

NTTデータ数理システム学生研究奨励賞

# テキストマイニングによる 口コミデータを用いた顧客満足度の分析

---

東京理科大学 経営学部 経営学科  
朝日研究室 学士3年 佐藤佑泉



# 目次

1. 背景と目的
2. 関連研究
3. 分析手法
4. 分析結果
5. 考察
6. 参考文献

# 1. 背景と目的

## 背景①：美容室の競争激化

市場規模は横ばいで推移しているが、新規開店数・施設数は共に増加しており、今後はますます競争が激化していくことが考えられる。

## 背景②：口コミの重要性

初回来店時は女性の21.0%、男性の17.2%がサロン選びでネットの口コミを重要視しており、リピートを決断する際にも女性の18.5%、男性の18.8%が口コミを事前に確認していることから、安定的な美容室経営を行うためには顧客の口コミの重要性が高い。

## 目的：顧客満足度につながる要素を分析

口コミデータから、顧客満足度につながる要素として重要な要素を取り上げ、美容室の安定的な経営のための戦略を考察する。

## 2-1. 関連研究 [ホットペッパービューティデータからの不満情報抽出と解消スタイリストのレコメンド]

———— ユーザが投稿した大量のレビューから不満情報を抽出し、  
その不満を解消するようなスタイリストをレコメンドする手法を提案する

- ① 不満について言及している投稿の抽出
- ② 1投稿あたりにおける文章の分割
- ③ 不満について言及している文章の抽出
- ④ 技術に関して言及している文章の抽出
- ⑤ 不満解消スタイリストの抽出

## 2-1. 関連研究 [ホットペッパービューティータからの不満情報抽出と解消スタイリストのレコメンド]

### ① 不満について言及している投稿の抽出

サービスポイント(接客サービス)、テクニックポイント(技術・仕上がり)ともに、5点満点中3点以下の投稿を不満について言及している投稿としてみなし、抽出する。

### ② 1投稿あたりにおける文章の分割

不満について言及している投稿では、1投稿のテキストの中に不満情報以外の情報も含まれていることがあるため、主に句点を基準にして文章を分割する。

### ③ 不満について言及している文章の抽出

満足情報を取り除くため、BERTを用いたネガポジ判定を行い、ネガティブ判定された文を不満に関する文章として抽出する。

## 2-1. 関連研究 [ホットペッパービューティータからの不満情報抽出と解消スタイリストのレコメンド]

### 【BERTの補足】

- 自然言語処理モデルの1つ
- 大量のテキストデータを用いて事前学習されており、文のマスキングや次の文予測などの学習タスクにより自然言語処理タスクにおいて高い精度を発揮する
- 文章全体の文脈を考慮して、単語の意味を正確に捉えることに優れている
- Pythonなどで実装可能

# 2-1. 関連研究 [ホットペッパービューティータからの不満情報抽出と解消スタイリストのレコメンド]

## ④技術に関して言及している文章の抽出

スタイリストの技術に関する文章を抽出するために、メニュー名とセットメニュー名を用いる。レビュー文の中にメニュー名が含まれていれば、その文章を技術に関して言及している文章として抽出する。

## ⑤不満解消スタイリストの抽出

抽出した技術に関する不満レビュー文に対して、BERTで類似度が最大となるアピール文をもつスタイリストを不満解消スタイリストとして抽出する。

表1：メニューデータを用いなかった場合の文章

レビュー
1 xxxxxxxxxxxx 髪がツンツン xxx
2 xxxxxxxxxxxxxxxxxxxx 技術に満足できずがっかりです xxxx
3 xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx1 時間以上放置された時間があつた xxxxxxxx
4 xxxxxxxx 待たされた時間があまりにも長く xxxxxxxxxxxx 帰ろうかとおもいました xx
5 xxxxxxxxxxxxxxxxxxxxxxxxxxxx 接客が非常にもったいない xxxxx
6 xxxxxxxx 話しかけられるのは得意でないので xxxxxxxxxxxx 終始話しかけてきたのが不愉快でした.
7 xx 他のスタイリストさん xxxxxxxx 睨みつけてきたのが気になりました xx
8 xxxxxxxxxxxxxxxxxxxx 早口だったり話し方 xxxxxxxx 少しせわしい感じ xxxxxxx
9 xxxxxxxx 営業をかけられている xxxx 印象 xxxxx
10 xxxxxxxxxxxx ゆったりした気分を味わいたかった xxxxxxx ゆったりはできなかった xxx

表2：メニューデータを用いた場合の文章

レビュー
1 xxx 炭酸泉洗浄 xxxxxxxxxxxxxxxxxxxx 残念ながら xxx
2 ノーマルパーマ xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx なかつた xxxxxxx
3 xx サラサラストレート xxxxxxxxxxxxxxxxxxxx パサパサでまったくまとまらない xxxxxxxxxxxx
4 シャンプー・カット・パーマ・トリートメント・ブロー xxxxxxxx 残念 xxxxx パーマ xxxxxxxx
5 シャンプー ブロー xxxxxxxxxxxxxxxxxxxxxxxxxxxx トリートメント xxxxxxxx 落ちてきて xxxxxxxxxxxx
6 カット (シャンプー・ブロー込) xxxxxx 放置しすぎ xxxxx
7 xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx シャンプー・マッサージ xxxxxxxx いまいちで xxxxx
8 xxxxxxxxxxxxxxxxxxxx シャンプー・カット・パーマ xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx 不満 xxxxxxxx
9 カット xxxxx ブロー xxxxxxxx メンズカット・シャンプー・ブロー xxxxx かかってしまって xxxxxxxx
10 シャンプー・ブロー xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx 残念 xxxxx

## 2-2. 関連研究 [レビュー投稿者の甘辛度を考慮した美容院評判情報と利用者情報の可視化]

———— レビュー情報を用い、ユーザーの特徴と美容院の評価情報を可視化することで美容院の検索を支援するインターフェースを提案する

- ①ユーザーごとの特徴語の抽出
- ②投稿者の好みの施術の抽出
- ③甘辛度の定義
- ④評点分布に基づく甘辛度の可視化方針
- ⑤可視化インターフェースの実装



## 2-2. 関連研究 [レビュー投稿者の甘辛度を考慮した美容院評判情報と利用者情報の可視化]

### ①ユーザーごとの特徴語の抽出

ユーザーが書いた全レビューを投稿者ごとにまとめ、各レビューに対し形態素解析を適用し名詞を抽出する。その際、事前に指定したストップワードは除去し、助詞の前に存在する名詞のみを抽出する。TF-IDF値が閾値(0.01)以上の値となった名詞を特徴語(例：主婦、娘、土日、都心など)として抽出する。

### 【補足】

#### ストップワード

一般的であるなどの理由で処理対象外とする単語。(例：接尾辞、非自立語、代名詞)  
頻繁に現れるため、テキスト検索処理に関連する内容をもたない。

#### TF-IDF値

文章内にたくさん出てくるが、レアな単語。

$TF(\text{単語の出現回数} \div \text{その文書内にある単語の数}) \times IDF(\text{単語が含まれる文章の数} \div \text{全ての文章の数の逆数})$

## 2-2. 関連研究 [レビュー投稿者の甘辛度を考慮した美容院評判情報と利用者情報の可視化]

### ② 投稿者の好みの施術の抽出

全レビューデータから名詞のみを抽出し、その出現回数をカウントしたリストを作成する。結果、施術に関する単語の中で頻度の高かったカット、カラー、パーマ、縮毛、アレンジ、マッサージの6種類の施術方法を分析対象とした。各ユーザーのレビューから、これらの単語の出現頻度をカウントすることで、そのユーザーがどの施術について興味・関心があるのかを定量的に評価する。

### ③ 甘辛度の定義

普段から高い評点をつける傾向にあるユーザーによる評点と、逆に普段から低い評点をつける傾向にあるユーザーによる評点では、同じ評定であってもそのもつ意味は異なる。この考えの下、各ユーザーの普段の評価傾向から、そのユーザーが甘口評価なのか辛口評価なのかを定量的に評価する。

## 2-2. 関連研究 [レビュー投稿者の甘辛度を考慮した美容院評判情報と利用者情報の可視化]

### 【補足】

稀につけた普段の評価傾向と異なる評点が大きく影響してしまうため、  
評点の平均値を甘辛度とする場合と最頻値を甘辛度とする場合で場合分けを行う

評点の最頻値の相対頻度

75%以上⇒甘辛度＝評点の最頻値

75%未満⇒甘辛度＝評点の平均値

## 2-2. 関連研究 [レビュー投稿者の甘辛度を考慮した美容院評判情報と利用者情報の可視化]

### ④ 評点分布に基づく甘辛度の可視化

甘口評価のユーザーがある店舗で評価した辛口評価や、辛口評価がある店舗で評価した甘口評価  
というように、各評点とその評点をつけたユーザーの甘辛度の相対的な関係を俯瞰できるように可視化する。

### ⑤ 可視化インターフェースの実装

店舗ごとのレビュー投稿者を、評点と甘辛度を用いて散布図により可視化し、さらに散布図上のユーザー  
をクリックすることで、そのユーザーの各施術に対する興味度と特徴語を円グラフで可視化する。

## 2-2. 関連研究 [レビュー投稿者の甘辛度を考慮した美容院評判情報と利用者情報の可視化]

選択された店舗に対する  
レビュー投稿者情報が右側に描画されている

左上の点…普段は辛口のユーザーが高い評点をつけている  
右下の点…普段は甘口のユーザーが低い評点をつけている

図1の可視化結果を得たうえで  
各レビュー投稿ユーザーの特徴と評価内容を  
確認することで、希望する店舗を探し出す

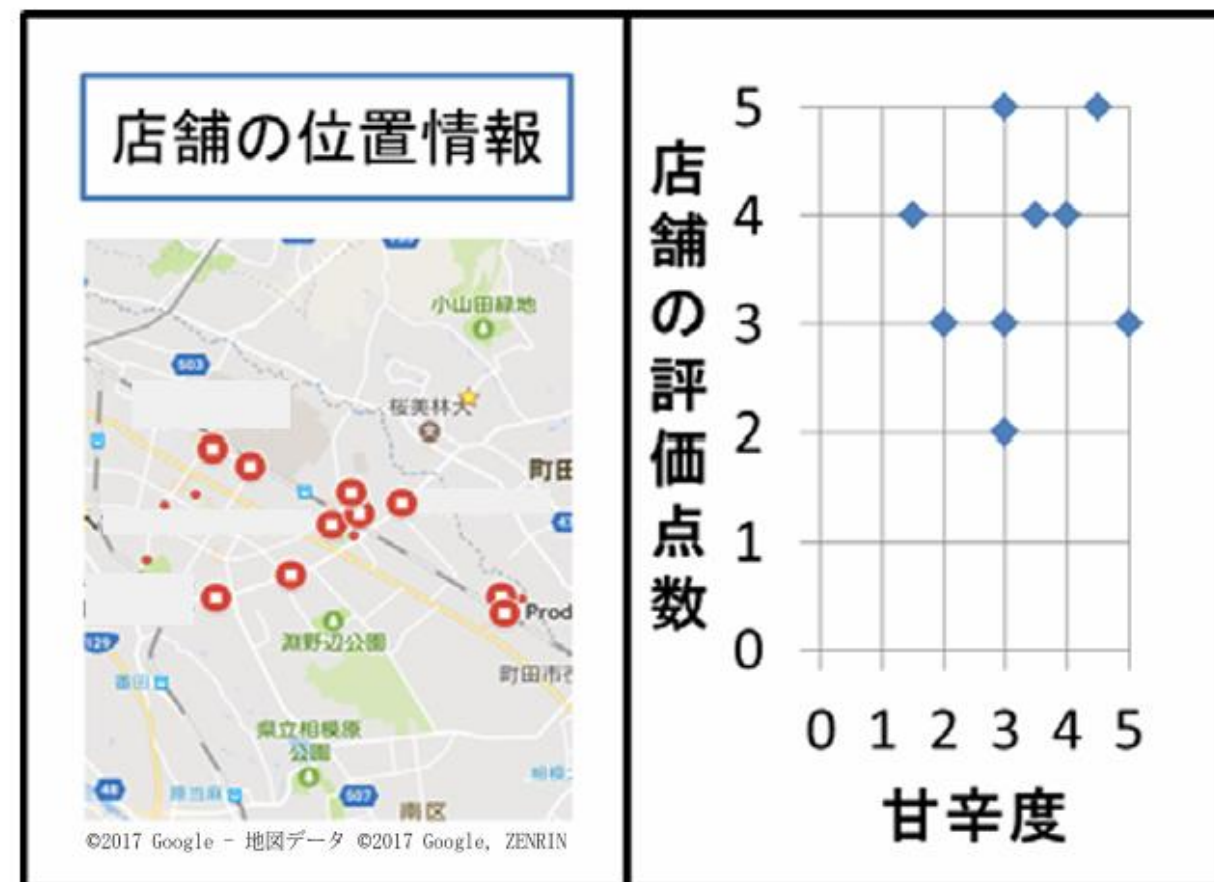


図1：甘辛度の可視化

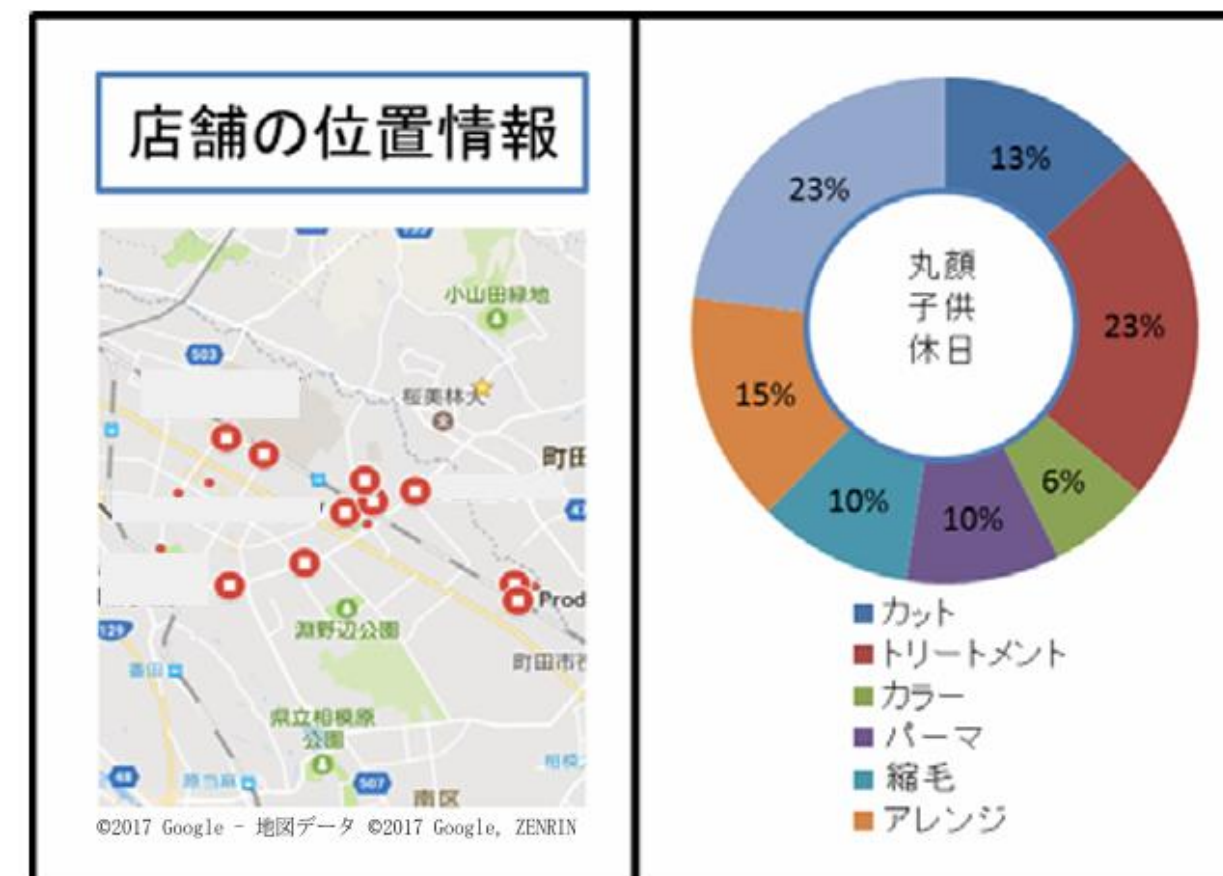


図2：好みの施術と特徴語の可視化

## 3. 分析手法 -TextMiningStudio-

### ① 特徴語抽出

総合評価が1点から5点の各評価における特徴語を抽出し  
不満足要因について言及している投稿と満足要因について言及している投稿を選別する

### ② 係り受け頻度解析

全ての口コミ投稿の中から頻度の高い係り受け表現を抽出し  
口コミに投稿されやすい単語や高評価または低評価につながる単語の抽出を行う

### ③ 評判抽出

口コミに投稿されている単語の中で高く評価されている単語と低く評価されている単語をそれぞれ抽出し  
口コミ投稿者が高く評価している傾向にある要素と低く評価している傾向にある要素を抽出する

### ④ 特徴表現抽出

総合評価が1点から5点の各評価における特徴表現を抽出し  
各評価において具体的にどの要素が低評価または高評価につながっているのかを抽出する

# 4. 分析結果 -特徴語抽出-

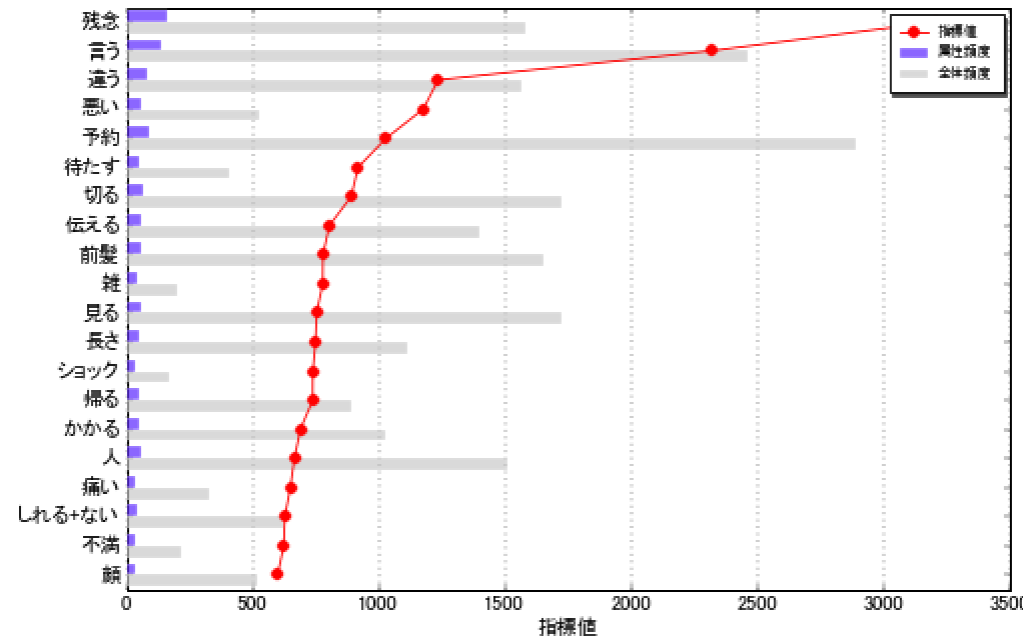


図1：総合評価1点の特徴語抽出

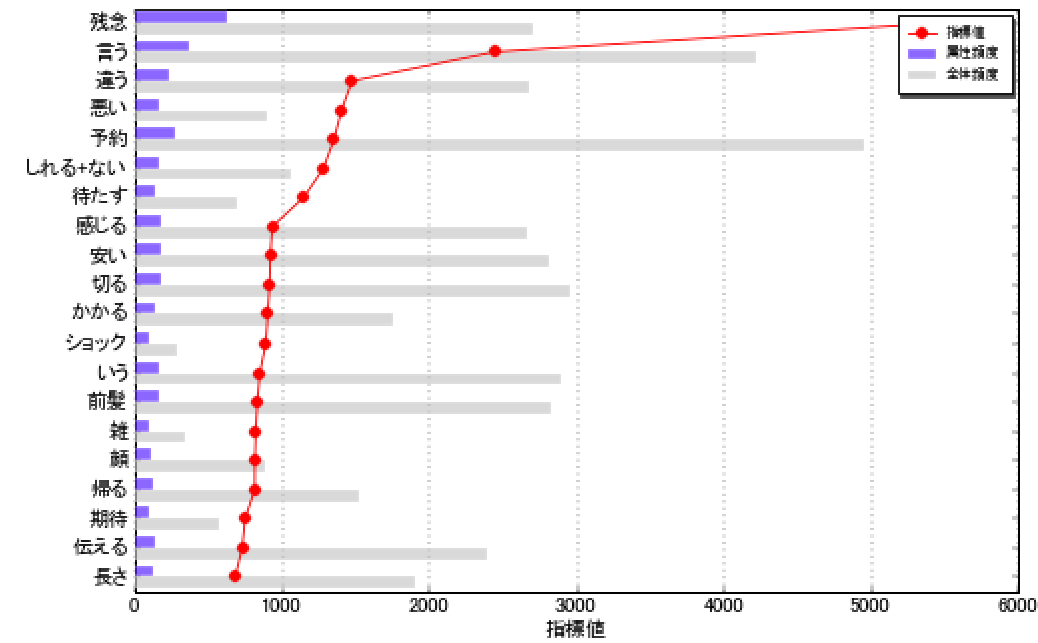


図2：総合評価2点の特徴語抽出

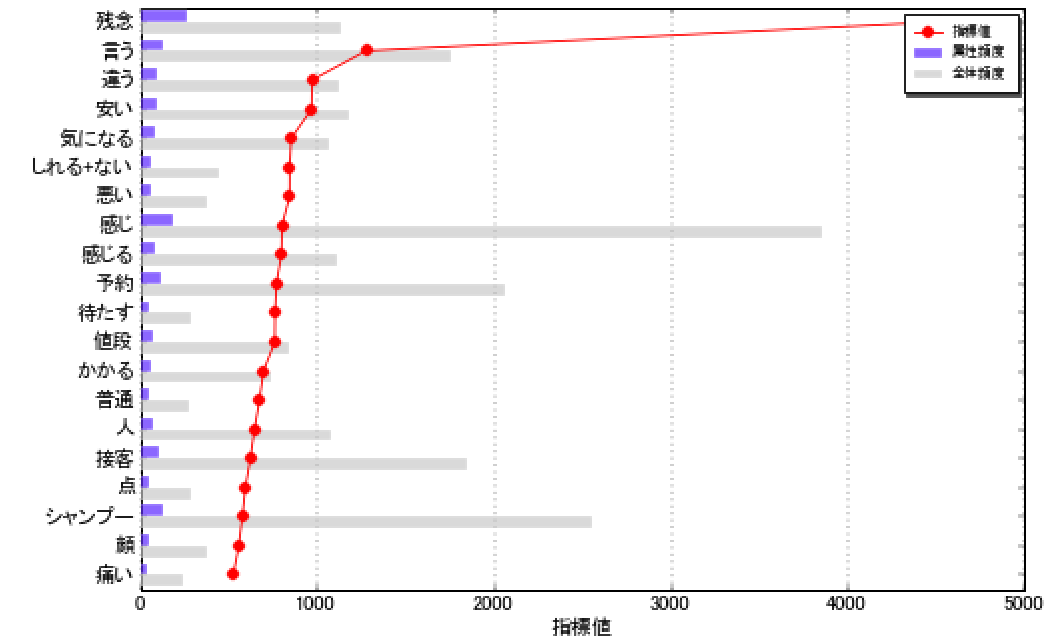


図3：総合評価3点の特徴語抽出

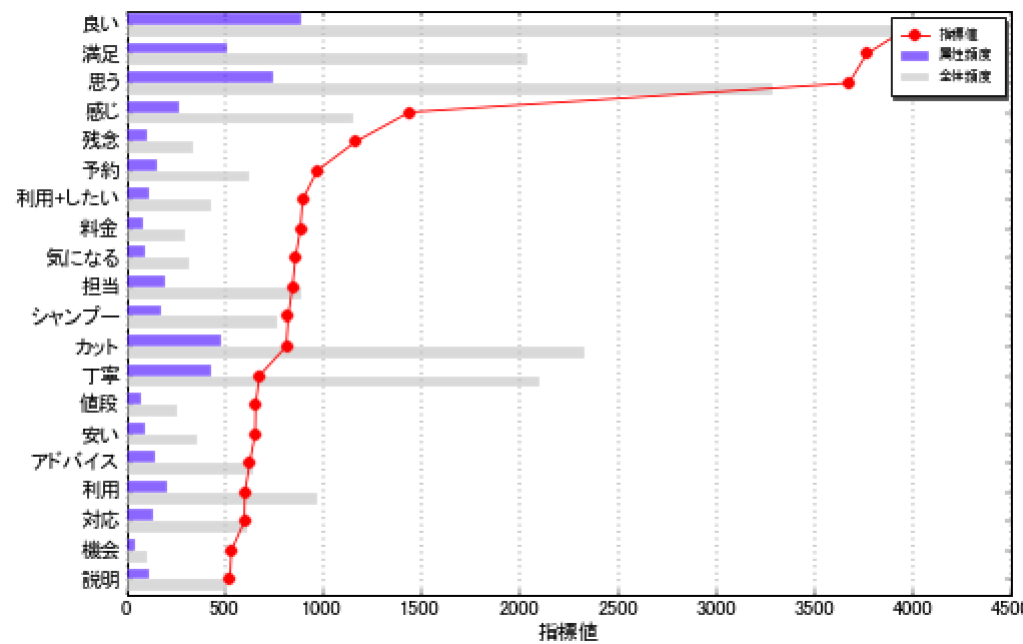


図4：総合評価4点の特徴語抽出

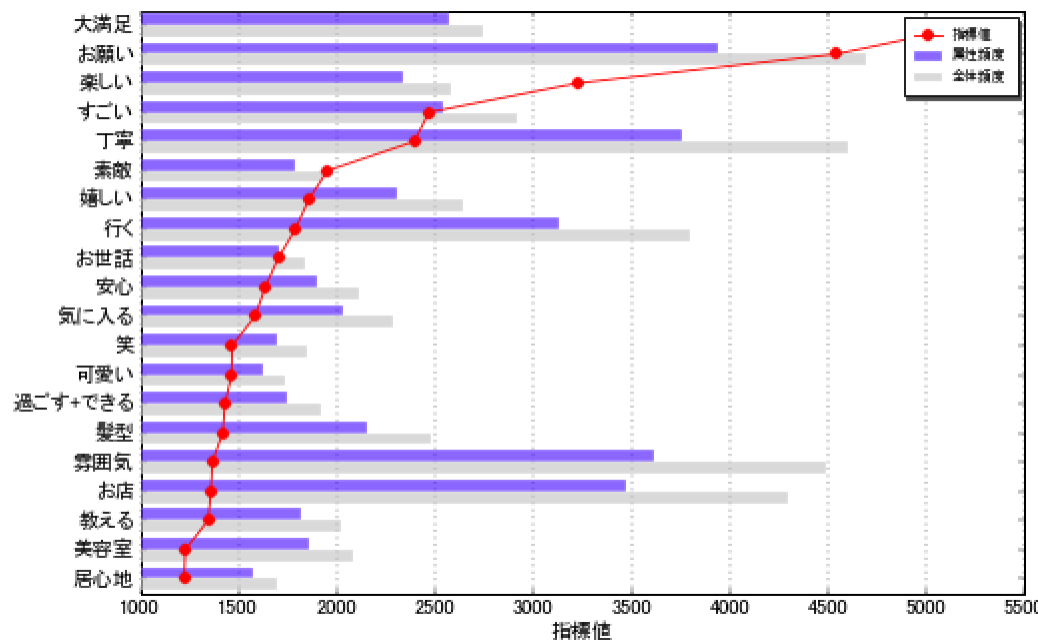


図5：総合評価5点の特徴語抽出

【総合評価が3点以下】  
不満要因について言及されている文章

【総合評価が4点以上】  
満足要因について言及されている文章

# 4. 分析結果 - 係り受け頻度解析と評判抽出 -

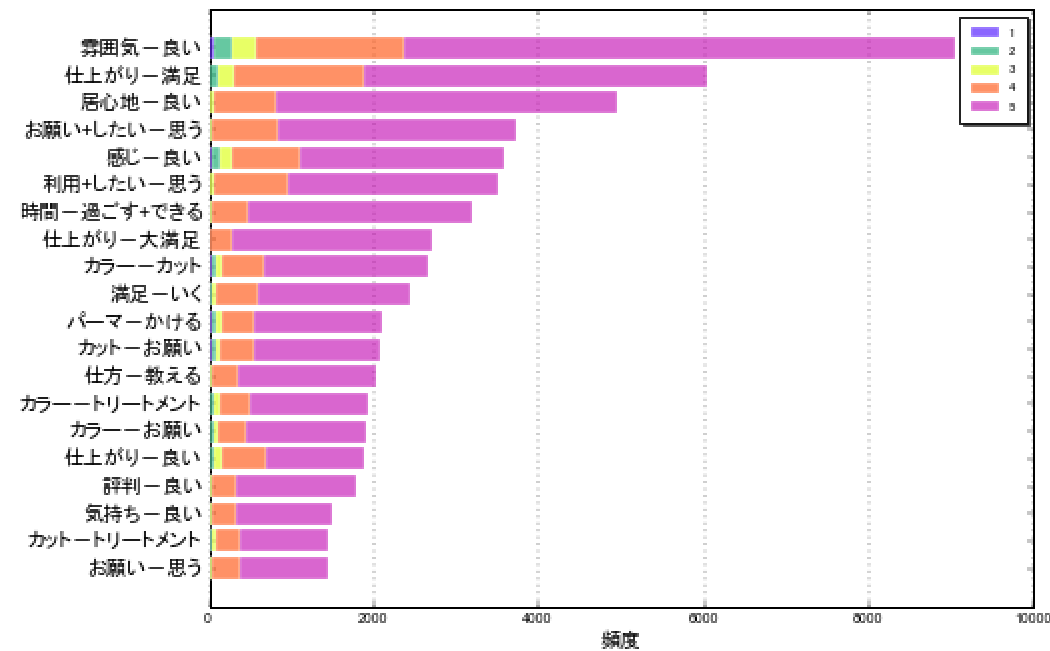


図6：総合評価と係り受け頻度解析

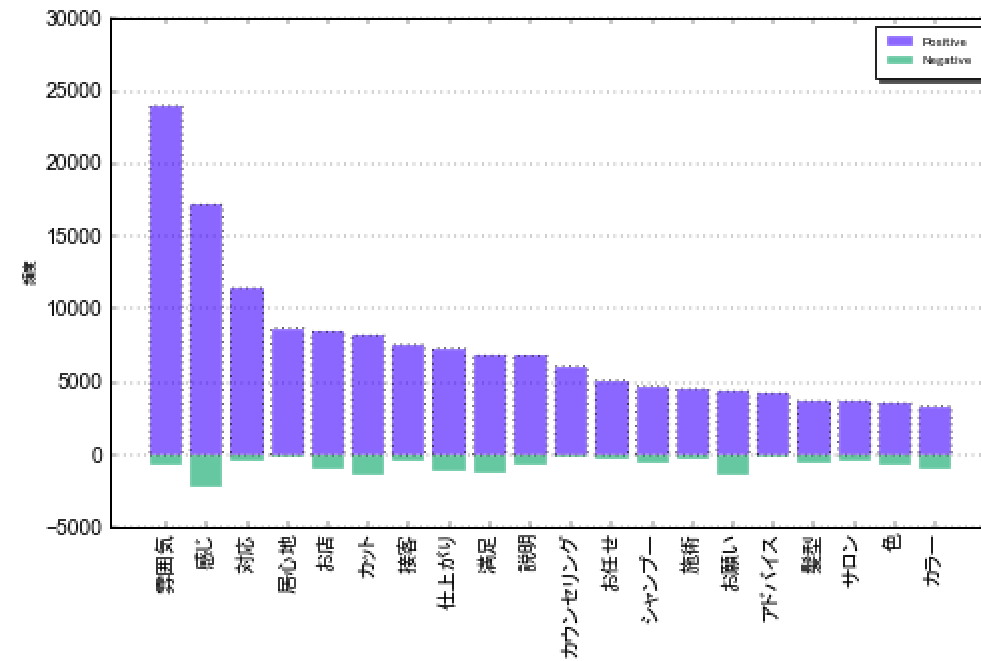


図7：好評語ランキング

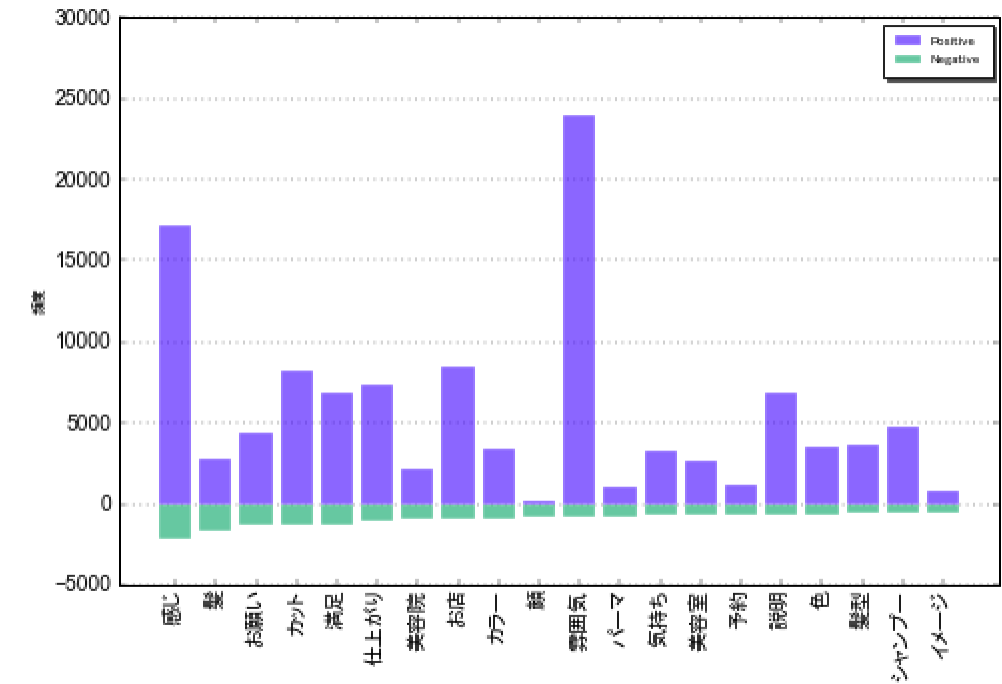


図8：不評語ランキング

「雰囲気」「仕上がり」「居心地」などの要素が口コミに投稿されやすい

お店の感じや対応、居心地などの**雰囲気**は好評であるが髪やカット、仕上がりなどの**技術面**は不評である



# 4. 分析結果 - 特徴表現抽出 -

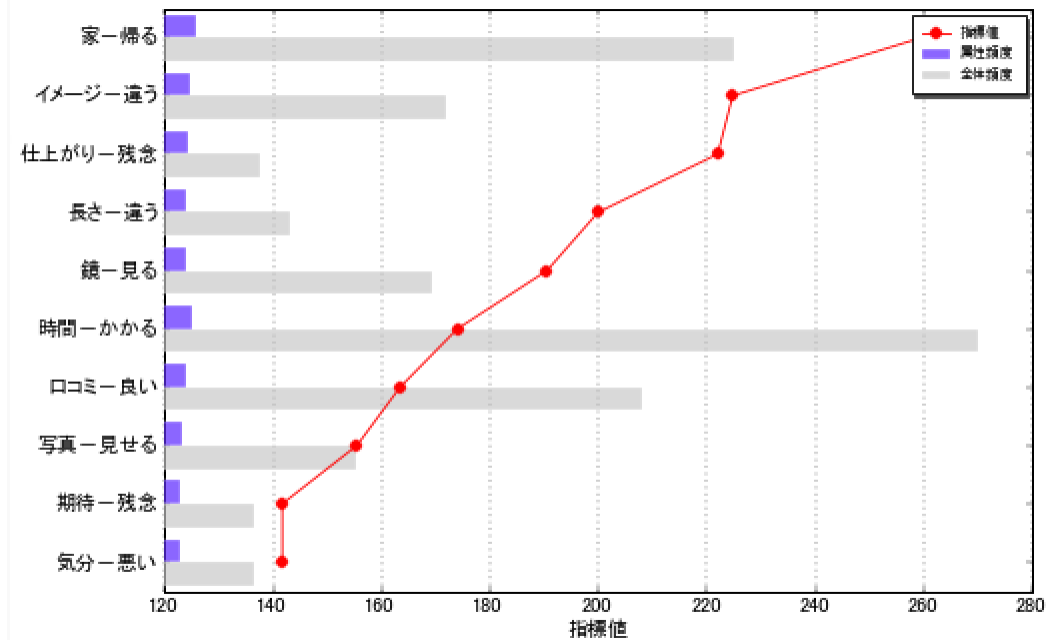


図9：総合評価1点の係り受け頻度解析

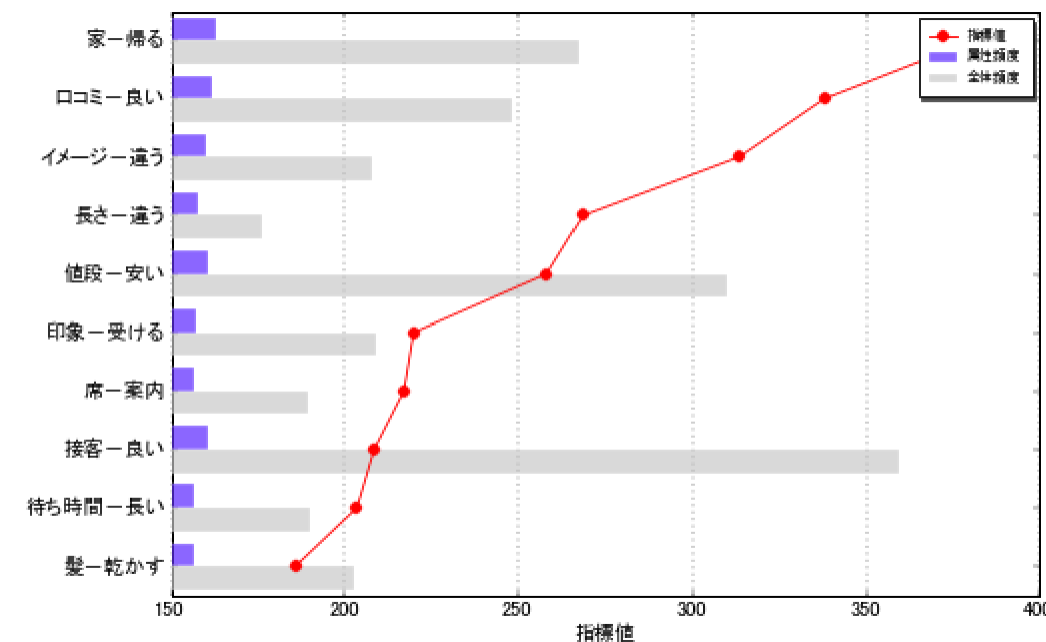


図10：総合評価2点の係り受け頻度解析

## 「家-帰る」の原文検索結果

- ・家に帰った後にスタイルが崩れた
- ・家に帰って確認すると思っていたスタイルと違った

## 「口コミ-良い」の原文検索結果

- ・口コミが良かったのに残念

総合評価が低い口コミでは  
家に帰った後のスタイリングにギャップを感じていることや  
期待をしていたことによる技術や接客のギャップが大きく影響している

## 5. 考察

### 考察：顧客満足度につながる要素

顧客満足度は「雰囲気」「仕上がり」「居心地」などの要素に影響されやすく、「雰囲気」「居心地」は高く評価される傾向にあるが「仕上がり」は低く評価される傾向にある。また、口コミやイメージとのギャップが大きいと低評価につながるため事前に顧客に適切な情報を提供することの重要性が高いと考えられる。

### 今後の展望：評点の甘辛度

普段から高い評点をつける傾向にあるユーザーによる評点と逆に普段から低い評点をつける傾向にあるユーザーによる評点では、同じ評定であってもそのもつ意味は異なる。この考えの下、各ユーザーの普段の評価傾向から、そのユーザーが甘口評価なのか辛口評価なのかを定量的に評価する必要がある。

## 6. 参考文献

- ✓ 株式会社リクルート.2014.ホットペッパービューティーデータ. 国立情報学研究所情報学研究データリポジトリ. (データセット). <https://www.nii.ac.jp/dsc/idr/recruit/>(参照2024-11-25)
- ✓ 株式会社リクルートホットペッパービューティーアカデミー.2024.数字で見る美容業界.  
[https://hba.beauty.hotpepper.jp/search\\_sp/](https://hba.beauty.hotpepper.jp/search_sp/)(参照2024-11-25)
- ✓ 株式会社リクルートホットペッパービューティーアカデミー.2023.「美容センサス2024年上期資料編(詳細版) <美容室編> 「15～69歳男女の美容サロン利用実態」. [https://hba.beauty.hotpepper.jp/wp/wp-content/uploads/2024/06/census\\_fullreport\\_hair\\_202406.pdf](https://hba.beauty.hotpepper.jp/wp/wp-content/uploads/2024/06/census_fullreport_hair_202406.pdf)(参照2024-11-25)
- ✓ 日高加菜,豊田哲也,大原合三.レビュー投稿者の甘辛度を考慮した美容院評判情報と利用者情報の可視化.情報処理学会第79回全国大会.2017.  
<https://cir.nii.ac.jp/crid/1050574047071951872> (参照2024-11-12)  
[https://www.nii.ac.jp/dsc/idr/userforum/poster/IDR-UF2017\\_P10.pdf](https://www.nii.ac.jp/dsc/idr/userforum/poster/IDR-UF2017_P10.pdf) (参照2024-11-12)