

F-010

レビューサイトにおけるレビュー属性間の評価差異の検証とその補正法の検討
The verification of the evaluation difference between reviewers' attributes on a review site
and the proposal of the correction method

椎田 太輝[†] 木村 文則[‡] 手塚 太郎[‡] 前田 亮[‡]
Takaaki Shiida Fuminori Kimura Taro Tezuka Akira Maeda

1. はじめに

近年、World Wide Web 上において商品/サービスへの評価を投稿できるレビューサイトが増加している。さらに、レビューの投稿者(レビューア)が年代/性別等の属性を登録することができ、属性ごとに評価結果を集計して閲覧できるようにしたサイトも存在する。

しかし、これらのサイトでは、属性別の集計結果だけが表示されている場合が多く、異なる属性を持つレビュー集団の間で商品/サービスへの評価にどれだけの差異が存在するのかわかることはできない。ゆえに、レビューを利用する際、自分とは異なる属性を持つレビューアによる評価をそのままでは活用できないことになる。

そこで、本研究では、全国の飲食店を評価対象としたレビューサイト上の採点方式レビューデータを用いて、レビューア属性間、特に、年代と性別における評価差異の検証を行う。また評価に差異が確認された場合、その各属性間での評価得点の補正法について検討する。

2. 実験データ

今回の実験データには、レビューデータの量、評価指標が豊富であるという点から飲食店レビューサイト“全国レストランガイド[食べログ.com]”(http://tabelog.com/)におけるレビューアの飲食店評価を利用した。

この飲食店評価には、総合評価、料理、サービス、雰囲気、コストパフォーマンス(CP)、ドリンク・酒の6種類の評価基準がある。レビューアは、飲食店に対してそれぞれ5点満点(0.5点刻み)で採点評価できる。今回の実験では、総合評価を除いた、他の5つの評価を用いた。

また、レビューアの属性は、レビューアの年齢を10歳ごとに区切った年代(20代、30代、40代)と性別(男性、女性)、居住地(大阪府)といった3つの条件に絞り込み、各属性50人(合計300人)のデータを用いて実験を行った(表1)。

なお、レビューア一人当たりの評価値は、そのレビューアの過去の採点評価値の平均点を利用した。つまり、この平均点をレビューア一人分の代表値とした。また、飲食店は、大阪府内のもののみを対象とした。

表1 各属性間における実験データ量(合計300人)

	男性	女性
20代	50人	50人
30代	50人	50人
40代	50人	50人

[†]立命館大学 理工学研究科

Graduate School of Science and Engineering,
Ritsumeikan University

[‡]立命館大学 情報理工学部

College of Information Science and Engineering,
Ritsumeikan University

3. 評価差異の検証

評価差異の検証には、t検定と区間推定を用いた。t検定は、属性間におけるデータ集合の各平均が一致しないことを確認するために用いる。なお、本論文では、信頼区間95%で両側t検定を適用した。結果を表2に示す。

区間推定は、t検定によって属性間に評価差異が確認された場合、そこにはどれ程の誤差が生じているのかを確認するための指標として用いた。結果の一部を表3に示す。

表2 属性間評価差異の検証結果

属性組合せ	①	②	③	④	⑤
男,女	×	×	×	×	×
20,30	○	○	○	○	○
20,40	○	×	×	○	○
30,40	×	○	○	×	×
40男,40女	×	○	×	○	○
40男,30男	×	×	×	×	×
40男,30女	×	×	×	×	×
40男,20男	○	○	○	○	○
40男,20女	×	×	×	×	○
40女,30男	×	○	○	○	○
40女,30女	×	○	○	×	×
40女,20男	×	×	×	×	×
40女,20女	×	×	×	×	×
30男,30女	×	×	×	×	○
30男,20男	○	○	○	○	○
30男,20女	○	×	×	○	○
30女,20男	○	○	○	○	○
30女,20女	○	×	×	○	○
20男,20女	×	×	×	×	×

①料理 ②サービス
③雰囲気 ④CP
⑤ドリンク・酒

表3 区間推定結果(料理)

属性	推定区間
全体	3.58±0.125
男性	3.58±0.120
女性	3.58±0.132
40代	3.55±0.108
30代	3.49±0.112
20代	3.68±0.139
40男	3.54±0.109
40女	3.58±0.110
30男	3.46±0.114
30女	3.52±0.111
20男	3.72±0.120
20女	3.63±0.158

表2より、20代と30代、40男と20男、30男と20男の各組合せに対して、全ての評価項目で有意差を確認した。この原因として、各属性におけるレビューアの背景にある生活環境(社会的地位による年収差、加齢による嗜好性の変化)が影響を与えていると推測できる。

また、表3より、どの属性においても平均からの推定誤差は、0.1点前後程度に留まることを確認できる。この推定誤差が小さいということは、個別の店の評価の値がばらつかないというよりは、同じ属性では、意見の食い違いが少ないということが考えられる。また、店の評価として捉えた場合、影響が少ないようにも考えられるが、全国の何万といった店舗数の各ジャンルのランキングを作成する場合には、ランキング評価の変動に大きな影響をもたらすものと考えられるため、有効な指標と考えられる。次節では、これらの属性間における評価値の有意差の一例を可視化して考える。

4. 評価差異の補正

例として、2つの属性間における評価差異の位置関係を数直線上で示す。図1上は、20代男性と30代男性、下は20代男性と40代男性の推定区間を表す。

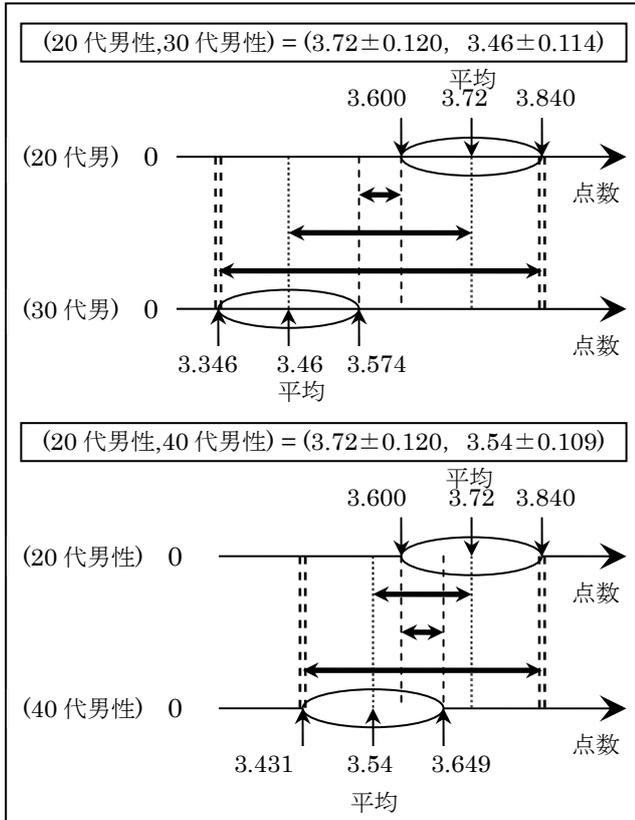


図1 属性間における区間推定結果の一例

図1より、評価差異の補正法について検討する。各属性における平均からの正負誤差を、その属性の評価範囲として捉えた場合、バイアスの強弱をつけることが可能となる。

例として、20代男性と30代男性の属性間(図1上)で考える。属性間のバイアスを最小にする場合、30代男性の評価値は3.574点、20代男性の評価値は3.600点となり、属性間には、0.026点分の差が生じる。この二者の属性を同一データで扱う際に、この差を考慮することで修正が可能となる。逆に、バイアスを最大にする場合、すなわち、30代男性の最小評価値3.346点と20代男性の最大評価値3.840点の場合には、最大で0.494点分の有意差をつけることが可能となる。この指標は、対象とする店のレビューが少なく、対象を評価しにくい場合に評価に極端な偏りを作り出しことで新たな目安を作り出すことが可能となる。また、属性間における平均値の差は、バイアスが均一に分散している状態と考えられる。

次に、20代男性と40代男性の属性間(図1下)で考える。この場合、前述の属性間と異なる点は、推定区間の一部が重複することである。前述の例と同様に考えた場合、バイアスの最大化と平均化は同じように考えられるが、最小化する場合には、一定の範囲においてバイアスが逆転してしまう。つまり、どちらの属性による評価が甘口か、辛口であるかという判断が一概にできない。このような場合、結果的にその重複する範囲を除いて考えることが望ましい。

以上より、バイアスの強弱を考えることで、実験により確認された有意差を補正できると考えられる。

また、上記とは別な方法として、各属性におけるデータの平均と分散を用いた正規化によるデータの補正法も考えられる。

5. 関連研究

小林ら[1]は、レビューサイトにおけるレビューの居住地周辺の生活領域を「生活圏」という言葉で表現し、さらに、その生活圏の規模を徒歩圏、自転車圏、自動車圏へと分類することで、ユーザに提示する飲食店情報について考慮している。また、小林ら[2]は、レビューの飲食店評価は、生活圏外と生活圏内とでは、異なることを示している。

Lauwら[3]はレビュー情報がバイアスを持つことに着目し、それを定量的に評価する手法を提案している。レビューが特定の商品が高く評価しがちである度合いを表す「バイアス」と双対の関係にあるものとして商品が異なる評価を与えられがちである傾向を表す「コントロールバー」を取り上げ、両者を求める計算をレビュー値のマトリクスを用いた固有値問題に帰着させている。さらに、Lauwら[4]は各レビューが商品全般に対してどの程度の高さのスコアをあげがちであるかを表すレニエンシー(寛容性)を考慮し、商品のクオリティーと同時に求める手法も提案している。

6. まとめと今後の課題

本研究では、インターネットユーザがより一層情報を正確に利用可能とするために、レビューサイトにおけるレビューの属性を考慮することで、それらの間での評価差異の検証とその補正法を検討した。

今後の課題として、より厳密なデータを提示するために、対象となるレビューの食の偏在性や採点評価の癖などを考慮したデータの取得法、補正法を検討する必要がある。また、レビューの地域性において、より広域化、もしくは、細分化することで生じる結果の差なども考慮することで、よりユーザの目的に応じた情報の提示法も検討していく。

参考文献

[1]小林 卓弥, 大島 裕明, 小山 聡, 田中 克己, "レビューアーの生活圏の特定に基づく CGM 評判情報評価", 情報処理学会研究報告, 2007-DBS-143, 211-216
 [2]小林 卓弥, 大島 裕明, 小山 聡, 田中 克己, "レビューアーのプロファイルと地域性に起因するバイアス補正に基づくレビュー情報の信憑性向上", 第 19 回データ工学ワークショップ(DEWS2008)論文集
 [3]H.W. Lauw, E.-P. Lim, and K. Wang, "Bias and controversy: Beyond the statistical deviation", in Proc. of the 12th ACM SIGKDD, 625-630, 2006
 [4]H. W. Lauw, E.-P. Lim, and K. Wang, "Summarizing Review Scores of "Unequal" Reviewers", in Proc. of the 7th SIAM Int'l Conf. on Data Mining, 2007