現語も含まないレビューであるが、レビュー6は評価表現語を含む。

6 妥当性の検証

抽出・分類したレビューについて、被験者実験にてユーザーにとって価値のあるレビューであるかどうかの

妥当性を検証した。

実験は2021年12月20日～2021年12月23日の3日間にわたって行い、被験者には実験の前にレビューに関するアンケートに回答してもらった。

被験者は大学1～4年生の20人（男性9人、女性11人）である。そのうち、レビューの閲覧頻度については13人が「よく見る」と回答し、「全く見たことがない」と回答した人はいなかった。一方で、レビューの投稿経験については投稿経験が「ない」と回答する人が11人と過半数を超えており、レビューを「よく見る」人が必ずしも投稿経験があるとは言えない結果となっている。また、本研究で用いたHot Pepper Beautyのレビューの閲覧経験に関する質問に対しては20人中16人が「閲覧経験がある」と回答し、Hot Pepper Beautyへの投稿経験があると回答したのは4人であった。

6-1 検証方法

20人の被験者を5人×4グループに分けた。それぞれのグループの被験者は「雰囲気」「接客・サービス」「技術・仕上がり」「メニュー・料金」の評価項目について主に言及された6件のレビューと特徴語や評価表現を含まないレビューからランダムに選ばれた1件のレビューに対する評価を行った。検証用のレビューのパターンごとの特徴はTable 2のとおりである。

Table 2 検証用のレビューのパターンごとの特徴

パターン	特徴語	評価表現語	特徴語と評価表現語の 係り受け関係 (評価表現フレーズ)		内容	phrase-score
			特徴語	評価表現語		
①	○	×	×	-	-	-
②	○	○	×	-	-	-
③	○	○	○	-	肯定的	肯定的内容の傾向が強い
④	○	○	○	-	肯定的	肯定的内容と否定的内容の混同もしくは肯定的な評価表現フレーズが少量含まれる
⑤	○	○	○	-	否定的	否定的内容の傾向が強い
⑥	○	○	○	-	否定的	肯定的内容と否定的内容の混同もしくは評価表現フレーズが少量含まれる
⑦	×	×	×	-	-	-

パターン①はFig. 1のレビュー1に該当し、特徴語は含むものの、評価表現フレーズは含まれない。式(1)で求められた各カテゴリのTFIDF平均とword-scoreが最も高く、1つのカテゴリの特徴語が多い特徴を持つ。

パターン②はFig. 1のレビュー2に該当し、特徴語と評価表現語を含むものの、特徴語と評価表現語が含まれているものの係り受け関係ではない文で構成されたレビューである。式(1)で求められた各カテゴリのTFIDF平均とword-scoreが最も高く、特徴語を多く含む。

パターン③はFig. 1のレビュー3に該当し、特徴語と肯定的な評価表現フレーズを含む。式(4)～(6)で求められたTFIDF平均とphrase-scoreが最も1に近く、肯

定的内容の傾向が強い特徴を持つ。

パターン④はFig. 1のレビュー4に該当し、特徴語と肯定的な評価表現フレーズを含む。式(4)～(6)で求められたTFIDF平均とphrase-scoreが最も0に近く、肯定的内容と否定的内容が混合しているが肯定がやや多い、あるいは肯定的な評価表現フレーズが少量含まれるという特徴を持つ。

パターン⑤はFig. 1のレビュー3に該当し、特徴語と肯定的な評価表現フレーズを含む。式(4)～(6)で求められたTFIDF平均とphrase-scoreが最も-1に近く、否定的内容の傾向が強い特徴を持つ。

パターン⑥はFig. 1のレビュー4に該当し、特徴語と否定的な評価表現フレーズを含む。式(4)～(6)で求められたTFIDF平均とphrase-scoreが最も0に近く、肯定的内容と否定的内容が混合しているが否定がやや多い、あるいは否定的な評価表現フレーズが少量含まれるという特徴を持つ。

パターン⑦はFig. 1のレビュー5に該当し、特徴語も評価表現語も含まないため、word-scoreとphrase-scoreはともに0である。

パターン④と⑥は肯定的な評価表現フレーズと否定的な評価表現フレーズが混在、もしくは特徴語の評価表現フレーズが少量含まれていることから、パターン③と④の結果と違いがあるのか比較する。

またパターン⑦についてはパターン①～⑥と比較して、それぞれの仮説の立証のために得た結果と差があるかどうかを検証した。Fig. 1におけるレビュー6は本研究では検証対象外とした。

6-2 仮説1の検証

ここでは仮説1「商品やサービスなどの特徴語を含んでいるレビューはユーザーの理解度が高い」の検証を行う。

質問文は以下の通りである。

与えられたレビューを読んで、それぞれ

- ・店舗の雰囲気(店内の飾り付けや清潔感, 空気感)
- ・店員の役職, 接客・サービス
- ・美容師(スタイリスト)の技術力や施術後の仕上がりの度合い
- ・施術メニューやメニュー料金

の4つの観点について理解できましたか。5段階評価で回答してください。なお、回答する際には同時に美容室の関連語と美容室の特徴を表していると思う「単語」(2つ以上も可能)にマーカーを引いてください。

理解度はグループごと5人の被験者によって、7件のレビューを上記の箇条書きの4つの観点で5段階評価(良い:5[よく理解できた]→悪い:1[まったく理解できなかった])した値の平均値から算出され、平均値が大きいほど理解度は高い。

パターン①～⑥は特徴語を含むレビューであるため、理解度は高くなるという仮説のもと検証を行った。パターン⑦は特徴語を含まないレビューであるために理解度は低くなるという仮説のもと検証を行った。

全 28 件のレビューの 4 つの観点における理解度の平均値は Table 3 のとおりである。

Table 3 レビューごとの理解度

グループ パターン	雰囲気	接客・ サービス	技術・ 仕上がり	メニュー・ 料金	P値
①	2.5	2.95	2.55	1.85	0.06243
②	3.2	4.25	2.65	4.30	0.008529
③	3.9	4.05	3.30	2.35	0.010996
④	3.3	3.80	3.95	3.95	0.000277
⑤	3.45	4.30	3.15	2.55	0.009767
⑥	3.45	3.50	3.65	3.75	0.001118
⑦	2.35	2.00	1.95	1.35	-
平均(①～⑥)	3.3	3.80833333	3.20833333	3.125	

すべてのレビューで特徴語を含まないパターン⑦のレビューの理解度の平均値よりも特徴語が含まれているパターン①～⑥の理解度の平均値の方が高い値をとっていることが確認できる。パターン①～⑥のレビューの理解度がパターン⑦の理解度と比較して、有意に大きいといえるかどうか確認するために有意水準 $\alpha = 0.05$ で Welch の t 検定 (片側検定) に行ったところ、パターン②～⑥のレビューでは特徴語を含むレビューが特徴語を含まないレビューよりも理解度が有意に大きいといえる。しかしパターン①の P 値が 0.06243 と 5% 有意にならなかった。よって、仮説 1 「商品やサービスなどの特徴語を含んでいるレビューはユーザーの理解度が高い」は先行研究では用いられなかった美容室のレビューデータでパターン①を除いて立証された。パターン①については対象となる他のレビューデータでも今後、検証が必要である。

6-3 仮説2の検証

次に、「特徴語と評価表現語が係り受け関係にある文章を含むレビューは特徴の良し悪しに対するユーザーの理解度をより高められる」の検証を行う。

質問文は以下の通りである。

それぞれのレビューには

- ・店舗の雰囲気(店内の飾り付けや清潔感, 空気感)
- ・店員の接客・サービス
- ・美容師(スタイリスト)の技術力や施術後の仕上がりの度合い
- ・施術メニューやメニュー料金

という 4 つの評価項目で 5 段階評価(1.0: 悪い → 5.0: 良い)の評価値が付与されています。それぞれ評価値を予想してください。なお、ある評価項目について全く言及されておらず、評価値をつけることができない場合に限り「評価できない」を選択す

ることも可能です。そして、回答する際は同時に根拠となった「フレーズ」(2 つ以上も可能)にマーカーを引いてください。

被験者が予想した 4 つの評価項目の評価値(以下、ポイントと表現)と実際のポイントが一致した割合(以下、一致率)を求めた。

パターン①②⑦は評価表現フレーズを含まないレビューであるため、一致率は低くなるという仮説のもと検証を行った。パターン③と⑤は肯定的な評価表現フレーズを含む肯定的内容のレビュー、パターン④と⑥は否定的な評価表現フレーズを含む否定的内容のレビューである。そのためパターン③～⑥は特徴の良し悪しが理解できるためポイント予測がパターン①と②に比べ容易であり、一致率は高くなるという仮説のもと検証を行った。

全 28 件のレビューの 4 つの評価項目の一致率の平均値は Table 4 のとおりである。

Table 4 レビューごとの一致率

グループ パターン	雰囲気	接客・ サービス	技術・ 仕上がり	メニュー・ 料金	平均
①	35%	20%	10%	10%	19%
②	40%	50%	25%	80%	49%
③	60%	50%	30%	5%	36%
④	45%	25%	65%	30%	41%
⑤	70%	30%	40%	40%	45%
⑥	55%	20%	65%	20%	40%
⑦	15%	0%	10%	0%	6%
平均(①～⑥)	51%	33%	39%	31%	

Table4 より全グループで特徴語と評価表現フレーズを含むパターン③～⑥の一致率は特徴語と評価表現フレーズを含まないパターン⑦の一致率よりも高い結果が得られた。特徴語を含むが評価表現フレーズを含まないパターン②の一致率について、「メニュー・料金」グループではパターン②が最も高く、「接客・サービス」グループでは評価表現フレーズを含むパターン③と評価表現フレーズを含まないパターン②の一致率が同じ値となっており、特徴語と評価表現語を含むが、係り受け関係でないパターン②が部分的に支持されたと言える。また、極性辞書の評価表現語が含まれていないレビューよりも評価表現語を含むレビューの方が一致率は高かった。

よって、特徴語と評価表現語が係り受け関係の「評価表現フレーズ」を含むレビューは比較的一致率は高くなったが、本研究の結果からは特徴語と評価表現語の係り受け関係の有無は一致率に決定的な影響を及ぼすとは言えない。

次に、レビューごとの「評価できない」と回答した割合(以下、回答不可率)と「評価できない」の回答を除いたポイント予測値と実際のポイントの差(以下、ポイント差)をそれぞれ Table 5, Table 6 に示す。

Table 5 レビューごとの回答不可率

グループ パターン	雰囲気	接客・ サービス	技術・ 仕上がり	メニュー・ 料金	平均
①	30%	50%	35%	65%	45%
②	20%	5%	50%	0%	19%
③	15%	20%	35%	65%	34%
④	15%	30%	15%	0%	15%
⑤	15%	15%	40%	10%	20%
⑥	15%	20%	25%	65%	31%
⑦	40%	70%	75%	75%	65%
平均(①~⑦)	18%	23%	33%	34%	

Table 6 レビューごとのポイント差の平均値*1

グループ パターン	雰囲気	接客・ サービス	技術・ 仕上がり	メニュー・ 料金
①	-0.2	-0.8	0.5	0.7
②	-0.8	-0.7	-1.8	-0.3
③	-0.6	-0.6	-0.3	0.4
④	-0.6	0.2	-0.4	-0.4
⑤	-0.4	-1.1	-0.6	-0.7
⑥	-0.4	-0.3	-0.3	-0.2
⑦	-1.6	-2.9	-0.6	-1.3

Table 5 の結果から評価表現語を含まないパターン①⑦よりもパターン③~⑥の回答不可率の方が低い結果となった。パターン②は一致率の結果と同様に、パターン③~⑥のレビューの回答不可率と同等もしくは低い結果を得ている。

Table 6 の結果からパターン③~⑥のレビューのポイント差の絶対値はパターン①②⑦の絶対値よりも比較的小さい結果となった。

Table 6 の絶対値が1よりも大きいレビューについて確認したところ、「雰囲気」グループのパターン⑦については投稿されたレビューで4項目とも評価5であったが、すべて1と評価した被験者がいた。このように、実際のポイントと大きく乖離して予測を行う被験者がいたために絶対値が大きくなるというような結果は「接客・サービス」のパターン⑤でも見受けられた。また、「評価できない」と回答した人が多い中、被験者の予測値が実際のポイントと大幅に異なるケースが「接客・サービス」グループのパターン⑦、「技術・仕上がり」グループのパターン②で見受けられた。「メニュー・料金」グループのパターン⑦については被験者間で予測ポイントが異なっていたためにポイント差の絶対値が大きかった。

仮説2の検証実験の結果から、レビュー文によってはパターン②のように特徴語と評価表現語が係り受け関係でなくても、特徴の良し悪しを文面から読み取れる場合があると言える。よって、「特徴語と評価表現語が係り受け関係にある文章を含むレビューはユーザーの特徴の良し悪しに対する理解度をより高められる」という仮説は本研究では完全に立証できなかったものの、部分的に支持される結果を得た。

6-4 仮説3の検証

最後に「特徴語と評価表現語が係り受け関係にある文章を含むレビューは参考になる」の検証を行う。

仮説2では4つの評価項目の観点でのレビュー文の理解度を測ったものの、仮説3では各グループの被験者に7パターンのレビューについて、レビュー文全体を読んで、参考になると思った順に1位から7位まで順位をつけるように指示をした。

評価表現フレーズを含まないパターン①②⑦よりも評価表現フレーズを含む③~⑥のレビューの方がより参考になるため、順位が高くなるという結果であれば仮説を肯定する。

アンケートの結果で得られた参考順位の平均をTable 7に示す。

Table 7 参考順位の平均

パターン	評価表現 フレーズ	雰囲気	接客・ サービス	技術・ 仕上がり	メニュー・ 料金	平均
①	×	5.8	4.2	5.8	5.6	5.35
②	×	4.2	2.6	6.2	1.4	3.6
③	肯定的	2	3.2	4.4	5.2	3.7
④	肯定的	5.2	3.4	1.4	2.4	3.1
⑤	否定的	1.2	3.2	4	2.2	2.65
⑥	否定的	3	5.4	4.2	4.2	4.2
⑦	×	6.6	6.8	7.8	7	7.05

Table 7より「技術・仕上がり」グループの評価表現フレーズを含むパターン③~⑥のレビューの順位は上位を占めた。「雰囲気」グループのレビューにおいても否定的表現の評価表現フレーズを含むパターン④のレビューを除いて評価表現を含むパターンのレビューが上位となった。「接客・サービス」「メニュー・料金」グループのレビューにおいては、パターン②が最も順位が高いレビューとなったが、パターン③~⑥は評価表現語を含まないパターン①⑦よりも順位が高かった。

以上の結果から、特徴語と評価表現語が係り受け関係にある文章を含むレビューが、評価表現語を含まないレビューよりも参考になるといえる。一方で、評価表現フレーズを含まないパターン②が最も参考になる結果が「接客・サービス」「メニュー・料金」の2グループで見られたが、全体的には評価表現フレーズを含むレビューの方が参考になりやすい結果となった。よって、本研究の結果からは仮説3は完全な立証ができなかったが、部分的に支持される結果となった。

6-5 マーカーテストの分析

ここでは仮説の検証をするだけでなく、被験者による妥当性の検証実験で良い結果を得られたレビューの特徴を知り、本研究で提案している特徴語や評価表現フレーズ以外に被験者が美容室の特徴を理解できる単

*1 ポイント差は「評価できない」の値を除いて評価値とポイントの差を求めてお

り、欠損値があるため平均値を求めていない

語やポイント予測をする際に根拠とするフレーズを分析することでよりよい抽出手法を検討する。

そこで質問 1 と質問 2 のマーカーテストの結果を分析した。

質問 1 では美容室の関連語や特徴を表す各グループのパターン①～⑦までのレビューにおいて、被験者が「特徴語」とみなした単語にマーキングした。分析の結果、提案手法で選出した多くの特徴語が、被験者にも特徴語とみなされていた。特に「カット」、「カラー」、「シャンプー」、「トリートメント」といった美容室特有の単語に関してはすべて 5 人中 4 人以上の被験者がマーキングした。

本研究での「特徴語」と被験者がマーキングした「特徴語」がほぼ一致しており、仮説 1 で示された理解度の高さにつながった。特徴語以外にマーキングされた単語には特定の美容室の特徴を表すユニークな単語が含まれているほか、ブリーチや炭酸泉スパといった施術メニューに関する語が見受けられた。さらに料金や色合いといった本研究の特徴語の類義語と思われる単語もマーカーが引かれた。マーカーで引かれた単語は、TFIDF は高くないが、これらの語も美容室に関連した語や類義語であり、特徴語の定義に関しては T 類義語も抽出する検討する必要がある。

質問 2 のマーカーテストでは、ポイント予測をする際に根拠としたフレーズに評価表現フレーズ以外にもマーキングされており、以下のように分類できた。

- ①特徴語と極性辞書に含まれない評価表現語が係り受け関係のフレーズ
- ②特徴語ではないが美容室の特徴を表す語であり、極性辞書に含まれる評価表現語と係り受け関係のフレーズ
- ③特徴語でも極性辞書に含まれる評価表現語でもない語が係り受け関係であるフレーズ

マーキングされた評価表現フレーズや上記①～③のフレーズの一部には投稿者の好みや施術前後の印象の変化に対する感想といった主観的なフレーズが含まれており、具体性に欠ける。そのような評価表現フレーズは内容が肯定的か否定的か判定する場合のみ有効であり、美容室の特徴を知るには不向きである。

反対に美容室の特徴を具体的に評価していることが読み取れる本研究で抽出した評価表現フレーズや①～③のフレーズは理解度を高めることができる。

7 まとめ

3つの仮説についてすべて立証することはできなかったが、被験者実験での検証方法及び得られた結果には Table8 で示すように相関関係があった。

Table8 実験用のレビューの検証結果の相関関係

	word-score	理解度	一致率	回答不可率	参考順位
word-score		0.559409	0.521000069	-0.4874479	-0.56694613
理解度			0.716213798	-0.7705956	-0.788766484
一致率				-0.7791186	-0.802318894
回答不可率					0.776171882
参考順位					

単語の特徴量を表す「word-score」と検証実験の結果には相関がみられるが、それ以上に検証結果同士の相関が強く、「理解度、一致率が高く、回答不可率が低いレビューは参考順位が高い」と言える。

中でも「メニュー・料金」グループのパターン②のレビューは本研究における評価表現フレーズは含まれていないものの、理解度と一致率、参考順位が 28 件中最も高く、回答不可率が 0%、さらに 4 カテゴリすべての特徴語が含まれている。そのため、最も価値のあるレビューは 4 カテゴリの特徴語を網羅しているレビューであると言える。今後は特徴語や評価表現語の数を増やすことによって、美容室のレビューデータにおいて質の高い評価表現フレーズを多く抽出することが可能であり、ユーザーにとって「価値のある」レビューの抽出数も増やすことが出来る。

8 考察

本研究では被験者実験から 3つの仮説をすべて立証できなかったが、抽出手法について見直すべきことや課題点、改善点に関する知見を得ることができた。

まず、特徴語については TFIDF を用いた手法によって名詞を抽出した。抽出された名詞をさらに手作業で分類したことで抽出されたほとんどの特徴語が被験者にとっても特徴語であると判断できたため、林・尾内⁹⁾の映画のレビューデータなどと同様に、美容室のレビューデータにおいても有効であると言える。

評価表現フレーズについては前処理の不足や係り受け関係の曖昧性が原因で CaboCha の結果が完璧とは言えず、一部の評価表現フレーズが抽出できなかった。また本来は評価表現フレーズに含まれないフレーズが評価表現フレーズと誤認識されたことによって抽出されたレビューもあった。そのため、日本語極性評価辞書による評価表現フレーズの抽出については、より適切な処理手法を検討し、結果の妥当性を再度検証する必要がある。

次に、特徴語および評価表現フレーズの「メニュー・料金」カテゴリは他 3 カテゴリに比べて、理解度や一致率、回答不可率が低い結果となった。その原因として、メニュー・料金カテゴリの特徴語の説明があいまいで、被験者に伝わりづらかったことやメニュー・料金カテゴリの説明にも当てはまることから被験者がカテゴリを誤認識した可能性が高いと考えられる。

Table9～11 は「メニュー・料金」グループの理解度・

一致率・回答不可率の結果である。Table. 9 は 3.2 を基準として特徴語が含まれているにもかかわらず基準値を下回っている場合は青色のセル, 特徴語が含まれていて基準値を上回っている場合は赤色のセル, 特徴語を含んでいないが基準値を上回っている場合は黄色のセルで表している。Table10 と 11 は 60%を基準値として評価表現フレーズが含まれているにもかかわらず基準値を下回っている場合は青色のセル, 評価表現フレーズが含まれていて基準値を上回っている場合は赤色のセル, 評価表現フレーズを含んでいないが基準値を上回っている場合は黄色のセルで表している。

Table9 「メニュー・料金」グループの理解度

評価項目 パターン	雰囲気	接客・ サービス	技術・ 仕上がり	メニュー・ 料金	平均
①	1.00	1.40	3.00	2.00	1.85
②	5.00	4.40	3.80	4.00	4.30
③	1.00	2.20	3.60	2.60	2.35
④	4.00	4.60	4.20	3.00	3.95
⑤	2.80	4.60	4.60	3.00	3.75
⑥	1.20	1.60	4.40	3.00	2.55
⑦	1.00	1.40	2.00	1.00	1.35

Table10 「メニュー・料金グループ」の一致率

評価項目 パターン	ムード	サービス	テクニック	メニュー	平均
①	0%	0%	20%	20%	10%
②	80%	80%	80%	80%	80%
③	0%	0%	0%	20%	5%
④	80%	0%	20%	20%	30%
⑤	40%	0%	60%	60%	40%
⑥	0%	0%	80%	0%	20%
⑦	0%	0%	0%	0%	0%

Table11 「メニュー・料金グループ」の回答不可率

評価項目 パターン	ムード	サービス	テクニック	メニュー	平均
①	100%	100%	0%	60%	65%
②	0%	0%	0%	0%	0%
③	100%	100%	0%	60%	65%
④	0%	0%	0%	0%	0%
⑤	20%	0%	0%	20%	10%
⑥	100%	100%	0%	60%	65%
⑦	100%	80%	20%	100%	75%

全体的に「メニュー・料金」および「メニュー」の評価項目の結果が低くなっており, 特に Table9 では「メニュー・料金」の評価項目がパターン②を除いて青色のセルとなっているのに対して, 「技術・仕上がり」は赤色もしくは黄色のセルとなっている。

このような結果から, 今後は単語やフレーズのカテゴリ分類も手作業ではなく, 分析で判定する処理手法を検討する必要がある。

また, 仮説 2 の被験者の検証実験ではポイント予測を行ったが, 被験者が過小評価する傾向がみられた。これは投稿者の評価スコアが全体的に高いことを意味しており, 実際にレビューデータのポイントは肯定的な評価が大半を占めていた。文脈から批判や不満を読み取ったとしても, 評価ポイントが高いといったケース

が多いと考えられ, 評価ポイントだけを頼りに肯定的・否定的な内容のレビューを抽出するよりも, 本研究のようにレビュー文を参照する手法の方がより適切であると言える。

そして, 抽出したレビューの文字数が大きな差があったため, 今後は分析・抽出する際に最適な文字数や情報量を検討しなければならない。

9 今後の展望

本研究で提案した抽出方法においては, ユーザーにとって価値のあるレビューを抽出するために TFIDF による特徴語の抽出が適切であることが確認できた。今後は Word2Vec による単語のベクトル情報から TFIDF の高い単語の類似語を抽出して新たな特徴語に追加する手法を検討する。

また, 極性辞書には含まれていない評価表現語を追加した美容分野に特化した辞書を作成して導入することで, 評価表現フレーズを増やし, 本研究の仮説でより妥当性の高い検証ができる。

さらに, 本研究で用いた TFIDF は教師なし学習ではあるが, 教師あり学習でもカテゴリや評価表現の判別が可能である。そのため, 新たな抽出手法を検討し, 最終目標であるレビュー推薦システムの構築においてもっとも適切なレビュー抽出・推薦方法を開発する。

参考文献

- 1) 三菱 UFJ リサーチ&コンサルティング: “ロコミサイト・インフルエンサーマーケティングに関するアンケート調査”, 1/28(2018)
- 2) 株式会社情報通信総合研究所:” GDP に現れない ICT の社会的厚生への貢献に関する調査研究”, 1/72(2016)
- 3) 神鷹:” 推薦システムのアルゴリズム(1)”, 人工知能学会誌, 22 巻 6 号, 826/837. (2007)
- 4) <https://www.amazon.co.jp/>
- 5) <https://www.rakuten.co.jp/>
- 6) 辻井, 津田:” テキストマイニングを用いた宿泊レビューからの注目情報抽出方法”, 情報処理学会デジタルプラクティス, 3 巻 4 号, 289/296. (2012)
- 7) 川又, 立澤, 岡, 森:” 飲食店ロコミサイトを対象にした評価文抽出”, 情報処理学会研究報告, 96 巻 9 号, 1/5. (2015)
- 8) 林, 尾内:” Web 上のレビューを利用した映画推薦システム”, 人工知能学会論文誌, 30 巻 6 号, 102/111. (2015)
- 9) <https://www.recruit.co.jp>
- 10) <https://beauty.hotpepper.jp>
- 11) <https://www.nii.ac.jp/dsc/idr/recruit/recruit.html>

- 12) 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一. :” 意見抽出のための評価表現の収集”, 自然言語処理, **12** 卷 3 号, 203/222. (2005)
- 13) 東山昌彦, 乾健太郎, 松本裕治:” 述語の選択選好性に着目した名詞評価極性の獲得”, 言語処理学会第 14 回年次大会論文集, 584/587. (2008)