

# レビューサイトにおける不均質性を考慮した 特異なレビュアー発見とレビューサマリの推測

川本 淳平<sup>†</sup> 浅野 泰仁<sup>††</sup> 俵本 一輝<sup>††</sup> 吉川 正俊<sup>††</sup>

<sup>†</sup>九州大学大学院システム情報科学研究所 〒819-0395 福岡市西区元岡 744

<sup>††</sup>京都大学大学院情報学研究科 〒606-850 京都市左京区吉田元町

E-mail: <sup>†</sup>kawamoto@inf.kyushu-u.ac.jp, <sup>††</sup>{asano,yoshikawa}@i.kyoto-u.ac.jp,

<sup>†††</sup>tawara@db.soc.i.kyoto-u.ac.jp

あらまし 本論文では、レビュアーサイトにおいてその不均質性を考慮した特異なレビュアー発見方法を提案する。レビューサイトにおける特異なレビュアー発見においては、レビュアー及びレビュー対象を頂点とする二部グラフを用いて反復改良を用いる手法が多く提案されている。このような手法は教師データを必要とせず、また文脈に依存せず動作する利点がある。その一方で、現実のレビューサイトのように、レビュー対象ごとのレビュー数と評価にばらつきがある及び特異なレビュアーは通常のレビュアーに比べて極めて少数であるという不均質性を持つ場合、正しく特異なレビュアーを発見できないことがある。本論文で我々は、こうした不均質性を取り扱うために、論争度と偏差希少度という二つの概念を導入し、これらを用いた新しい特異なレビュアー発見手法を提案する。また、Amazon.com から取得したレビューデータを用いて、提案手法の有効性を評価した。さらに我々の提案手法は、不均質性を持つレビューサイトにおいて、特異なレビュアー発見だけでなく少数のレビューから将来的なレビューサマリを推測する場合においても有効であることを確認した。

キーワード レビューサイト, 二部グラフ, レビュー予測

## 1. はじめに

多くのウェブサイトが何らかのレビューを提供している。これらのレビューは消費者だけでなく市場調査にとっても有益なものである。特に、レビューのサマリは個別のレビューを読まずともおおよその評価を手軽に得ることができるため、よく商品購入などの意思決定に用いられている。典型的なサマリは、例えば Amazon.com において各商品に付けられた五つ星評価値の平均値を商品評価のサマリとして用いる方法などであるが、その他にも様々なサマリ計算方法が提案されている [7, 11]。なお本稿におけるレビューとは、レイティングのような実数値として表されるものとする。

こうしたレビューにおいて、他のレビューと掛け離れたレビューを投稿する 特異なレビュー は問題の一つである。悪意あるレビューや、スパマー、ステルスマーケティングなども特異なレビューの一種である。また、アーリーアダプターや専門家も特異なレビューになり得る。専門家の意見はヘビーユーザにとってしばしば有益であるものの、ライトユーザにとって必ずしも有益とは言えないからである [15]。この特異なレビュー問題は、商品やサービスが公開されてすぐの 初期レビュー においてより顕著である。なぜなら、初期レビューではレビュー総数が未だ少ない状態であり、特異なレビューが多数を支配しやすいからである。そのため、初期レビューから求めたサマリは長期的なレビューから求めたサマリとは掛け離れることが多い。一方、初期レビューは消費者や市場調査において重要であるため、初期レビューにおける特異なレビュー問題を解決する

必要がある。

Amazon.com など多くのウェブサイトでは、ユーザに「このレビューは参考になりましたか？」などと問い投票してもらうことでスパムレビューを削除している。しかし、こうしたクラウドソースによる特異なレビュー削除には、ある程度の投票を待つ必要があり、初期レビューに対しては有効では無い。そこで、我々はレビュー総数の大小に関わらず特異なレビューを発見する方法を研究している。特異なレビュー発見は、特異なレビューによるレビューを取り除き 一般的なレビュー によるレビューを求めることができるため、初期レビューから長期的なレビューサマリの計算にも有効である。初期レビューは新しい商品やサービスに対するレビューであるため、過去に提供されている商品やサービスとは異なる評価を得ることが多い。そのため、長期的なレビューのサマリを求めるために、過去の事例を元にした教師あり学習を用いることは適していない。したがって、文脈に依存せず教師なしで長期的なレビューのサマリを求める必要がある。

レビュー評価における、文脈に依存しない教師なし学習では、二部グラフによってレビューと商品の関係をモデル化し、反復改良原理 (principle of repeated improvement) [2] を用いる方法が効果的である。この二部グラフでは、片方の頂点集合がレビューをもう片方の頂点集合が商品集合を表す。そして、レビューが商品に対してレビューを投稿した場合に限り、そのレビューを表す頂点と商品を表す頂点を枝で結ぶ。なお、各枝にはレビューの評価値を表す値をラベルとして持たせることが多い。このモデルの下で反復改良を行う方法が幾つか提案されてい

る [5, 12, 14] . しかし, これらの方法ではどれも実際のレビュー投稿サイトにおける 不均質性を十分に取り扱えているとは言い難い. 逆の状況, すなわちレビューも評価対象も本質的に均質であると仮定している例として, Lauw らによる論文査読システムにおける寛容性評価 [5] がある. この例では, すべての論文は同人数の査読者によって査読されることを仮定しており, また査読者は性善説に則り振る舞うことを仮定している. また, 寛大なレビューと厳格なレビューの間に対称性を仮定している. しかしながら, 現実的な商品やサービスのレビューについて考えてみると, こうした均質性は一般的に成り立たない. 幾つかのレビュー対象が大量のレビューを得る一方, ほとんどレビューが付かないレビュー対象もある. 特異なレビューは性善説に則り行動しているとは言いがたく, また特異なレビューと通常のレビューの間に対称性は仮定できない. したがって, レビューサイトにおける特異なレビュー発見においては, こうした不均質性をいかに取り扱うかが重要である.

本稿で我々は, 不均質性を取り扱うために, (I) 各レビュー対象における評価値の 論争度 と (II) 評価値の 偏差希少度 という二つのアイデアを下に反復改良を行う方法を提案する.

各商品ごとに設定される 論争度は, その商品に対するレビュー数と分散によって求められる. レビュー数が大きく分散が小さい場合, 論争度は小さくなる. この特徴は, 長期的にレビューがある評価に収束する場合をうまく表現している. この収束値と掛け離れたレビューを投稿するレビューは特異であると判断できる. レビュー数と分散が共に大きい場合, 論争度は大きくなる. この時, 投稿されたレビューがサマリと掛け離れていたとしても, そのレビューを投稿したレビューは特異であると判断することは難しい. 最後に, レビュー数が少ない場合, 論争度はニュートラルな値となる. 初期レビューはこのケースに含まれるが, この場合, このレビュー対象に対するレビューのみで特異なレビューであるか否かの判断は行わない. 以上をまとめると, 論争度は特異なレビュー検知において各レビュー対象をどの程度考慮するかを測る指標である.

一方偏差希少度は, 評価値ごとに設定される. レビュー  $r$  の商品  $p$  に対する評価を  $\text{rate}(r, p)$  と書き, 商品  $p$  の評価のサマリを  $\text{summary}(p)$  と書くとする. この時, このレビューによる評価の偏差は  $|\text{rate}(r, p) - \text{summary}(p)|$  と計算できる. 簡単に述べると, 偏差希少度はこの偏差がどの程度頻繁あるいは稀に発生するかを測る指標である. 反復改良を用いた従来の手法 [12, 14] では, どの偏差も等しく発生することを暗に仮定していた. しかし, 実際のレビューを調査した結果, この仮定は多くの場合成り立たないことが分かった. そして, 上記の仮定を採用した場合, 特異でないレビューに対して過剰なペナルティが与えられていることが分かった. 具体的には, 小さな逸脱は大きな逸脱よりも頻繁に発生しており, 偏差に対して線形にペナルティを与えると小さな偏差に対して過剰なペナルティとなってしまう. そこで我々は, 偏差希少度を導入し, 偏差の発生頻度を考慮してペナルティを与えることにした.

最後に, 本稿の成果をまとめる.

- 現実的なレビューにおいて特異なレビューを発見するた

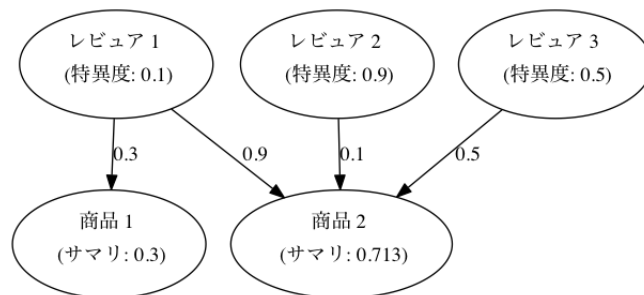


図 1: 二部グラフの例.

めには, 不均質性の取り扱いが重要であることを示した.

- 不均質性を取り扱うために, 論争度 と 偏差希少度 という二つの新しい概念を示した.
- 上記二つの概念を下に, 反復改良を用いた文脈に依存しない教師なし学習手法と提案した.
- Amazon.com より得られた実データセットを用いて, 提案手法が既存の文脈非依存の教師なし学習手法 [12, 14] より効果的であることを示した.
- 特異なレビューによる影響を削減し, 初期レビューから求めた評価のサマリが長期的なレビューから得られるサマリの予測として効果的であることを確認した.

## 2. 二部グラフモデル

本節では, レビューと商品を同時に扱う二部グラフモデルを導入する. 二部グラフは二種類の頂点集合と片方の集合に属する頂点からもう片方の集合に属する頂点への枝のみからなるグラフである. そのため, レビューと商品そしてレビューによるその商品に対する評価を表現するに適している. 従って, 我々が利用する二部グラフではレビューを一種類の頂点とし, 商品をもう一種類の頂点とする. レビューによる商品の評価は, そのレビューから商品へ枝を張りそのラベルとして表現する. また, レビューと商品にもラベルを持たせる. このラベルはそれぞれレビューの特異度と評価のサマリを表している. 我々は, この特異度をもとに評価のサマリを計算する.

[例 2.1] 図 1 は, 我々が用いる二部グラフの例を示したものである. この図は, 3人のレビューと2つの商品からなるレビュー関係を表しており, 3人のレビューの特異度はそれぞれ 0.1, 0.9, 0.5 となっている. すなわち, レビュー 2 は3人の中でもっとも特異な評価を行うレビューと判断されている. この時, 商品 2 の評価のサマリは各レビューからの評価に対し特異度で重みを付けた平均として求める.

次に, この二部グラフの形式的な定義を与える. 今,  $n$  人のレビューと  $m$  個の商品があるとすると, レビューを表す頂点集合を  $R = \{r_1, r_2, \dots, r_n\}$ , 商品を表す頂点集合を  $P = \{p_1, p_2, \dots, p_m\}$  と書く. なお本稿では, レビューは各商品に対して高々一回の評価を行うものとする. この仮定の元で, レビューは評価を行った商品とのみ枝で結ばれることになる. 従って, レビューを表す枝集合は  $E = \{(r, p) | r \in R, p \in P, r \text{ が } p \text{ を評価している場合}\}$  と定義する. 前述のように, 各

枝はその評価値を表すラベルを持つ．評価値は  $[0, 1]$  に正規化するものとし， $\text{rate} : R \times P \rightarrow [0, 1]$  と定義する．よって， $\text{rate}(r, p)$  にてレビュー  $r$  の商品  $p$  に対する評価値を表す．また，レビューはそれぞれ特異度を持つ．この特異度も  $[0, 1]$  に正規化するものとし， $\text{anomalous} : R \rightarrow [0, 1]$  と定める．評価値の定義より，各商品の評価のサマリも  $[0, 1]$  上に定義されることになり， $\text{summary} : P \rightarrow [0, 1]$  と定める．よって， $\text{anomalous}(r)$  と  $\text{summary}(p)$  はそれぞれレビュー  $r$  の特異度と商品  $p$  の評価のサマリを表す．

以上より，本稿で用いる二部グラフを

$$G = (R, P, E, \text{rate}, \text{anomalous}, \text{summary})$$

と定める．

以降の議論を簡潔に記すため，次の記号を定義する． $E_r$  はレビュー  $r$  が評価した商品集合を表すものとする．また逆に， $E_p$  は商品  $p$  を評価したレビュー集合を表すものとする．

### 3. 不均質性を考慮した反復改良

本稿で我々は，前節にて定義した二部グラフ上で動作するアルゴリズム *RIH* (*repeated improvement considering heterogeneity*) を提案する．*RIH* は反復改良原理 (principle of repeated improvement) [2] を利用しており，不均質性を扱うために二つの概念，論争度と偏差希少度を利用している．本節では，まず始めに反復改良を説明し，その後我々が導入する二つの概念を説明する．

#### 3.1 反復改良

反復改良はグラフにおける二つのプロパティ  $X$  と  $Y$  の間に，プロパティ  $X$  はプロパティ  $Y$  を持つ頂点からサポートされ，プロパティ  $Y$  はプロパティ  $X$  を持つ頂点からサポートされるという関係がある時に用いられる．*HITS* [4] や *PageRank* [1, 10] が代表的な利用例である．

この反復改良を我々の二部グラフに用いるために，特異度と評価のサマリの間に，次の二つを仮定する [12]．

[仮定 3.1] (i) 特異なレビューは評価のサマリから掛け離れたレビューを行う，(ii) 評価のサマリは普通のレビューによる評価を代表する値であるべきである．

この仮定により，特異度と評価のサマリは交互に更新を繰り返し最終的な値を求める．レビュー  $r$  の更新された特異度は，そのレビューによる各レビューがそれぞれの商品における評価のサマリとどれくらいかけ離れているかを調べ，それらを集約して求める．我々は，レビュー  $r$  による商品  $p$  へのレビューがどれくらいサマリとかけ離れているを示す指標として 部分特異度 を導入し， $a(r, p)$  と書く．一方，評価のサマリの更新は，対象の商品に対するレビューを各レビューの特異度で重みをつけた平均値として求める．これらの詳細は以降の節で説明する．

二部グラフにおける仮定 3.1 はレビューや商品が多数のレビューを持っていれば期待通りに動作する．我々はここに，現実のレビューサイトで起こりうる不均質性を取り扱うために，(I) 各商品の評価に対する 論争度 と (II) 評価の 偏差希少度 という二つの概念を導入する．本稿では，これらの概念を特異度

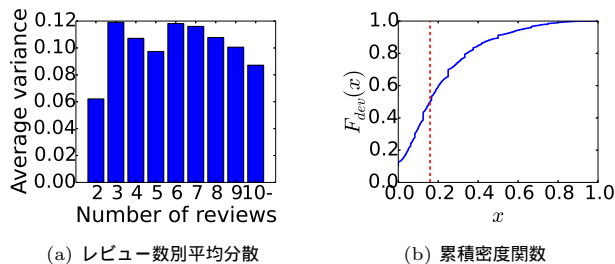


図 2: Amazon データセットの特徴．

の更新方法とともに説明する．

### 3.2 特異度の更新

#### 3.2.1 論争度

論争度に関するいくつかのコンセプトは評判分析の文脈で，*credibility* や *trustness* を扱う方法として議論されてきた [13]．我々は，レビューサイトにおける不均質性を扱うために 論争度 (*controversiality*) を導入する．レビュー  $r$  の商品  $p$  に対する部分特異度は，このレビューによる評価がどれくらい評価のサマリとかけ離れているかを示す値である．ここで，我々は商品ごとに商品によっては投稿されたレビュー数やその分散が異なるという不均質性を考慮して部分特異度を求める必要がある．投稿されたレビュー数が十分多く分散が小さければ，それらから得られた評価のサマリをそのまま用いてレビューの特異度を計算することができる．しかし，レビュー数は多いものの分散が大きければ，言い換えれば評価が割れるような商品の場合，評価のサマリから掛け離れたレビューを行ったレビューが即特異なレビューとは言い切れない．同様のことはレビュー数が少ない場合にも考えられる．そこで我々は，レビュー数とその分散を 4 つの場合に分けて考え，レビュー数が小さく分散が大きい場合，レビュー数が小さく分散が小さい場合，レビュー数が多く分散が大きい場合，レビュー数が多く分散が小さい場合の順に小さくなるような値として論争度を定義する．

本稿で我々が用いる論争度は，シグモイド関数を用いて次のような形とする．

$$1 - 1/(1 + |E_p| \sigma_p^2 - \bar{\sigma}^2) \quad (1)$$

ここで， $\sigma_p^2$  は商品  $p$  に対する評価値の不偏分散を表し，

$$\sigma_p^2 = \frac{1}{|E_p| - 1} \sum_{r \in E_p} \left( \text{rate}(r, p) - \sum_{r' \in E_p} \frac{\text{rate}(r', p)}{|E_p|} \right)^2$$

と計算されるものとする．また， $\bar{\sigma}^2$  は後で詳しく述べるが平均分散を表す．この論争度において，もし  $\sigma_p^2 - \bar{\sigma}^2 > 0$  であるなら，商品  $p$  に対するレビューの分散は相対的に大きいと判断できる．逆に， $\sigma_p^2 - \bar{\sigma}^2 < 0$  であるなら，相対的に小さいと言える．これにより，式 (1) は，レビュー数と分散について期待通りの関係を持つ．

最後に，平均分散について説明する．レビューサイトにおける不均質性を考えると，評価値の分散の平均はレビュー数によっても変わってくる．図 2a は，本稿の評価実験でも用いる

Amazon データセットにおける、レビュー数と評価値の分散を調べた結果であり、実際にレビュー数によって平均分散が異なることが見て取れる．そこで我々は、レビュー数ごとに異なる平均分散を用いることにする． $P_{|E_p|} = \{p' \in P \mid |E_{p'}| = |E_p|\}$  を商品  $p$  と同じレビュー数を持つ商品の集合を表すとする．そして、 $P_{|E_p|}$  における平均分散を  $\bar{\sigma}_{|E_p|}^2$  と表すことにする．すなわち、

$$\bar{\sigma}_{|E_p|}^2 = \frac{1}{|P_{|E_p|}|} \sum_{p' \in P_{|E_p|}} \sigma_{p'}^2$$

である．

以上を用いて、論争度を次のように定義する．

[定義 3.1] (論争度)  $\text{cont}_\alpha(p) : P \rightarrow [0, 1]$  を商品  $p$  の論争度とすると、

$$\text{cont}_\alpha(p) = \begin{cases} 0.5 & (|E_p| = 1 \text{ の場合}) \\ 1 - 1/(1 + |E_p|^{\alpha(\sigma_p^2 - \bar{\sigma}_{|E_p|}^2)}) & (\text{それ以外}) \end{cases}$$

と定める．

ここで、 $\alpha$  は評価の分散と平均分散との差  $(\sigma_p^2 - \bar{\sigma}_{|E_p|}^2)$  の影響を調整するパラメータである．なお、レビュー数が 1 の場合、論争度の判断は難しいため中立的な値として 0.5 を選んでいる．

### 3.2.2 偏差希少度

我々のこれまでの成果 [12, 15] では、レビュー  $r$  の商品  $p$  に対する部分特異度を単純に評価値とサマリの差  $|\text{rate}(r, p) - \text{summary}(p)|$  として定義していた．しかし、特異なレビューは通常のレビューに比べて少ないという不均質性があるため、この方法では通常のレビューに対するペナルティが大きすぎるということが分かった．そこで、評価値とサマリの差の分布を考慮した 偏差希少度 (*deviation rarity*) を定義する．

レビュー  $r$  が商品  $p$  に与えた評価の偏差を  $\text{deviation}(r, p)$  と書き、 $\text{deviation}(r, p) = |\text{rate}(r, p) - \text{summary}(p)|$  と定める．そして、この偏差が相対的に大きい小さいかを調べたい．なぜなら、偏差が相対的に大きいレビューは特異なレビューである可能性が高いからである．最も簡単に偏差の大小関係を調べる方法は、偏差の平均値と比較することである．偏差が平均値よりも大きければ相対的に特異なレビューであり、平均値よりも小さければ相対的に通常のレビューと言える．しかし、評価実験でも用いる Amazon.com に投稿されたレビューを調査した結果、この方法では通常のレビューに対するペナルティが大きいたことが分かった．Amazon.com に渡航されたレビューでは、偏差の平均は 0 に近い．今、偏差の分布を平均値で二分することを考える．すでに述べたように、平均値より小さい部分の幅は、平均値より大きい部分の幅より小さい．そのため、偏差が小さい場合、平均値の差分は負の小さな値となる．一方、偏差が大きい場合、平均値の差分は正の大きな値となる．その結果、特異なレビューによる影響は通常のレビューに比べて大きくなってしまふ．この問題は、特異なレビューが多数となった場合に顕著になる．特異なレビューが多数をとった場合、通常のレビューはレビューのサマリから離れた値になるが、偏差の平均値との差は大きな正の値となってしまう．反復改良原理を利用する目的は、通常のレビューによる影響を伝播させるこ

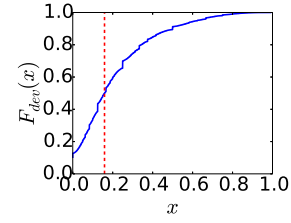


図 3: Amazon.com に投稿されたレビューの累積密度関数.

とであり、これでは逆の結果になってしまう．通常のレビューから得られる報酬が負の小さな値であるため、一度正の大きな値を得てしまうと、それを打ち消すためには数多くの通常のレビューが必要となってしまうためである．

そこで我々は、この問題を解決するために各レビューに偏差希少度を設定する．偏差希少度は、各偏差がどれほど稀で大きい、あるいは一般的に小さいかを測る指標である．もしあるレビューが偏差希少度が稀かつ大きいレビューを行えば、そのレビューは特異であると判断できる．しかし、同時に一般的に小さい偏差希少度を持つレビューを複数行っていれば、そのレビューは特異であると判断すべきではない．この偏差希少度を計算するために、まず共通の偏差を持つレビューの集合を定義する．すなわち、 $E_{\text{dev}}(v)$  にて偏差が  $v$  であるレビューの集合

$$E_{\text{dev}}(v) = \{(r, p) \in E \mid |\text{rate}(r, p) - \text{summary}(p)| = v\} \quad (2)$$

を表すことにする．この集合を用いて、偏差の分布に関する累積密度関数を定義する．

[定義 3.2] (累積密度関数) 偏差の分布に関する累積密度関数  $F_{\text{dev}}(x)$  は

$$F_{\text{dev}}(x) = \frac{1}{|E|} \int_0^x |E_{\text{dev}}(v)| dv$$

となる．

図 3 は、Amazon.com に投稿されたレビューにおける累積密度関数を表したものである． $F_{\text{dev}}(x)$  は、偏差が  $x$  であるレビューの割合を示している．例えば、 $\bar{\delta}$  偏差の平均

$$\bar{\delta} = \frac{1}{|E|} \sum_{(r,p) \in E} \text{deviation}(r, p),$$

を表すとすると、 $F_{\text{dev}}(\bar{\delta})$  は平均以下の偏差を持つレビューの割合を表し 0.5 となる．なお図では、 $\bar{\delta}$  を赤い点線で示している．従って、ある偏差  $x$  に対して累積密度関数の差分  $F_{\text{dev}}(x) - F_{\text{dev}}(\bar{\delta})$  を用いることで、その偏差が稀で大きい値か一般的で小さな値かを調べることができる．すなわち、この差分が正の値であれば、偏差  $x$  は稀で大きい値であり、負の値であれば一般的で小さい値である．偏差を直接用いる代わりに、この累積密度関数を用いることで、一般的なレビューに対する過剰なペナルティ問題を解決した．

[例 3.1] 二つのレビュー  $\text{rate}(r_1, p)$  と  $\text{rate}(r_2, p)$  の偏差がそれぞれ 0.34 と 0.050 であるとする．つまり、前者は特異なレビューであり後者は一般的なレビューである．もし、 $\bar{\delta} = 0.16$  であるなら、単純な差分は  $r_1$  にたいして 0.18 であり、 $r_2$  に

対して 0.11 より大きい値となる．一方，図 3 に示した累積密度関数を用いた場合， $|F_{\text{dev}}(\text{deviation}(r_1, p)) - F_{\text{dev}}(\bar{\delta})|$  は  $|F_{\text{dev}}(\text{deviation}(r_2, p)) - F_{\text{dev}}(\bar{\delta})| \approx 0.3$  と近い値になる．

以上より，累積密度関数による差分を用いて偏差希少度を定義する．

[定義 3.3] (偏差希少度) レビュアー  $r$  の商品  $p$  に対するレビューの偏差希少度を  $C(r, p)$  と書くと時，

$$C(r, p) = F_{\text{dev}}(\text{deviation}(r, p)) - F_{\text{dev}}(\bar{\delta})$$

と定める．

### 3.2.3 特異度の更新

前節までに導入した論争度と偏差希少度を元に，本節では，レビュー  $r$  による商品  $p$  に対する部分特異度  $a(r, p)$  を定義し，その後特異度の計算方法について説明する．

まず，各商品に対して論争度が 1 である場合を考えよう．この時，すべての商品は同等に扱われ，部分特異度は偏差希少度に対して線形に増加する関数となる．すなわち，偏差希少度が正の大きい値である時，その評価値は稀かつ大きく他と離れているため部分特異度は大きくなる．逆に，偏差希少度が負の大きい値である時，その評価値は他の評価値と近く部分特異度も小さくなる．

次に，不均質性を持つ現実のレビューについて考え，論争度が商品ごとに異なるとする．今，論争度  $\text{cont}(p)$  が大きいとしよう．すなわち，評価数が少ないか分散が大きいか，その両方である場合である．この時，偏差希少度が正の大きい値を取ったとしても部分特異度は大きくすべきではない．つまり，論争度は，特異度の計算において，ある商品に対する評価値をどの程度考慮すべきかということを表していると言える．したがって，我々は  $1 - \text{cont}(p)$  を偏差希少度の増幅に用いることにする．

[定義 3.4] (部分特異度) レビュアー  $r$  の商品  $p$  に対する部分特異度を  $a(r, p) : R \times P \rightarrow [0, 1]$  と書くと，

$$a(r, p) = \frac{1}{1 + \exp(-\beta \times (1 - \text{cont}(p)) \times C(r, p))}$$

によって求められる．なお， $\beta$  は論争度と偏差希少度の影響度を調整するためのパラメータである．

特異度計算における最後のステップは，部分特異度の集約である．我々は，新しい特異度として部分特異度の平均値を用いることにした．

[定義 3.5] (特異度の更新) レビュアー  $r$  の新しい特異度を  $\text{anomalous}_{\text{new}}(r)$  と書くと，

$$\text{anomalous}_{\text{new}}(r) = \frac{1}{N_r} \sum_{p \in E_r} a(r, p)$$

によって求められる．

### 3.3 評価のサマリの更新

評価のサマリを計算するために，我々はレビューの特異度を利用する．我々は，最も単純な方法として，評価のサマリの更新値に特異度による重みをつけた評価値の平均を用いる．具体的には，レビュー  $r$  に対して  $1 - \text{anomalous}(r)$  を重みとし

---

## Algorithm 1 RIH

---

**Require:** 二部グラフ  $G = (R, P, E, \text{rate})$  とパラメータ  $\alpha, \beta$ .

```

/* 初期化 */
for each product  $p$  do
  summary( $p$ )  $\leftarrow \sum_{r \in E_p} \text{rate}(r, p) / |E_p|$ 
end for
repeat
  /* 特異度の更新 */
  累積密度関数  $F_{\text{dev}}(x)$  の計算
  for each reviewer  $r$  do
    anomalous( $r$ )  $\leftarrow \frac{1}{N_r} \sum_{p \in E_r} a(r, p)$ 
  end for
  /* 評価のサマリの更新 */
  for each product  $p$  do
    summary( $p$ )  $\leftarrow \frac{\sum_{r \in E_p} (1 - \text{anomalous}(r)) \times \text{rate}(r, p)}{\sum_{r \in E_p} 1 - \text{anomalous}(r)}$ 
  end for
until anomalous( $r$ ) と summary( $p$ ) が収束する
return anomalous( $r$ ) 及び summary( $p$ )

```

---

て用いる．その結果，特異度が大きいレビューの影響を小さくする．

[定義 3.6] (サマリの更新) 商品  $p$  の新しい評価のサマリを  $\text{summary}_{\text{new}}(p)$  と書くと，

$$\text{summary}_{\text{new}}(p) = \frac{\sum_{r \in E_p} (1 - \text{anomalous}(r)) \times \text{rate}(r, p)}{\sum_{r \in E_p} 1 - \text{anomalous}(r)}$$

によって計算する．

## 3.4 アルゴリズム

我々が提案する RIH (Repeated Improvement considering Heterogeneity) アルゴリズムをアルゴリズム 1 に示す．なお，左向きの矢印 ( $\leftarrow$ ) は更新された新しい値の割り当てを表すものとする．このアルゴリズムは，2. 節で定義した二部グラフ  $G$  と 3.2 節で説明した二つのパラメータ  $\alpha, \beta$  を受け取り，レビューアと商品に対してそれぞれ特異度と評価のサマリを計算する．

RIH アルゴリズムは，まず初期値としてそれぞれの商品に対して与えられた評価値の平均を初期サマリとして割り当て，その後反復改良を行う．累積密度関数  $F_{\text{dev}}(x)$  の計算には，式 (2) にて定義した評価のサマリから  $v$  だけ離れたレビューの数  $E_{\text{dev}}(v)$  の計算が必要である．反復改良の試行ごとに評価のサマリは更新されるため，累積密度関数も反復改良の試行ごとに再度計算する必要がある．累積密度関数の計算が終わると，定義 3.5 にて定めた方法により，各レビューアの特異度を更新する．そして，定義 3.6 にて定めた方法により，各商品ごとに評価のサマリを更新する．

反復改良は，この更新幅が無視できるほど小さくなった時，終了する．RIH アルゴリズムは，最後に得られたレビューアごとの特異度と商品ごとの評価のサマリを出力する．

## 4. 評価実験

我々は，Amazon.com におけるレビューを用いて，RIH アルゴリズムにおけるパラメータの影響を調べ，特異なレビュー発見及び長期的な評価のサマリ推測の精度を既存手法と比較した．

表 1: レビューとレビュー数.

2004 年以前に投稿されたレビュー数	1555315
2005 年以降に投稿されたレビュー数	613265
2004 年以前に投稿を行ったレビュー数	730667

#### 4.1 データセット

本評価実験では、2006 年までに Amazon.com へ投稿されたレビューからなるデータセット [3] を用いた。その中から、書籍カテゴリに属する商品のみを対象とした。一般的に、短期間で陳腐化してしまう家電やデジタル製品と比べて、書籍は長期間評価が一定である傾向が強く、長期間評価値の予測に適しているため書籍のみを対象としている。以降では、本実験で用いるデータセットを単に *Amazon* データセットと呼ぶことにする。

我々は、2004 年以前に投稿されたレビューを用いて、レビューの特異度を計算・評価した。また、同時に 2004 年以前に投稿されたレビューを初期レビューとみなして評価のサマリを計算する。そして、2005 年以降に投稿されたレビューも含めて、Amazon データセットに含まれるすべてのレビューを用いて計算したサマリを長期的なレビューサマリとみなし、初期レビューから計算したサマリの精度を比較した。なお、長期的なレビューからサマリを求める方法は、単純な平均を用いた。

一般的に、電子商取引サイトはスパムレビューを削除しており、Amazon データセットにおいても同様であると考えられる。そのため、Amazon データセットに含まれる特異なレビュー数は現実よりも少なくなっていると言える。そこで、次の方法で人工的に特異なレビューを追加することにした。我々は、特異なレビューとして独立したレビューと結託するレビューの二種類を考える。独立した特異なレビューは、その名の通り特異なレビューを独立に行動する。一方、結託する特異なレビューは、特異なレビューを複数人で結託して投稿し、対象となる商品のレビューにおいて多数を取ろうとするものである。ひとたび特異なレビューが多数を取ると、通常のレビューが特異なレビューに見えてしまうため、特異なレビュー発見は困難となる。ソーシャルスパマー やステルスマーケティングなどでしばしば見られる方法であり、こうした結託する特異なレビュー発見はレビューの価値を保つために重要である。

我々は、Amazon データセットにおけるレビューの 5% に当たる 18,300 の独立した特異なレビューを人工的に追加した。また、同様に 18,300 の結託する特異なレビューも追加した。Amazon データセットに元々含まれていたレビューから一人あたりの投稿レビュー数の分布を作成し、独立した特異なレビューは、この分布に従う乱数によって投稿するレビュー数を決定した。また、投稿先の商品は一律ランダムに選ぶことにした。投稿する特異なレビュー値は、一般的なレビューによる投稿レビュー値の平均が 3.5 以上であれば 1 とし、3.5 より小さい時 5 とした。すなわち、一般的なレビューと真逆の行動を取るように定めた。

結託する特異なレビューは、4 人のレビューをグループとして扱った。すなわち、4575 の結託するレビューグループを作

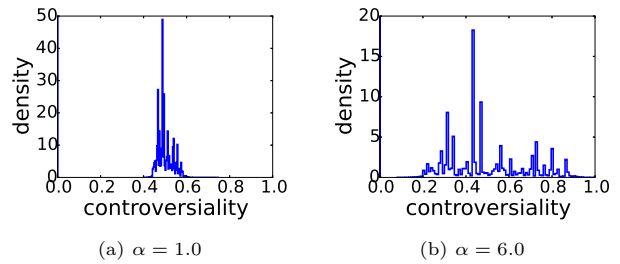


図 4: パラメータ  $\alpha$  による論争度分布の比較.

成した。各グループは、2 人のレビューからのみ評価されている書籍をランダムに 2 冊選び、この 2 冊に対して結託して特異なレビューを投稿することとした。特異なレビューの値は独立した特異なレビューの場合と同様に、投稿レビュー値の平均が 3.5 以上であれば 1 とし、3.5 より小さい時 5 とした。したがって、この 2 冊においては結託するレビューは必ず多数となる。大量の特異なレビューが結託する事象は我々の知る限り稀であり、また大規模に特異なレビューを投稿する行為は一般的なスパム検知手法 [8,9] が利用出来る。そのため、本評価実験では少数グループによる結託の発見を課題として設定した。

最後に、本評価実験にて用いた提案手法 (RIH) に対する比較手法について説明する。ONE は、文献 [6] にて提案されている別の特異度を用いた方法であり、反復改良は行わず一度の計算で特異度を求める手法である。MRA は、我々の既存手法 [12] であり、反復改良を用いているが不均質性を考慮していない手法である。RSD は、文献 [14] にて提案されている反復改良を用いる異なる手法である。RSD は、タイムウィンドウを用いてレビューの変遷を取り扱えることが特徴となっている。RSD には、タイムウィンドウ幅を決めるパラメータがある。以降では、最も結果の良かった 0.001 を用いた場合のみを掲載する。

#### 4.2 パラメータ決定

始めに、提案の RIH アルゴリズムにおける Amazon データセットに適したパラメータの決定を行う。アルゴリズム 1 に示したように、RIH アルゴリズムには二つのパラメータ  $\alpha, \beta$  がある。パラメータ  $\alpha$  は定義 3.1 にて定めた論争度 (controversiality) において、分散幅の制御に用いられている。図 4 に、 $\alpha$  の値による論争度分布の違いを示す。横軸が論争度を表し縦軸が密度である。なお、この分布図は総ビン数は 100 であり総面積が 1 となるように正規化している。図 4a によると論争度は [0.4, 0.8] の間のみ分布しているが、図 4b では [0.1, 1.0] の間に分布していることが分かる。結果のみ述べると、論争度は [0, 1] の間に広く分布している方が好ましいため、以降では  $\alpha = 6$  を用いる。

次に、定義 3.4 にて定めた部分特異度を用いるパラメータ  $\beta$  について考える。パラメータ  $\beta$  は、計算された特異度の精度及び求めた評価のサマリと長期レビューサマリとの誤差という二つの目的を考慮して選ぶ必要がある。RIH アルゴリズムでは、各レビューに特異度を割り当てる。我々は、簡単にその特異度

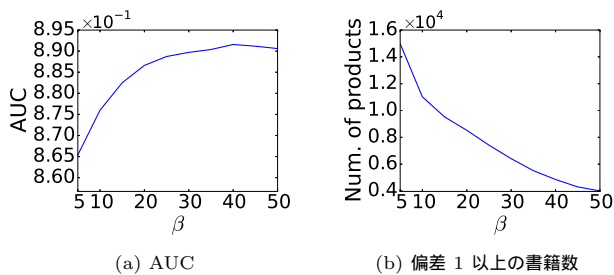


図 5: パラメータ  $\beta$  による影響の比較.

表 2: 特異なレビュー検出における AUC の比較.

アルゴリズム	AUC
RIH	0.8916
ONE	0.8697
MRA	0.9092
RSD	0.6481

が閾値以上であれば、すなわち

$$\text{anomalous}(r) > \theta \quad (3)$$

であれば特異なレビューであるとした。そして、閾値を  $[0, 1]$  の間で動かし ROC 曲線の AUC (area under curve) を求めた。図 5a は、パラメータ  $\beta$  による AUC の比較を示したものであり、 $\beta$  を  $[5, 50]$  の間を 5 刻みで動かして作成した。初期レビューから求めたサマリと長期レビューから求めたサマリの誤差については、我々は誤差が 1.0 以上のものについてのみ考慮する。Amazon.com は五つ星レビューシステムを採用しているため、利用者にとっての差は一つ以上の星が異なる場合であるからである。図 5b は、誤差が 1.0 以上となった書籍数を表したものである。これらの結果を踏まえ、最もレビューサマリの誤差が少ない  $\beta = 40$  を以降の実験で用いることにする。

### 4.3 比較結果

前節で定めたパラメータを用いて、RIH アルゴリズムを Amazon データセットに用いた。図 6 は、各レビューの種類ごとに特異度の分布を表したものである。また、点線は平均特異度を表している。この図によると、ほとんどの一般的なレビューは平均より小さな特異度となっており、独立した特異なレビューと結託する特異なレビューは共に平均より大きな特異度となっていることが分かる。結果的に、我々は特異度を用いて一般的なレビューと特異なレビューを区別することができる。

一方、ONE 及び MRA では多くの特異なレビューが平均以上の特異度となっているが、少なくない一般的なレビューの特異度も平均以上となっている。表 2 は各手法における二種類の特異なレビューそれぞれの検出における AUC を比較したものである<sup>(注1)</sup>。RSD は特異なレビューを十分区別しているとは言いがたく、また AUC も同様に高くない結果となった。

(注1): RIH アルゴリズムのパラメータは、AUC よりも初期レビューからのサマリ計算精度を重視して選んでいる。AUC が最適となるようにパラメータは、 $\alpha = 1.0, \beta = 40$  であり AUC は 0.9392 となる。

次に、初期レビューからのサマリ計算における誤差を比較する。なお、Amazon データセットのレビューは五つ星評価であるが、RIH アルゴリズムの出力は  $[0, 1]$  である。そのため、 $[0, 5]$  区間の値となるように調整を行っている。

図 7 は、初期レビューから求めたサマリと長期レビューから求めたサマリの誤差の分布を記したものである。図からは RIH と MRA は共にべき乗則に近い形となっていることが見て取れる。中でも RIH は MRA よりも小さい平均偏差となっている。MRA では、いくつかの書籍が偏差 2.0 以上となっているが、RIH では偏差 2.0 以上の書籍は MRA と比べて少ない。図 8 は図 7 において、偏差 1.0 以上の部分を拡大したものである。また、合わせて偏差が 1.0 以上となった書籍数も記している。RIH では 4854 冊の書籍が偏差 1.0 以上となり、また偏差 3.5 以上の書籍は無かった。一方、MRA では、約 2 倍の書籍が偏差 1.0 以上であり、いくつかの書籍では偏差 3.5 以上となっていた。この点からいうと、RIH は MRA よりも誤差の少ないサマリが計算できていると言える。最後に、ONE と RSD の結果であるが、これらは共に偏差の極値が 1.0 以上のところにある。そして、2 万冊以上の書籍が偏差 1.0 となっていた。その理由は、この二手法とも特異なレビューを発見する手法でありサマリ計算を主な対象とはしていないからだと考えられる。以上より、この実験では RIH が最も良く初期レビューから長期的なレビューサマリを計算できていた。

## 5. まとめと今後の課題

本稿では、ウェブサイトに掲載されたレビューの不均質性を考慮した、特異なレビュー検出とレビューサマリの計算方法を提案した。Amazon.com に投稿されたレビューのように現実には投稿されたレビューを調べることで、多くのレビューを集める商品がある一方でほとんどレビューが投稿されない商品もあること、特異なレビューは一般的なレビューに比べて極端に少ないことという不均質性を発見した。これらの不均質性を取り扱うため、我々は 論争度 と 偏差希少度 という二つの概念を導入し、反復改良を用いた特異なレビュー発見手法を提案した。また、評価実験により結託するような特異なレビューであっても我々の手法は発見できること及び特異なレビューを含む初期レビューから長期的なレビューサマリを予測できることを示した。

今後の課題としては、まず評価が割れるような商品の取り扱いがあげられる。本稿では、各商品に対するレビューのサマリは一つの値に収束すると仮定しているが、そうではない商品に対する扱いを考える必要がある。また、本アルゴリズムを知っている特異なレビューに対する堅牢性についての議論も必要であると考えられる。最後に、反復改良アルゴリズムの収束速度に関する議論も今後の課題の一つである。

### 文 献

- [1] S. Brin and L. Page. The Anatomy of A Large-scale Hypertextual Web Search Engine. *Computer Networks and ISDN Systems*, 30(1-7):107-117, 1998.
- [2] D. Easley and J. Kleinberg. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, 2010.

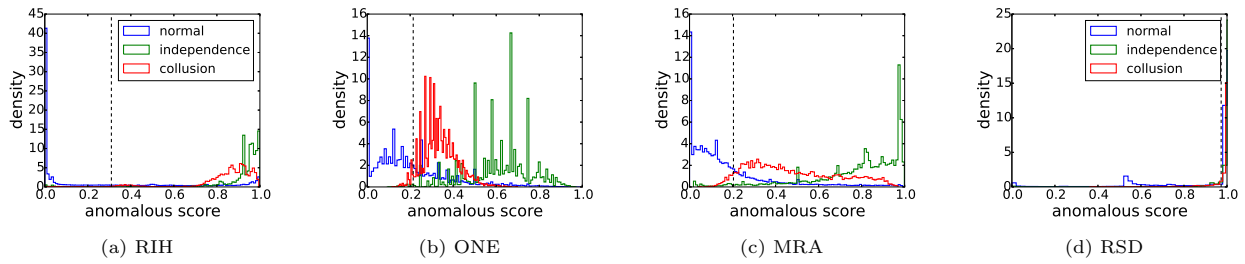


図 6: 特異度の分布.

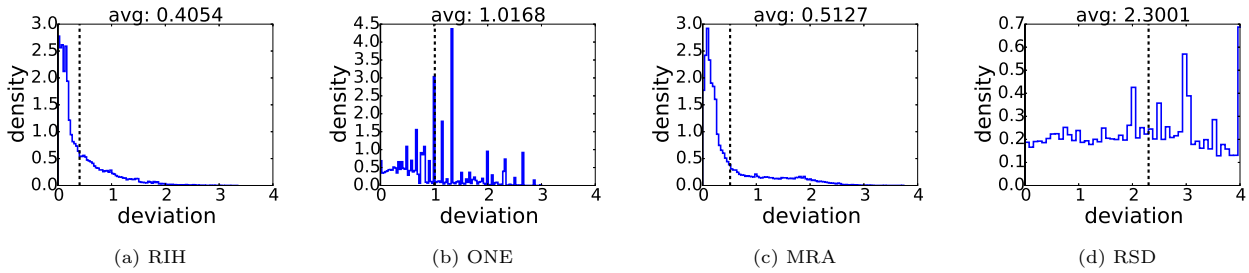


図 7: 正解評価値との偏差の分布.

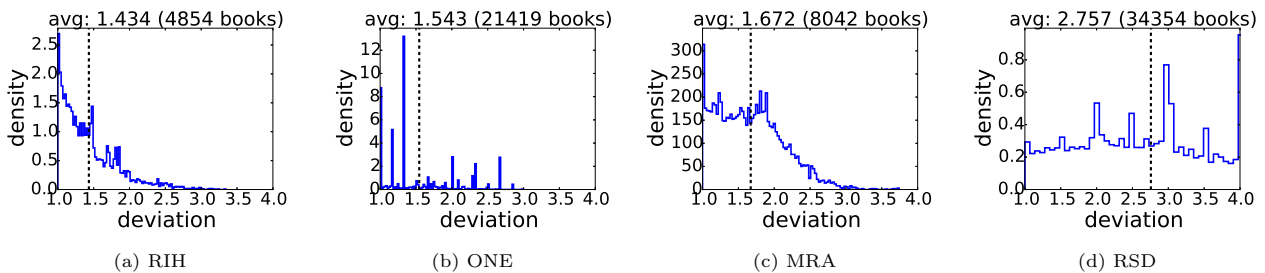


図 8: 正解評価値との偏差の分布. (偏差が 1.0 以上のもの)

- [3] N. Jindal and B. Liu. Opinion spam and analysis. In *Proc. of the 2008 International Conference on Web Search and Data Mining*, pages 219–230, Palo Alto, California, USA, feb 2008. ACM Press.
- [4] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5):604–632, sep 1999.
- [5] H. W. Lauw, E. Lim, and K. Wang. Summarizing review scores of “unequal” reviewers. In *Proceedings of the Seventh SIAM International Conference on Data Mining, April 26–28, 2007, Minneapolis, Minnesota, USA*, pages 539–544, 2007.
- [6] E.-P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw. Detecting Product Review Spammers using Rating Behaviors. In *Proc. of the 19th ACM International Conference on Information and Knowledge Management*, pages 939–948, Toronto, ON, Canada, oct 2010. ACM Press.
- [7] B. Liu, M. Hu, and J. Cheng. Opinion Observer: Analyzing and Comparing Opinions on the Web. In *Proc. of the 14th International Conference on World Wide Web*, pages 342–351, Chiba, Japan, may 2005. ACM Press.
- [8] A. Mukherjee, B. Liu, and N. Glance. Spotting Fake Reviewer Groups in Consumer Reviews. In *Proceedings of the 21st International Conference on World Wide Web*, pages 191–200. ACM, 2012.
- [9] A. Mukherjee, B. Liu, J. Wang, N. S. Glance, and N. Jindal. Detecting group review spam. In *In Proc. of the 20th International Conference on World Wide Web (Companion Volume)*, pages 93–94, 2011.
- [10] L. Page, S. Brin, R. Motwani, and T. Winograd. The PageRank Citation Ranking: Bringing Order to the Web. Technical report, Stanford University, 1998.
- [11] M. S. Pera, R. Qumsiyeh, and Y.-K. Ng. An Unsupervised Sentiment Classifier on Summarized or Full Reviews. In *Proc. of the 11th International Conference on Web Information Systems Engineering*, pages 142–156, Hong Kong, China, dec 2010. Springer-Verlag.
- [12] K. Tawaramoto, J. Kawamoto, Y. Asano, and M. Yoshikawa. A Bipartite Graph Model and Mutually Reinforcing Analysis for Review Sites. In *Proc. of the 22nd International Conference on Database and Expert Systems Applications*, pages 341–348, Toulouse, France, 2011. Springer.
- [13] P. Victor, C. Cornelis, M. De Cock, and A. M. Teredesai. Trust-and-distrust-based recommendations for controversial reviews. *IEEE Intelligent Systems*, 26(1):48–55, 2011.
- [14] G. Wang, S. Xie, B. Liu, and P. S. Yu. Review Graph Based Online Store Review Spammer Detection. In *Proc. of the 11th IEEE International Conference on Data Mining*, pages 1242–1247, Vancouver, BC, Canada, dec 2011. IEEE Computer Society.
- [15] 川本 淳平, 俵本 一輝, 浅野 泰仁, 吉川 正俊. 初期レビューを用いた長期間評価推定. In 第 7 回データ工学と情報マネジメントに関するフォーラム, pages D3–6, 2015.