

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

日高 加菜[†] 豊田 哲也[†] 大原 剛三[†]

[†]青山学院大学 理工学部

1. はじめに

近年の Web サービスの普及により、実際に商品や店舗を利用した消費者によるレビューが数多くインターネット上に公開され、新たに利用する人の参考情報として利用されることが普通になりつつある。しかし、大量に存在するすべてのレビューを閲覧することは困難である。このような問題を解決するために、これまで商品の評判情報の抽出と可視化 [1]や、5段階評価だけでなくレビューから得られた多角的な分析結果の可視化 [2]などが提案されている。これらの既存研究では、大量の情報から商品の評価を見やすくすることができるが、レビュー投稿者ごとの価値観や好みの違い、評価基準の違いを反映させることは難しい。

そこで本研究では、美容院情報サイトを対象に、レビュー投稿者の特徴を表す単語や注文する施術の頻度、さらには評価の甘口/辛口を表す甘辛度を考慮した店舗評価を可視化することで、ユーザの店舗検索を支援するインタフェースを提案する。データとしては、国立情報学研究所を通して公開されている Hot Pepper Beauty¹のレビューデータを使用する。

2. 提案システム

本システムでは、美容院予約サイトである Hot Pepper Beauty のレビュー情報を用い、ユーザの特徴と美容院の評価情報を可視化する。具体的には、各ユーザのレビュー情報から特徴語の抽出と、好みの施術方法を抽出し、さらにユーザの甘口/辛口評価の指標をレビューの甘辛度と定義して算出する。以下、それぞれの抽出方法を説明する。

2.1 ユーザごとの特徴語の抽出

ユーザの特徴を表す特徴語をレビュー文から抽出するために、まず、ユーザ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、精度向上のため事前に指定したストップワードは除外し、さらにより特徴的な単語を抽出するために助詞の前に存在する名詞のみを抽出する。次に、1ユーザの名詞のリストを1ドキュメントとして各名詞に対する TF-IDF 値[3]を計算し、閾値以上の値となった名詞を特徴語として抽出する。本研究では、予備的な解析の結果より、その閾値を 0.01 に設定した。

2.2 投稿者の好みの施術の抽出

まず、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。対象施術の候補を限定するために、一定頻度以上出現する名詞から施術に関する単語を抽出・分類した。今回用いたデータでは、出現回数が 1,000 以上の名詞群から施術に関する単語を手動で抽出し、その結果、頻度の高かったカット、トリートメント、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。提案手法では、各ユーザのレビューから、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。

2.3 甘辛度の定義

今回用いる Hot Pepper Beauty のレビューでは、ユーザは総合、雰囲気、接客サービス、技術・仕上がり、メニュー・料金の 5 項目に対して 1 から 5 までの 5 段階の評点をつける。ただし、普段から高い評点をつける傾向にあるユーザによる評点と、逆に普段から低い評点をつける傾向にあるユーザによる評点では、同じ評点であっても、そのもつ意味は異なると考えられる。たとえば、普段は低い評点をつける頻度の高い辛口評価のユーザがつける高い評点は、普段から高い評点をつける甘口評価のユーザがつける同じ評点よりも、実際にその項目の評価が高いことを強く示唆するものと捉えることができる。このような考えの下、提案手法では各ユーザの普段の評価傾向から、そのユーザが甘口

Visualizing Reputation of Hair Salon and Reviewers' Features Considering Their Acridity

Kana Hidaka[†], Tetsuya Toyota[†] and Kouzou Ohara[†]

[†]College of science and engineering, Aoyama Gakuin University

¹<http://www.nii.ac.jp/dsc/idr/recruit/recruit.html>

評価なのか辛口評価なのかを定量的に評価する指標として甘辛度を導入する. 具体的には, いま, あるユーザがつけた評点 i の頻度を p_i , そのユーザのレビュー投稿総数を $p = \sum p_i$ としたとき, そのユーザの甘辛度 A を以下のように定義する.

$$\begin{cases} A = i & \text{if } \exists i \in \{1, \dots, 5\} \frac{p_i}{p} \geq 0.75 \\ A = \sum_{i=1}^5 \frac{ip_i}{p} & \text{otherwise} \end{cases}$$

2.4 評点分布に基づく甘辛度の可視化方針

Hot Pepper Beauty 等の美容院予約サイトに投稿されるレビューの評点は高くなる事が多く, 実際に Hot Pepper Beauty における評点の平均は 4.5 である. そのため, 単に店舗ごとにレビュー投稿ユーザの甘辛度を示しても偏りが生じ, あまり参考にならない. このような理由から, 提案システムでは, 甘口評価のユーザ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.

3. 可視化インタフェースの実装

提案システムでは, 前述の方針に従い, まず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をクリックすることで, そのユーザ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. それぞれの可視化イメージを図 1 と図 2 に示す. 各図の左側には各店舗の位置情報が地図上に描画されており, 図 1 では, 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. この散布図では, 左上の点ほど普段は辛口のユーザが高い評点をつけていることを表し, 右下の点ほど普段は甘口のユーザが低い評点をつけていることを表す. 具体的には, まず地域と自分の好みの施術を選択することで対象店舗を絞り込み, 図 1 の可視化結果を得た上で, 図 2 に示したような各レビュー投稿ユーザ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.

4. まとめ

本研究では, ユーザのレビュー内容, 数値評価評点分布を分析することでユーザの甘辛度や, 特徴語, 好みの施術を抽出した. そして, それらを可視化することで, 美容院の検索を支援す

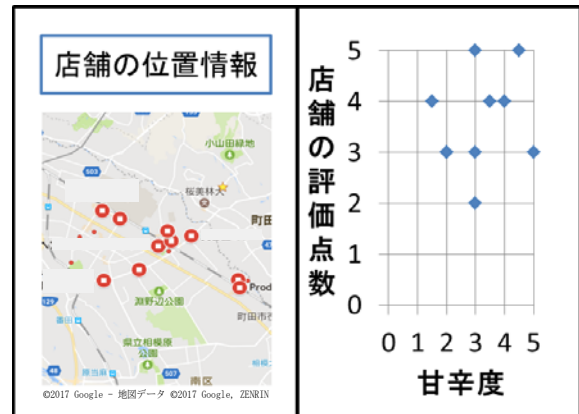


図 1. 甘辛度の可視化

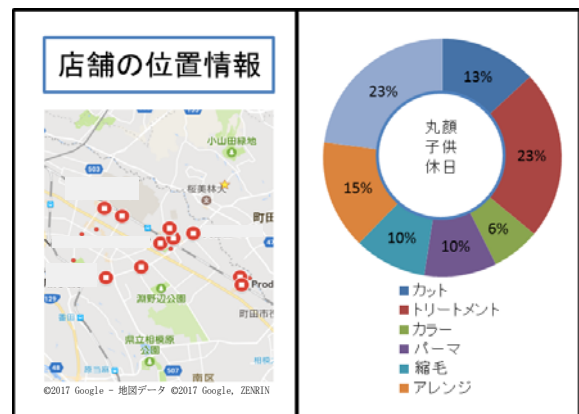


図 2. 好みの施術と特徴語の可視化

るインタフェースを提案した. これにより, 自分の要望や価値観に近いユーザを発見しやすくなり, さらにはそれらのユーザの情報を参考にすることで, 美容院選択の手助けになることが期待される. 今後, 被験者実験を通して提案インタフェースの有効性を検証する必要がある.

謝辞

本論文を執筆するにあたっては, 株式会社リクルートホールディングス様に貴重なデータを提供していただきました. 心より感謝致します.

参考文献

- [1] 金山博: テキストを用いた評判と嗜好の分析, IPSJ Magazine, Vol. 48, No. 9, pp. 1001-1007 (2007).
- [2] 打田裕樹, 吉川大弘, 古橋武, 平尾英司, 井口浩人: Web ユーザレビューにおける評価情報の時系列変化の可視化, 日本知能情報ファジィ学会誌, Vol. 22, No3, pp. 337-389 (2010).
- [3] 森純一郎, 松尾豊, 石塚満: Web からの人物に関するキーワード抽出, 人工知能学会論文誌, Vol. 20, No. 5, pp. 337-345 (2005).