

# レビューサイトにおける レビューおよびレビューアの信頼性評価

中里 拓哉<sup>†1</sup> 奥野 峻弥<sup>†2</sup> 山名 早人<sup>†3†4</sup>

<sup>†1</sup> 早稲田大学基幹理工学部 〒169-8555 東京都新宿区大久保 3-4-1

<sup>†2</sup> 早稲田大学基幹理工学研究科 〒169-8555 東京都新宿区大久保 3-4-1

<sup>†3</sup> 早稲田大学理工術院 〒169-8555 東京都新宿区大久保 3-4-1

<sup>†4</sup> 国立情報学研究所 〒101-8430 東京都千代田区一ツ橋 2-1-2

E-mail: † {takuya0723, o\_syunya, yamana}@yama.info.waseda.ac.jp

**あらまし** 近年、レビューサイトが多くの人に利用されるようになった。しかし、レビューは個人の主観的な意見や感想であり、信頼性の低い情報とされている。そのため、不当な評価の検出やレビューの品質評価など、レビューの信頼性に関する研究は多く行われている。しかし、レビューの信頼性を具体的に示す数値的指標が提示されるに至っていない。そのため、レビューを参考にするか否かの判断はユーザの感覚に委ねられている。本研究では、レビューの信頼性をレビューアの信頼度を算出することで定義する。提案手法ではレビューアが投稿した各レビューをスコア付けし、全レビューのスコアからレビューアの信頼度を数値として算出し、評価する。

**キーワード** レビュー, 信頼性評価, スпам

## 1. はじめに

誰もが Web を通じて情報を発信することのできる時代となり、個人が情報を発信することを容易にする Web サービスが急速に普及している [1]。例えば、飲食店のレビューサイトである食べログ<sup>1</sup>の月間の利用者数は約 4,800 万人、PV 数は約 11 億にのぼっている [2]。しかし、レビューは個人が商品やサービスに対して感じたことを投稿する主観的評価情報である。そのため、レビューに含まれている情報は信頼性が低く、ともすれば誤った情報をユーザが信じてしまうという問題を引き起こしかねない。一方で、主観的評価情報を扱う Web サービスは年々需要を増している。実際、商品やサービスの購入時に、普段からクチコミを参考にして決めている人は全体の 4 割を占めるというデータも存在する [3]。つまり、多くのユーザにとって、主観的評価情報は信頼性の低さを加味しても利用したい、価値のある情報になっているといえる。

そこで、本稿ではレビューの信頼性を評価する指標の定義を行うことで、主観的評価情報であるレビューに客観性と根拠を付加し、レビューの価値を高めることを目標とする。

レビューに関する研究は主に、レビューおよびレビューアの品質を評価する研究と、不正な評価の検出に関する研究の 2 つに分類できる。レビューおよび、レビューアの情報の品質を評価する研究は多々存在する [4][5]。しかし、これらの研究では正確かつ、信頼性が担保された数値的評価指標は存在しない。そのため、信頼性の評価自体の信頼性がない、評価が正確でない

などの欠点がある。また、不正な評価の検出に関する研究も多々存在する [6][7]。しかし、不正評価の検出対象となるレビューを限定している、レビューの投稿時間のみを用いて不正な評価の検出を行うため、誤検出の可能性が大きいなどの欠点がある。

このように、レビュー、レビューアの品質を評価する研究や不正評価を検出する研究は多く行われている。しかし、レビュー、およびレビューアがどれだけ正確な評価を行っているかを示す、客観的かつ正確な数値指標を定義する研究はいまだ存在しない。そこで本研究では、レビューおよび、レビューアの信頼性を正確に評価する指標を作成することを目的とする。提案手法では、レビューおよびレビューアの信頼性評価と不正な評価を行っている疑いがあるスパムの検出を行う。はじめにレビューが急激に投稿されている期間を用いたスパム検出手法により、スパムレビューアおよび、スパムレビューアの投稿した全レビューを評価の対象から除外する。次にレビューに付加される数値的な評価である評価値を用いて、レビューアおよびレビューの信頼性を、それぞれレビューア信頼度とレビュー参考度という数値の形で出力する。

本稿では以下の構成をとる。まず第 2 章でレビューについて説明する。第 3 章ではレビューの信頼性評価における関連研究、スパム検出における関連研究を紹介し、第 4 章では提案手法について説明する。第 5 章では提案手法を用いて実験を行い、その結果および評価について述べる。最後に 6 章でまとめを行う。

<sup>1</sup> <http://tabelog.com/>

## 2. レビューについて

本章ではレビューについて説明する。

### 2.1. レビューとは

レビューとは商品やサービスといった対象への論評や批評のことである。その中でも、本稿では Web 上で個人が自由に投稿できるものをレビューとして扱う。またレビューはクチコミとも呼ばれるが、本稿ではクチコミもレビューと統一して表記を行う。

#### 2.1.1. 用語解説

本項では、レビューに関する用語について定義し、説明する。

##### 評価対象

レビュアーによって評価がなされる、商品やサービス、店舗などの対象物。

##### 評価値

投稿されたレビューに付加される、評価対象への数値的に取り扱える評価。例として、グルメレビューサイトの食ベログにおける、星を用いた 1.0~5.0 の間での評価が挙げられる。この数値のことを評価値とする。

##### 参考票

レビューに対して、閲覧者が当レビューを信頼できるものとして評価したことを表す票。基本的にはレビューサイトでユーザ登録を行った場合に投票することができる。

## 3. 関連研究

本章では、3.1 節でレビューおよびレビュアーの品質を評価する研究、3.2 節で不正評価の検出に関する研究について説明する。

### 3.1. レビューおよびレビュアーの品質を評価する研究

レビューに関する研究として、多く行われているもののひとつが、レビューに含まれる情報の品質や、レビュアーの情報源としての品質の評価についての研究である。そこで、レビューの品質を評価している研究として、3.1.1 項で、レビューサイトの閲覧者が付加した参考票を用いる小倉ら[4]の研究を紹介する。つぎに、レビュアーの品質を評価している研究として、3.1.2 項で同一の評価対象に投稿された各レビューの評価値を用いて品質の高いレビュアーを抽出する Tracy[5]らの研究を紹介する。

#### 3.1.1. レビューへの参考票を使用した手法

小倉ら[4]は、多くの閲覧者から参考票を付加されているレビューを良質なレビューと仮定した上で、参考票を用いてレビューの品質を算出し、選出した品質の高いレビューを用いて評価対象の品質の数値的な評

価を行った。具体的には、レビューに参考票を付加した人数と参考票を付加しなかった人数の差である参考度を用いてレビューの品質を評価する。レビューの評価値を用いて品質の高いレビューのみを利用し、評価対象そのものの品質についてのランキングを作成した。その結果、評価対象に対してより正確な評価を行っている、つまり品質の高いレビューが抽出できることを示した。

ここで、小倉らの研究では、レビューの閲覧者は、レビューが正しい情報を持つと思った場合にのみ参考票を付与すると仮定した上で、レビューの品質や信頼性の評価を行っている。しかし、レビューに対して実際に参考票を付加する閲覧者が、正確な情報を流す情報源であるかどうか、といった閲覧者の品質の評価が行われていない。そのため、悪意のある閲覧者が意図的にレビューの評価を操作することが可能であり、そのような場合には提案手法が機能しない。つまり、「レビューに対する参考票自体の信頼性が担保されていない」という問題が存在する。

#### 3.1.2. レビューの評価値を用いたレビュアーの評価手法

Tracyら[5]は、「評価対象へ下した評価が、評価対象に付加されている評価値の平均値に近いレビュアーほど信頼できるレビュアーである」という前提のもと、レビュアーの品質評価を行った。具体的には、評価対象のスコアを、当評価対象に付加されている全ての評価値から計算する。そして、レビュアーが過去に評価を行った全評価対象のスコアと、各評価対象へ当レビュアーが付加した評価値の差を算出し、その平均を 1 から引いた値が、レビュアーの基本スコアとなる。このようにして求めた各レビュアーの基本スコアに加え、レビュアーが過去に投稿した総レビュー数を用いてスコアを算出することで、各レビュアーに対する最終的な品質評価を算出する。これにより、より高い精度で正しい評価を行うレビュアーを抽出できていることを示した。

ここで、Tracyらは評価実験において、良質なレビュアーの判断基準として、当レビュアーが投稿したレビューに付加される参考票の数量のみを用いている。しかし、レビューに対する参考票自体の信頼性が担保されていない以上、参考票が多く付加されるレビューが良質であるとは限らない。当然、そのようなレビューを多く投稿するレビュアーが良質なレビュアーであるとは限らないため、Tracyらによる評価実験は十分なものとはいえない。また、Tracyらの前提と反し、実際には同一のレビュアーが投稿したレビューの品質が常に一定であるとは限らない。つまり、同一のレビュアーが投稿した各レビューの信頼性について、それ

ぞれ別々に算出する必要があるといえる。

### 3.2. 不正な評価を検出する研究

レビューについても、評価対象の不正な評価の操作を検出する研究は多く行われている。そこで本節では、不正評価を検出する研究として、3.2.1 項でレビューの投稿時間を用いる Sihong ら[6]のレビューサイトにおける不正評価検出の研究を紹介する。つぎに、3.2.2 項でレビューが急増した期間のレビュー増加率を用いる、平手ら[7]のインターネットオークションにおける不正評価検出の研究を紹介する。

#### 3.2.1. 投稿期間を用いた不正評価検出手法

Sihong ら[6]は、スパムレビュアーが、評価を不正に操作したいと考える人間に雇われた場合、不正に評価の操作を行う評価対象へ短期間に多量のレビューが投稿されると考えた。その際、スパムレビュアーはスパムの投稿によって、評価対象への評価の操作を行っていることが露見しないよう、同一の ID を用いたレビューの投稿は 1 件に留めるはずである。Sihong らはそのようなレビューを Singleton Review(SR)と定義した。さらに、各評価対象において 1) 投稿されるレビュー数、2) 全レビューにおける SR の比率、3) 評価対象へ付加された全ての評価点の平均値の 3 つが同時に、かつ急激に上昇している期間を検出し、その期間に当該評価対象へ投稿された SR をスパムと判定することで、スパムによって評価を操作している店舗の検出を行った。評価実験の結果、スパムレビューによって不正に評価の操作を行っていると思われる店舗を 75.86%の適合率で検出可能であることを示した。しかし、不正評価を行うレビュアー自体の検出は行っておらず、不正評価を行うレビューは SR に限定されると仮定しているため、複数の評価対象へレビューを投稿しているレビュアーについて考慮していない。それゆえに、複数のレビューを投稿するレビュアーについても信頼性評価を行う本手法には、そのまま適用することが出来ない。

#### 3.2.2. インターネットオークションにおける不正評価の検出手法

平手ら[7]は、サイト運営側から規約違反を行った ID と判定されたなどの理由により無効にされた ID を持つユーザを不正評価 ID として、評価実験を行った。まず、収集したユーザ ID 集合のうち、提案手法で算出した乖離率の大きさ上位 1,000 件に対して不正評価 ID の判定を行ったところ、1,000 件のうちに 54.9%の不正評価 ID が含まれる結果となった。また、乖離率の大きさ上位 1000 件に含まれる ID をシード ID として、不正評価を行う ID のコミュニティを抽出した結果、抽出したコミュニティに含まれる ID において、98.4%の適合率で不正評価 ID を検出する結果となった。この研究の問題点として、評価数における増加率の

乖離の大きさのみを用いているため、レビューサイトを対象とする本手法においては、評価対象のメディアなどへの露出といった、不正な評価の操作以外の要因による急激なレビューの増加も、不正評価によるものとして検出してしまう可能性がある。

### 4. レビューおよびレビュアー信頼性評価指標定義手法

本章では、レビューの信頼性はレビュアーが投稿した過去のレビューに依存するという仮定のもと、レビュアーの信頼性評価を行うことで、レビューの信頼性を評価する手法について述べる。まず、4.1 節で提案手法の概要について説明する。次に、4.2 節でレビュー投稿期間を用いたスパムレビューの検出手法について述べ、4.3 節ではレビュアーがどれだけ信頼できるかを示す指標であるレビュアー信頼度の算出手法について述べる。そして、4.4 節でレビューがどれだけ参考になるかを示す指標であるレビュー参考度の算出手法について述べる。

#### 4.1. 概要

第 3 章において、レビューに関する既存研究は、1) レビューおよびレビュアーの信頼性評価手法、2) 不正な評価の検出手法の 2 つに分類できることを述べた。しかし、それぞれの研究において、

- 1-i. 信頼性評価に利用した参考票の信頼性が担保されていない[4]
- 1-ii. レビューごとの信頼性のばらつきを考慮していない[5]
- 2-i. スパムの検出対象が過去のレビュー投稿数 1 件のレビュアーのみ[6]
- 2-ii. レビューの増加率のみを用いているため、スパム以外の要因によってレビューが増えた場合に誤検知する可能性がある[7]

といった問題点が存在する。そこで本稿では、それらの問題点に対応した、スパム検出手法、およびレビューおよびレビュアーの信頼性評価手法を提案する。

提案手法では、これらの問題について、以下のように解消する。

- 1-i. 信頼性の評価に参考票を利用しない
- 1-ii. レビューの投稿した全レビューについて、それぞれ信頼性を評価する
- 2-i. レビュー投稿数によらず、全てのレビュアーをスパムレビュアー判定対象とする
- 2-ii. スパム候補レビューの比率を用いてスパムレビュアーを検出し、レビューが急激に投稿されている期間に投稿されたレビューのうち、スパムレビュアーが投稿したレビューのみをスパムとする

具体的な提案手法の手順としては、以下の手順に沿う。ここで、step1 および step2 がスパム検出手法、step3、step4 および step5 がレビューおよびレビューアの信頼性評価手法に対応する。

step1. レビューの急激な投稿を示す burst の発生した期間に投稿されたレビューはスパムである疑いがあるため、burst の発生した期間に投稿されたレビュー全てに対し、スパム候補レビューのラベルを付与する。

step2. スпам候補レビューを複数投稿しているレビューアを除外するため、レビューアの投稿した全レビューにおけるスパム候補レビューが過半数を超えた場合、当該レビューアをスパムレビューアとし、当該レビューアの投稿した全てのレビューをスパムレビューとして検出し、除外する。

step3. レビューに付加される評価値が、他の多数のレビューアが付加した評価値とどの程度一致するかを示す数値的指標であるスコアをレビューごとに算出する。

step4. レビューアが投稿した全レビューについて、step3 で算出したそれぞれのスコアの平均値を求めることで、当該レビューアがどの程度多くの人と同じ評価を下してきたレビューアであるかを示す数値指標である、レビューア信頼度を算出する。

step5. step4 で算出したレビューア信頼度に対し、当該レビューアの投稿した全レビューにおける各レビューのスコアの偏差値を用いて重み付けを行った値を、各レビューがどれだけ参考になるかを示す数値指標であるレビュー参考度として算出する。

## 4.2. レビュー投稿期間を用いたスパムレビューアの検出手法

本節では、提案手法であるレビューの投稿期間を用いたスパムレビューアの検出手法について説明する。我々の提案手法は、1) 短期間で多数のレビューが投稿された期間である burst の検出、2) 検出した burst の期間に投稿されたレビューを用いたスパムレビューア判定、の2つの手順で行う。

まず、短期間で多数のレビューが投稿された burst が発生した期間を検出する。具体的には、以下の手順に従う。

step1. 評価対象  $o$  へ最初にレビューが投稿された日付から、最後にレビューが投稿された日付までを全投稿期間  $s(o)$  とし、一定の期間  $w$  で  $s(o)$  を分割する。

step2. 分割された各期間、および全投稿期間  $s(o)$  でのレビュー投稿数がどの程度増加したかを示す、レビュー増加率を計算する。

step3. 各期間でのレビュー増加率と、全投稿期間でのレビュー増加率との差の大きさを示す、乖離率を計算する

step4. 乖離率が閾値を超える期間について、burst が発生した期間とみなす。

はじめに、step1 で述べたレビュー投稿期間の分割について説明する。我々は、評価対象  $o$  へ最初にレビューが投稿された日付から、最後にレビューが投稿された日付までを全投稿期間  $s(o)$  とし、この期間をスパムレビューアの検出の際に用いる期間とした。ここで、投稿期間を  $d$  日ごとに分割する。そのため、投稿期間の分割数  $div$  は

$$div = \begin{cases} s(o)/d, & s(o) \bmod d = 0 \\ 1 + s(o)/d, & s(o) \bmod d \neq 0 \end{cases} \quad (1)$$

となる。ここで、 $s(o)$  の日数が  $d$  日で割り切れない場合、分割された投稿期間のうち、最も新しい投稿期間についての日数は  $d_{mod} = s(o) \bmod d (0 < d_{mod} < d)$  となる。

つぎに、step2 で述べたレビュー増加率の算出方法について説明する。step1 で分割された投稿期間を  $w_1, w_2, \dots, w_{div}$  とする。このとき、 $i$  番目のウィンドウ  $w_i$  の期間に投稿されたレビュー数を  $post_{w_i}$  とすると、 $s(o) \bmod d = 0$  の場合、投稿期間  $w_i$  のレビュー増加率は

$$increase(w_i) = \frac{post_{w_i}}{d} \quad (2)$$

となる。また、 $s(o) \bmod d \neq 0$  の場合、 $div - 1$  番目までは式(2)の通り求め、 $div$  番目のウィンドウ  $w_{div}$  における投稿期間を  $d_{mod}$  とすると、 $w_{div}$  における増加率は

$$increase(w_{div}) = \frac{post_{w_{div}}}{d_{mod}} \quad (3)$$

となる。同様に、全投稿期間  $s(o)$  の期間に投稿されたレビュー数を  $post_{s(o)}$ 、 $s(o)$  の日数を  $days(s(o))$  とすると、全投稿期間  $s(o)$  でのレビュー増加率は

$$increase(s(o)) = \frac{post_{s(o)}}{days(s(o))} \quad (4)$$

で求めることができる。step2 では、このレビュー増加率を分割された各期間、および全投稿期間  $s(o)$  についてそれぞれ算出する。

つぎに、step3 で述べた各期間のレビュー増加率と、全投稿期間でのレビュー増加率との乖離率計算方法について説明する。ウィンドウ  $w_i$  における乖離率は、平均的なレビュー増加率  $increase(s(o))$  を用いて式(5)のように表せる。

$$diff_{w_i} = \frac{increase(w_i)}{increase(s(o))} \quad (5)$$

この乖離率  $diff_{w_i}$  を、全てのウィンドウについてそれぞれ算出する。

最後に step4 で述べた、burst があつた期間の検出についての説明を行う。step3 で算出した各投稿期間における乖離率  $diff_{w_i}$  の値が閾値  $\alpha = 5.0$  よりも大きい場合、 $w_i$  は burst のあつた期間であるとする。ここで閾値  $\alpha$  は、

実験において求めた乖離率の平均から設定した値である。

上記の手順で *burst* が発生したと検出された期間に、評価対象 *o* へ投稿されたレビューに対し、スパム候補レビューのラベルを付加する。このスパム候補レビューが、レビュアーの投稿した全レビューのうち半数以上を占める場合、当レビュアーをスパムレビュアーとし、当該レビュアーおよび当該レビュアーの投稿した全てのレビューを、以後の信頼性評価の対象から除外する。

### 4.3. レビュー信頼度の算出手法

本節では、レビュアーがどれだけ信頼できるかを示す指標であるレビュー信頼度の算出手法を提案する。我々の提案手法は、1) レビューアーが投稿した各レビューのスコアの計算、2) 算出したレビューのスコアを用いたレビュー信頼度の計算、の2つの手順で行う。

我々は、多数のレビュアーが付与した評価値こそが、評価対象に対しての正しい評価結果であるものと考えた。そこで、信頼性評価の対象とするレビューと同一の評価値を付与したレビュー投稿数の、同一の評価対象に対し投稿された全てのレビュー投稿数に対する比率を用いたスコア付けを行う。レビューサイトの評価値を  $rate_1, rate_2, \dots, rate_m$  の  $m$  段階としたとき、評価対象  $o$  に対して、評価値  $rate_i$  の評価を行ったレビューの数を  $infl_{(o,rate_i)}$  とする。そして、レビュアー  $x$  が評価対象  $o$  に対し、評価値  $rate_i$  を付加したレビューを  $r(x, o, rate_i)$  とする。評価対象  $o$  に対して、評価値  $rate_i$  と評価をしたレビューが、評価対象  $o$  に対する全レビューの評価の中で占める割合から、レビュー  $r(x, o, rate_i)$  のスコアは

$$s(r(x, o, rate_i)) = \frac{infl_{(o,rate_i)}}{\sum_{j=1}^m infl_{(o,rate_j)}} \quad (7)$$

と求められる。ただし、レビュー数が極端に少ない場合は、割合は正しくスコアとして機能しない。そのため、 $s(r(x, o, rate_i))$  は評価対象  $o$  に投稿されているレビュー数が4以下の場合には求めず、 $r(x, o, rate_i)$  は除外する。

このように算出されたレビューのスコアは、どれだけ多くのレビュアーと同じ評価を行っているレビューであるかを示している。すなわち、スコアの平均が高いレビュアーは、高い確率で多くのレビュアーと同じ評価のレビューを投稿している、つまり高い確率で正しい情報を含むレビューの投稿を行っているレビュアーだと言える。そこで、我々はレビュアーの情報源としての信頼性を、スコアの平均値を用いて評価した。具体的には、式(7)を用いて算出される、レビューのスコアを用いてレビュー信頼度の計算を行う。レビュー信頼度は、レビュアー  $x$  が過去に投稿した全てのレビューのスコアの平均値によって求める。そこで、

レビュアー  $x$  が  $j$  番目に投稿したレビューを  $r_j(x, o, rate_i)$ 、レビュー数を  $num_x$  とすると、レビューアー信頼度  $trust_x$  は

$$trust_x = \frac{\sum_{j=1}^n s(r_j(x, o, rate_i))}{num_x} \quad (8)$$

という0から1の値となる。

### 4.4. レビュー参考度の算出手法

本節では、レビューの信頼性を示す指標である、レビュー参考度の算出手法について説明する。我々はレビュー参考度の算出にあたり、各レビューの信頼性はレビュアーに依存するが、同一のものではないと仮定した。そこで、レビューアー信頼度に対し、各レビューのスコアの偏差を用いて重みづけを行うことでレビュー参考度を算出した。我々の提案手法は、1) レビューアーが過去に投稿した全レビューのスコアにおける、各スコアの偏差値の算出、2) 算出した偏差値とレビューアー信頼度を用いたレビュー参考度の算出、の2つの手順で行う。

まず、4.3節で算出した各レビューのスコアを用いて、レビュアーが投稿した全レビューにおける、各レビューの偏差値を求める。各レビューの偏差値を求めるために、まずレビュアー  $x$  が投稿した全レビューのスコアにおける、各スコア

$$s(r_1(x, o, rate_i)), s(r_2(x, o, rate_i)), \dots, s(r_n(x, o, rate_i))$$

の標準偏差  $sd(s(r(x, o, rate_i)))$  を求める。式(8)より、

$trust_x$  はレビュアー  $x$  が投稿した全レビューのスコアの平均値なので、

$$\begin{aligned} & sd(s(r(x, o, rate_i))) \\ &= \sqrt{\frac{\sum_{j=1}^n (s(r_j(x, o, rate_i)) - trust_x)^2}{num_x}} \quad (9) \end{aligned}$$

となる。求められた標準偏差  $sd(s(r(x, o, rate_i)))$  を用いて、レビュー  $r_j(x, o, rate_i)$  の偏差値

$dev(s(r_j(x, o, rate_i)))$  は式(10)のように求める。

$$\begin{aligned} & dev(s(r_j(x, o, rate_i))) \\ &= 50 + \frac{10(s(r_j(x, o, rate_i)) - trust_x)}{sd(s(r(x, o, rate_i)))} \quad (10) \end{aligned}$$

式(10)で求めた各レビューのスコアの偏差値は、各レビューのスコアが、当レビュアーが通常投稿しているレビューのスコアとどの程度異なるかを示す指標である。我々は、レビューの信頼性は、当レビューを投稿したレビュアーに依存すると考えた。しかし、レビュアーが常に一定の信頼性を持ったレビューを投稿するとは限らない。そこで、レビューごとの信頼性のば

らつきを考慮し、上記で求めたレビュー  $r_j(x, o, rate_i)$  の偏差値  $dev(s(r_j(x, o, rate_i)))$  と 4.3 節で算出したレビューア  $x$  のレビュー信頼度  $trust_x$  を用いて、レビュー  $r_j(x, o, rate_i)$  についてのレビュー参考度  $ref(r_j(x, o, rate_i))$  を式(11)を用いて算出する。

$$ref(r_j(x, o, rate_i)) = \frac{50(trust_x)}{|dev(s(r_j(x, o, rate_i))) - 50|^2 + 50} \quad (11)$$

レビュー参考度  $ref(r_j(x, o, rate_i))$  は 0 か 1 の値となる。

## 5. 実験・評価

本章では、提案手法の実験結果と評価について述べる。まず、5.1 節で実験に利用するデータセットについて述べる。つぎに 5.2 節で、提案手法を用いたスパム検出、レビュー信頼度、およびレビュー参考度の算出結果について述べる。そして、5.3 節で、協力者による評価実験を行い、提案手法によって検出したスパム、およびレビュー信頼度、レビュー参考度との比較を行う。

### 5.1. データセット

本稿では、グルメレビューサイトの食べログ<sup>2</sup>に投稿されたレビューを収集し、データとして用いた。データの収集期間は 2013 年 12 月 20 日から 2014 年 1 月 24 日に行い、その結果として 10,064 人のレビューアにより投稿された、1,515, 720 件のレビューを収集した。そして、得られたレビューをレビューアごとに分類し、各レビューが評価対象とする店舗における、評価点の分布、および全レビューの投稿時間を取得し、データセットとする。

### 5.2. 提案手法を用いたスパム検出および信頼度算出結果

5.1 節で作成したデータセットに含まれるレビューアから、ランダムに抽出した 1,000 人のうち、評価対象に 1 件でも評価数が 4 以上の店舗があった 909 人のレビューアに対し、提案手法を用いて各レビューアのレビュー信頼度を算出した。その後、算出した全レビューアの信頼度における、標準偏差  $\sigma$  および平均値  $\mu$  を求め、算出されたレビューア信頼度によって、評価 1)  $\mu - 2\sigma$  未満、評価 2)  $\mu - 2\sigma$  以上かつ  $\mu - \sigma$  未満、評価 3)  $\mu - \sigma$  以上かつ  $\mu + \sigma$  以下、評価 4)  $\mu + \sigma$  より大きくかつ  $\mu + 2\sigma$  以下、評価 5)  $\mu + 2\sigma$  より大きい、の 5 つのグループに分類する。5 段階に分割した信頼度ごとのレビューア数を図 1 に示す。同様に、抽出した 909 人のレビューアが投稿した 64,895 レビューに対し、提案手法を用いてレビュー参考度を算出し、レビューア信頼

度と同様に 5 つのグループに分割した、参考度ごとのレビュー数の分布を図 2 に示す。さらに、4.2 節で説明した、提案手法であるスパムレビューアの検出手法を用いて、スパムレビューアと判定された 24 人のレビューアのレビューア信頼度ごとの分布を図 1 に示す。

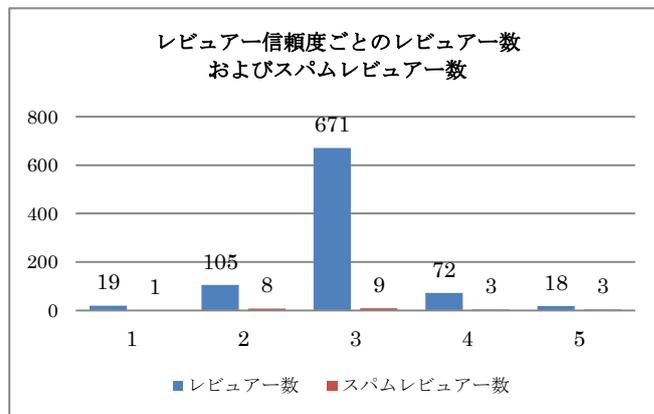


図 1 レビューア信頼度、スパムレビューアの分布

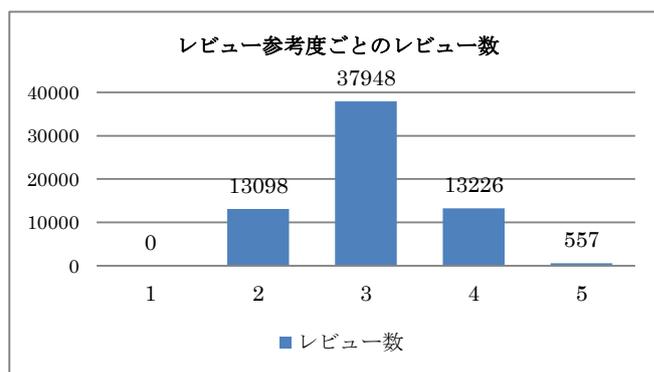


図 2 レビュー参考度の分布

### 5.3. 評価実験

5.2 節で算出したスパム検出結果について、本研究室に所属する大学生である協力者 3 人に

1. レビューアがスパムレビューアであるかどうかのスパム判定
2. レビューアがどの程度信頼できるレビューを投稿しているレビューアであるか、5 段階での評価
3. レビューが、同一の評価対象に投稿されているレビューの中で、どの程度参考になるレビューであるか、5 段階での評価

を行ってもらった。

はじめに、1 の実験について述べる。まず、提案手法を用いて、全てのレビューアを、スパムレビューア、非スパムレビューアの 2 つのグループに分類する。そして、各グループから 10% のレビューアをランダムにサンプリングを行い、3 人のスパムレビューア、88 人

<sup>2</sup> <http://tabelog.com/>

の非スパムレビュアー、合計 91 人のレビュアーの抽出を行った。つぎに、協力者に対しレビュアーがスパムであるか伏せた上で、91 人についてレビュアーごとに全レビューを提示し、レビュアーがスパムレビュアーであるか判定を行ってもらった。3 人の協力者のうち 2 人がスパムレビュアー、非スパムレビュアーと判定したレビュアーを、それぞれ正解、不正解とする。また、提案手法を用いてスパムレビュアー、非スパムレビュアーと判定したレビュアーを、それぞれ検出、不検出とする。協力者によるスパム判定の結果と、提案手法によるスパム判定結果を比較した結果を表 1 に示す。表 1 から、協力者によってスパムレビュアーと判定されたスパムレビュアーは存在せず、提案手法を用いたスパムレビュアーの検出を行うことはできなかった。

つぎに、2 の実験について述べる。まず、1 の実験でサンプリングされた 88 人の非スパムレビュアーについて、レビュアー信頼度によって 5 段階で分類を行い、評価 1) 2 人、評価 2) 10 人、評価 3) 67 人、評価 4) 7 人、評価 5) 2 人のレビュアーを抽出した。つぎに、抽出した各レビュアーが投稿した全レビューを協力者に提示し、各レビュアーが、どの程度信頼できるレビュアーであるか 5 段階での評価を行ってもらった。提案手法による各レビュアーの信頼度と全協力者の評価の平均におけるピアソンの相関係数および、ケンドールの順位相関係数を表 2 レビュアー信頼度の評価結果に示す。表 2 レビュアー信頼度の評価結果から、ピアソンの相関係数において協力者のレビュアーの信頼性評価結果と、提案手法によって算出したレビュアー信頼度には弱い相関があることがわかる。

最後に、3 の実験について述べる。まず、5.2 節で抽出した 909 人のレビュアーによって投稿された全レビューのうち、スパムレビューと判断されたものを取り除いた、64,895 レビューについてもレビュー参考度によって 5 段階の分類を行い、各評価について 0.2% のレビューをランダムにサンプリングを行い、評価 1) 0 レビュー、評価 2) 26 レビュー、評価 3) 76 レビュー、評価 4) 26 レビュー、評価 5) 1 レビューの計 129 レビューを抽出した。抽出した各レビューがどの程度参考になるレビューであるかを、当該レビューおよび、当該レビューが投稿された評価対象に投稿されている全レビューを提示し、評価対象協力者に当該レビューがどの程度参考になるレビューであるか 5 段階での評価を行ってもらった。提案手法によって算出した各レビューの参考度と、全協力者の評価の平均におけるピアソンの相関係数および、ケンドールの相関係数を表 3 レビュー参考度の評価結果に示す。表 3 レビュー参考度の評価結果から、ピアソンの相関係数において協力者によるレビューの信頼性評価の結果と提案手法によ

って算出したレビュー参考度にはほとんど相関がみられなかった。この結果から、レビュアーにより投稿されたレビューの信頼性のばらつきは、レビュアーに依存していなかったと考察される。そのため、偏差値を用いたレビュー参考度の算出は適切ではなかったと考えられる。

以上の 3 つの実験から、スパム判定についてはデータセットのサンプリング方法を再考する必要があることがわかった。また、レビュアーの信頼性評価については弱い相関がみられ、レビューの参考度の算出手法については、ほとんど相関のみられない結果となった。

表 1 協力者によるスパム判定結果(例)

提案手法	協力者評価	
	正解 (スパム)	不正解 (非スパム)
検出 (スパム)	0	3
不検出 (非スパム)	0	88

表 2 レビュアー信頼度の評価結果

	協力者 1	協力者 2	協力者 3	平均
ピアソンの相関係数	0.133	0.252	0.027	0.232
ケンドールの相関係数	-0.210	-0.098	-0.339	-0.049

表 3 レビュー参考度の評価結果

	協力者 1	協力者 2	協力者 3	平均
ピアソンの相関係数	0.124	0.135	0.011	0.176
ケンドールの相関係数	-0.171	-0.440	-0.502	0.012

## 6. おわりに

本稿では、レビューに記載される情報の正確性、つまりレビューの信頼性は、当レビューを投稿したレビュアーが過去に投稿したレビューの信頼性に依存するという仮定のもと、レビュアーのスパム判定、レビュアーの信頼性評価およびレビューの信頼性評価を行う手法を提案した。具体的には、レビューが投稿されている期間を利用することでスパムレビュアー判定を行った。さらに、レビュアーが投稿した全レビューの評価値を用いて、当レビュアーのレビュアー信頼度を算出した。そして、算出したレビュアー信頼度に対し、レビューごとに重み付けをすることで、各レビューの信頼性を、レビュー参考度として算出した。提案手法を用いることにより、レビュアーの信頼性評価は協力者による評価の平均と 0.232、レビュー参考度は 0.176

の相関係数となった。結果として、提案手法は更なる精度の向上が必要な結果となった。

本研究ではスパムレビュアーの検出の評価実験について、適切なデータセットを選択することが出来ていなかった。そのため、スパムレビュアーが含まれるようなサンプリングの手法をとる必要がある。また、スパムレビュアー特有の投稿パターンについて、投稿時間以外の考慮をしていない。そのため、スパムレビュアーが長期間による評価の操作を行っている、スパムレビューの投稿時間にばらつきがある、など提案手法では対応できない場合が多々存在する。レビュアー信頼度についても、評価対象へ投稿した時間や、評価を行う際に他のレビューから受ける影響など考慮できていない場合が多々存在した。レビュー参考後については、偏差値を用いた信頼性のばらつきの考慮が効果的でないことを示す結果となってしまった。

今後の課題としては、提案手法を用いて検出したスパムから、評価文に頻出する単語や、スパムレビュアーが多く評価を行っている評価対象を抽出し、投稿時間以外の投稿パターンを用いることで、対処しきれていないスパムレビュアーも検出できるようにすることが挙げられる。また、レビュアー信頼度についても、スコアの算出に投稿時間や、他のレビューによる影響

を反映させることが考えられる。レビュー参考度については、評価対象の評価値の割合が分散する場合を考慮することで、ばらつきを考慮することが挙げられる。

## 参 考 文 献

- [1] “What Is Web 2.0 - O'Reilly Media”, <http://oreilly.com/web2/archive/what-is-web-20.html> (2014年1月25日アクセス).
- [2] “食ベログ媒体資料 13年10月～13年12月分”, [http://corporate.kakaku.com/material/pdf/201310\\_tabelog\\_ad.pdf](http://corporate.kakaku.com/material/pdf/201310_tabelog_ad.pdf) (2014年1月25日アクセス).
- [3] NTT レゾナント株式会社: “「購買行動におけるクチコミの影響」に関する調査” (2014年1月25日アクセス).
- [4] 小倉達矢, 宍戸開, 今藤紀子, 山口実靖, 浅谷耕一: “レビューサイトにおける良質なレビューの特性とそれを考慮した評判情報の抽出に関する一考察”, DEWS2008 (2008).
- [5] Tracy Riggs, and Robert Wilensky: “An algorithm for automated rating of reviewers”, Proc. of the 1st ACM/IEEE-CS joint conference on Digital libraries, ACM (2001).
- [6] Sihong Xie, Guan Wang, Shuyan Lin, Philip S. Yu: “Review spam detection via temporal pattern discovery”, KDD'12 (2012).
- [7] Yu Hirate, Akira Aiyoshizawa, Shorei O, Yuichi Ioku, Fuyuko Kido and Hayato Yamana: “System for Extracting Auction Fraud Communities in Internet Auctions”, Proc. of ICISTM 2008 (2008).