

口コミサイトにおける評価済み口コミ文章を用いた 評価無し口コミ文章の評価数の予測

岡部将之^{†1} 山本堅嗣宣^{†2} 山本純子^{†2} 大貫勉^{†2} 床井真理子^{†2}
原紳^{†3} 渡邊信一^{†3}

概要 : 口コミレビューサイトになどの Web サービスでは、ユーザーからの多数の口コミが投稿されているが、口コミの量が膨大なため、ユーザーが本当に欲しい情報を見つけられないことがある。そのため、ユーザーに応じた適切な口コミや評価の高い有用な口コミを表示することは満足度の向上につながる。本研究では評価済み口コミを使用し口コミがどのような評価を得るかの予測をする手法について提案する。

キーワード : ランダムフォレスト法, テキスト分類

Forecast of review sentence on word-of-mouth website using number of evaluated

MASAYUKI OKABE^{†1} MITSUNOBU YAMAMOTO^{†2}
JUNKO YAMAMOTO^{†2}
TSUTOMU OHNUKI^{†2} MARIKO TOKOI^{†2}
SHIN HARA^{†3} SHINICHI WATANABE^{†3}

Abstract: Many reviews from users are posted on web service such as word of mouth review website, but because the amount of word of mouth is enormous, users may not be able to find the information they really want. Therefore, displaying satisfactory reviews that are appropriate for the user and highly valuable reviews of the reviews leads to an improvement in satisfaction. In this research, we propose a method to predict what kind of evaluation is obtained by using the evaluated reviews.

Keywords: RandomForest, Text classification

1. はじめに

近年、インターネットにおける口コミ投稿サービスを使用する消費者は増加しており、消費者による口コミ・評価が、消費者の商品・サービス選択において有用な情報源のひとつとなっている[1][2].

口コミ投稿サービスの運用上の問題のひとつとして、ステルスマーケティングと呼ばれるものがあり、いくつかの手口が存在している。その中でサクラレビューについて着目する[3].

- サクラレビューにはいくつかの種類があり、その中でも
- 自らの店舗に都合のいい口コミを投稿すること、もしくははさせること
 - 自らの店舗に都合のいい口コミの評価ボタンをクリ

ックし、サクラレビューが多くの人に支持されているように見せること

の二つについて、本件では口コミの評価ボタンのクリック数 (=評価数)について着目し、その数を予測するし実際の評価数と比較することでサクラの有無を調査することができると考え、評価数を予測する手法について検討する。

2. 提案手法の概要

提案手法の概要図を図1に示す。口コミ投稿サイトに投稿された口コミ文章とその口コミに対する評価数をデータセットとし、評価数をラベルとして学習を行い、評価されていない口コミの分類を行うことで評価数の予測を行う。

^{†1} 宇都宮大学大学院工学研究科機械知能工学専攻
Department of Mechanical and Intelligent Engineering, Utsunomiya University.

^{†2} ヤマゼンコミュニケーションズ株式会社
Yamazen Communications Co., Ltd.

^{†3} 宇都宮大学工学部附属ものづくり創成工学センター
Utsunomiya University Innovation Center for Research and Engineering
Education

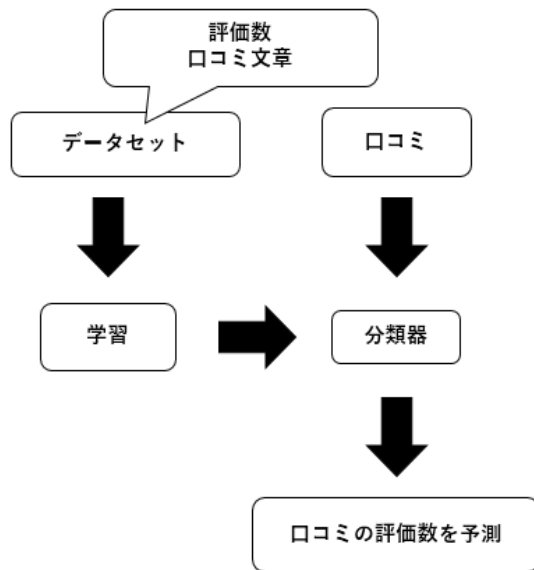


図 1 概要図

3. 実験

3.1 実験

ロコミレビューの文章の評価された回数を予測させるプログラムを作成し、その精度を計測する。今回は、分類器として、ランダムフォレスト法を使用する。

3.2 データセット

栃ナビ!内のロコミを 5000 件抽出し、データセットとして使用する。栃ナビ!とは、栃木県のヤマゼンコミュニケーションズ株式会社が運営するロコミ投稿サイトであり、飲食店や小売店、レジャー施設等のロコミが投稿されている。なお、抽出方法はある期間に掲載されたロコミを時間順に 5000 件とした。

抽出内容としては

- ロコミ文章
- ロコミに対する評価数

以上の二つを抽出した。

ロコミに対する評価数として、「グッときた」の数を使用する。「グッときた」とは栃ナビ!のサイト上のサービスのひとつであり、これは、ロコミを閲覧するユーザーが投稿されたロコミに対し、グッときたと感じた場合にサイト上のボタンをクリックすることで投稿されたロコミに対しポジティブな評価をすることができるものであり、閲覧者がサイト上で評価数を視認することができる。

3.3 実験方法

データセットのロコミを「グッときた」の個数でクラス

分けをし、それぞれの文章に形態素解析を行い、単語に分割する。分割したもので BOW (Bag of words) を作成し特徴ベクトルを生成する。なお、データセット内で出現回数が一定回数以下、および一定割合以上の登場回数の単語はフィルタをかけ取り除く。生成した特徴ベクトルに対し機械学習を行い、ランダムフォレスト法を用いてクラスを予測する。

4. 結果と考察

結果を図 2 に示す。正解率はいくつかのパターンすべてにおいて約 50%ということになった。結果から推測できることとして、あくまで文章内容のみを使用しているため、文章内容、とりわけどの単語が文章中に含まれているかは評価数に影響が無いのではないかと考えられる。

n	0	10	20	0	10	20	0	10	20	0	0	0
m	100	100	100	90	90	90	80	80	80	70	60	50
正解率	0.50	0.47	0.48	0.47	0.46	0.45	0.51	0.46	0.47	0.47	0.50	0.47

図 2 n 回未満の出現回数および m% 以上の出現率の単語を取り除いた正解率

また、栃ナビ!内のロコミの評価数は文章内容では無い部分が影響を与える可能性があると考えられるため、

- 掲載日
- 掲載日からの時間経過による評価数の変化
- ロコミ投稿者の性別、居住地、年齢といった閲覧者が視認可能なユーザー情報

などの文章以外の情報を考慮に入れ正解率を向上させたいと考えている。さらに、ロコミ文章と評価数の相関関係の有無を検討する必要がある。

5. おわりに

本件での提案システムはロコミ文章の分析によって評価数の予測を目指したが実用的な効果は得られなかった。今後はロコミ文章と評価数の関係性の有無、および評価数を決定付けるパラメータの発見が必要である。

謝辞 本研究ではヤマゼンコミュニケーションズ株式会社より提供していただいたデータセットを使用させていただきました。ここに謝意を表します。

参考文献

- 1) “平成 26 年度消費者白書” 消費者庁
http://www.caa.go.jp/information/hakusyo/2014/honbun_1_2_1_1.html#m01
- 2) “ロコミサイトと運営者の責任” 上机 美穂, 札幌大学
- 3) “ステルスマーケティングに着目した Web レビューの信憑性判断支援”. 三上 智行, 柴 直樹, 齋藤 敏雄