

映画レビューデータの統計的分析

総合情報学部 社会情報学科 水谷 直樹 (水谷研究室)

Keywords: マーケティング、データサイエンス、映画レビュー

1. 研究目的

WWW では、財・サービスについてのレビューを投稿できる各種のレビューサイトが運営されている。これらのサイトは、特定の財・サービスについて消費者の意見を収集できる場として企業から注目されている。

ところが、レビューデータには信頼できない情報が存在するとともに、投稿者によって評価基準や採点基準が異なることなどが、レビュー情報から経営上有用となる資料を得る上での労力を増大させている。

そこで、レビューデータの特徴を明らかにするとともに、マーケティングに役立つ情報を抽出する方法について研究を進めている。

2. レビューデータの特徴

レビューサイトの中で、投稿数が多く、レビューデータごとに投稿者IDがついているサイトとしてYahoo!映画がある。マーケティング活動において、IDつきレビューデータは有用な情報源になりうる。以下、2005年から2014年までに投稿されたレビューデータ約 105万件から得られたレビュー行動の特徴を示す。

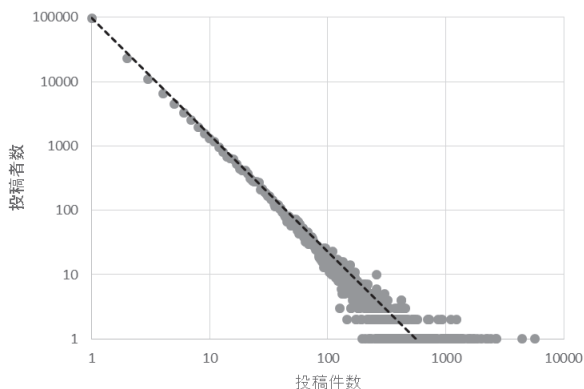


Fig. 1 投稿件数と投稿者数との関係

Fig.1 は、横軸に投稿件数、縦軸に投稿者数を取り、投稿件数ごとに投稿者数をプロットしたものである。たとえば、100件の投稿を行っ

た者は約20人存在する。投稿数が最も多い人は、5,664本の映画レビューを投稿しており、1,000本以上の映画作品のレビューを投稿した人は41名であった。投稿者数の合計は166,325名で、グラフから投稿件数と投稿者数との間に冪乗則がみてとれる。

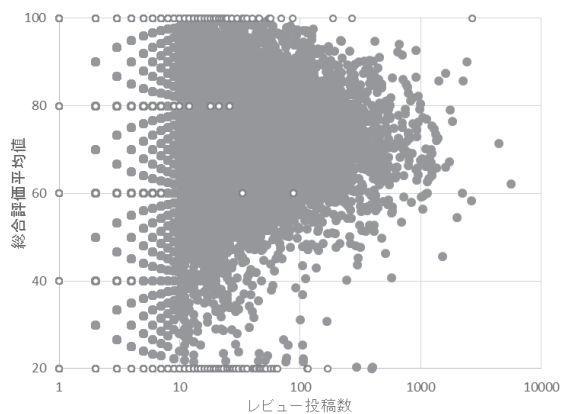


Fig. 2 投稿数と総合評価平均値との関係

このレビューサイトでは、映画に対する総合評価を5段階評定で入力する。Fig.2 は、横軸にレビュー投稿数、縦軸に総合評価の平均値をとり、レビュー投稿者ごとに総合評価平均値をプロットしたものである。たとえば、最も投稿件数が多い投稿者は、映画5,664本の総合評価の平均が62.2点である。Fig.2 から、採点基準は投稿者によって大きく異なることが推測される。また、投稿者の中には、どの映画についても同じ評価値をつける投稿者が存在することが確認できた。

消費者からの財・サービスの評価をレビュー情報から推測する場合、このような投稿者の存在を踏まえておくことが必要である。

3. まとめ と 今後の予定

映画レビューサイトにおけるレビューデータの特性について簡単に紹介した。今後はマーケティング・リサーチの観点からレビューデータの分析手法の開拓に取り組む予定である。