

初期レビューを用いた長期間評価推定

川本 淳平[†] 俵本 一輝^{††} 浅野 泰仁^{††} 吉川 正俊^{††}

[†]九州大学大学院システム情報科学研究院 〒819-0395 福岡市西区元岡 744

^{††}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: [†]kawamoto@inf.kyushu-u.ac.jp, ^{††}tawara@db.soc.i.kyoto-u.ac.jp,

^{†††}{asano,yoshikawa}@i.kyoto-u.ac.jp

あらまし レビューサイトの評価は消費者にとって有益である。しかし、商品やサービスの提供開始すぐの段階では限られた数のレビューしか投稿されず、これらの初期レビューは長期的な評価とは異なっていることがある。特に、レビュー数が少ない場合、悪意あるレビュアーなどによる特異なレビューの影響を受けやすい。本研究では、初期レビューから長期的な評価を予測する手法を提案する。

キーワード レビューサイト, 早期レビュー, 予測

1. はじめに

多くのウェブサイトが種々の対象物、例えば商品や商店、サービスに関する人々の評価（レビュー）を掲載している。これらの評価は、顧客や販売者の意思決定に対して重要な影響を与えていると考えられている。例えば、顧客はいくつかの商品に関して Amazon.com に投稿された評価を読み最も優れた商品を購入しようとする。また、“The Internet Movie Database (IMDb)”^(注1) のような映画に関する評価を集めたサイトを読み、次に視聴する映画を決定するだろう。一方で、販売者はこうした人々の評価からフィードバックを得てマーケティングに反映するだろう。

ある対象物が公開されてから短い期間に投稿された評価、すなわち初期レビューは一般的に限定された数の評価のみからなる。初期レビューは、顧客や販売者の意思決定について重要な影響を与えるが、十分時間が経過するまでに投稿された評価の傾向とは異なる場合がある。評価数が少ないため悪意のある評価結果の影響を受けやすく、またステルスマーケティングのように意図的な評価結果操作の対象となり得る。Social spammers や Crowd turbing workers のように、複数の評価者が結託し定期的に特別な目的を持って評価を行う例も報告されている。その他にも、一般的な人々とは異なる特別な主張を持つ評価者による評価の影響も大きい。こうした悪意は無いものの特異な評価である例の一つとして、アーリーアダプターかつエキスパートの評価があるだろう。彼らの評価は一般的な人々が求めている評価とはしばしば異なる。例えば、日常生活で使用する新しいデジタルカメラを探している通常の人にとって、高価で多機能なカメラを勧める評価は求めているものと異なるであろう。対象物の公開から時間が経ち、十分な量の評価が集まれば、こうした特異な評価の影響は小さくなる。したがって、特異な評価は初期レビューに対して特に問題である。また、Amazon.com のような多くのサイトでは他の利用者からの連

絡を受けてスパムのような特異な評価を削除している。しかし、これも時間を必要とする。そのため、この努力も最新の商品に対するスパムには効果がないことも多い。それゆえ、そうした評価から一般的な人々に対して適正な評価結果を求める方法が求められている。

本稿では、初期レビューから長期間経過後の評価を推測する方法を提案する。提案手法は、仮に特異な評価が含まれていても推測することができる。初期レビューからだけでは長期間経過後の評価を求めるに十分な情報を得ることは難しい、そこで各評価者の他の評価対象に対する評価を再帰的に利用する。評価値の再帰的な計算は、二部グラフにおける *Repeated improvement* [1] として知られる手法であり、我々は評価者と評価対象を頂点とする二部グラフを用いて利用する。多くのスパムレビュアー検知手法 [2-5] は十分な学習用データを利用するが、我々の提案手法は学習データを必要としない。レビューの解析という観点では類似の二部グラフや三部グラフを用いた手法が提案されている [6-8]。Lauw ら [6] によるモデルは、学術論文の査読プロセスに焦点を当てており、すべての評価者が正当に振舞うものと仮定されている。Wang ら [8] によるモデルは、限られた期間内に行われた商店の評価分析を対象としている。俵本ら [7] は、レビューサイトにおいて特異な評価者を発見し評価値の要約を計算する手法を提案している。この手法は比較的少量のデータからでも特異な評価者を発見することができる。このモデルを基に、本稿ではより適正なモデルを提案する。特に、現実のレビューサイトにおいて十分な時間経過後の評価を予測するために評価値のばらつきを考慮する二つの新しい方法を提案する。一つは評価結果の信頼度であり、もう一つは特異度の重み付けである。

本稿では、Amazon.com による書籍評価データを実データとして用い提案手法による長期間経過後評価の予測値を評価し、特異な評価者の検出結果を評価する。また、提案手法は既存研究 [7] による結果に比べて良いことを示す。特に、我々の提案手法は特異な評価者が結託している場合において他の手法に比べてそうした評価者を良く発見できることを示す。

(注1): <http://www.imdb.com/>

本稿の構成は次の通りである．まず，2. 節で関連研究について紹介する．そして，3. 節にて本稿で利用する二部グラフモデルを導入する．4. 節では，Repeated improvement について紹介し提案の初期レビューから長期間経過後の評価値を推測する提案アルゴリズムを説明する．5. 節では実データを用いた評価実験の概要とその結果を説明する．最後の 6. 節では，まとめと今後の課題について述べる．

2. 関連研究

我々の提案手法は，*Repeated improvement* の原理 [6–9] を利用している．Lauw ら [6] は学术论文の査読プロセスをモデル化しており査読者の甘さを計算している．甘い査読者は他の査読者に比べて高い点数を各論文に与えがちであり，厳しい査読者は他の査読者に比べて低い点数を与えがちである．この研究ではすべての査読者は正当に振舞うことを仮定している．つまり，査読者は特定の論文について意図的に高いまたは低い点数を付けることはないとは仮定している．一方，我々が対象としているレビューサイトでは，特異な評価者はこの仮定を満足しない．例えば競合他者については通常とは異なる振る舞いをすることはあり得る．したがって，彼らの成果をそのまま我々の問題に適用することは難しい．

依本ら [7] と Wang ら [8] は独立に二部グラフを用いてレビューサイトにおける評価者と評価対象の関係をモデル化する手法を提案した．この二つの二部グラフモデルでは，評価者を表す頂点集合と評価対象を表す頂点集合の二種類の頂点集合があり，評価者と評価対象の間に評価を表す枝が張られる．なお，Wang らは彼らのモデルを評価者，評価そして評価対象という三つの異なる頂点集合からなる三部グラフであるとしているが本質的に差はない．彼らのモデルでは，評価者の信頼性と評価値の公正さ，評価対象の信頼性という再帰的かつ相互に関係する三つの指標を用いている．そして，

- (1) 信頼性の高い評価者は公正な評価を行う．
 - (2) 公正な評価は信頼性の高い評価者からの評価と類似している．
 - (3) 公正さの計算は高い信頼性または低い信頼性の評価対象の計算に際して重要となる．
 - (4) 多くの信頼出来る評価者が良いと判断した評価対象は信頼性が高く，悪いと判断した評価対象の信頼性は低い．
- ということを仮定している．仮定 (2) を実現するために，彼らの手法は “agreement score” を導入している．似た評価が多い場合この agreement score は高くなる．したがって，この手法では結託するスパマーへの耐性は低いと思われる．依本ら [7] の手法は Repeated improvement を用い，特異な評価者の検出と評価値の要約を行う．この手法では，長期的な評価値の推定は行っておらず，実データを用いた評価も行われていない．

特異値あるいは外れ値の検出は様々な分野で長く研究されている．多くの場合では，グラフ上 [10–12] での特異値検出は考えられていない．これらは文字列解析における特異値検出と組み合わせることはできるが，二部グラフモデルへの適用は難しい．いくつかのグラフにおける特異値検出 [13–15] は提案さ

れている．Noble と Cook [13] は，グラフにおける特異な部分グラフの検出方法を提案している．また，Chakrabarti [14] は特異な枝の検出方法を提案している．これらの二つの手法は残念ながら特異な頂点の検出は行えない．そのため，特異な評価者を検出するという我々の問題に直接適用することは難しい．Sun ら [15] は，二部グラフを用いて特異な頂点を検出する方法を提案している．彼らの手法は，始めにランダムウォークを用いて似ている頂点を探索し，その後特異な頂点を求める．彼らは，この二部グラフモデルはネットワークにおいて著者と論文，著者と会議，役者と映画などのように二つの事柄の関係を表現し特異なものを検出することに適していると主張している．しかしながら，彼らの手法はオンラインにおけるレビューサイトを対象とはしておらず，関係間の強度即ち枝に付加された値を考慮していないため，評価値が重要となる我々の問題へそのまま適用することは難しい．

いくつかの研究では，スパムレビューやスパム評価者の検出を試みている [2–5]．スパム評価者は常に特異な評価者であるが，その逆は常には成り立たない．特異な評価者は他の人々とは異なる評価を行うエキスパートなども含む．それゆえ，特異な評価者の検出問題はより一般的な問題である．そして，一般的な人々のための長期的な評価の予測においてはより利便性が高い．

3. 二部グラフモデル

本節では，評価者と評価値の両方を同時に扱うための二部グラフモデルを導入する．二部グラフモデルは，本研究のように評価者，評価値そして評価対象の関係を表すことに適したデータ構造である．我々の二部グラフでは，評価者と評価対象という二種類の頂点を用い，二頂点間の枝に評価値を表すラベルを付加することにする．

本稿では，評価者を表す頂点の集合を $V_R = \{p_1, p_2, \dots, p_{|V_R|}\}$ で表し，評価対象を表す頂点の集合を $V_O = \{q_1, q_2, \dots, q_{|V_O|}\}$ で表すことにする．また，枝は二種類の頂点間のみ張られるものとし，枝の集合を $E = \{(p, q) | p \in V_R, q \in V_O\}$ と表す．なお，評価者は一つの評価対象に対して最大で一つの評価値のみを与えるものとする．評価者 p が評価対象 q に評価値を与えた場合 p と q は枝 (p, q) にて結ばれることになる．そして，評価値を表すためにそれぞれの枝はラベルを持つ．本稿では，評価値は実数値であるとし， $e(p, q)$ で評価者 p による評価対象 q の評価値を表すことにする．

次に，評価者の特異度を導入する．本稿では，特異度は実数値であり評価者 $p \in V_R$ の特異度を $a_p \in \mathbf{R}$ で表す．また，評価対象に対して評価の要約を導入する．評価の要約は，それぞれの評価対象に与えられたすべての評価値から算出される値であり，評価対象 $q \in V_O$ に対する評価の要約を s_q で表す．

以上を元に，本稿で用いる二部グラフを次のように定義する．

定義 3.1 V_R 及び V_O をそれぞれ評価者並びに評価対象の表す頂点集合とし， E を頂点集合とする．ま

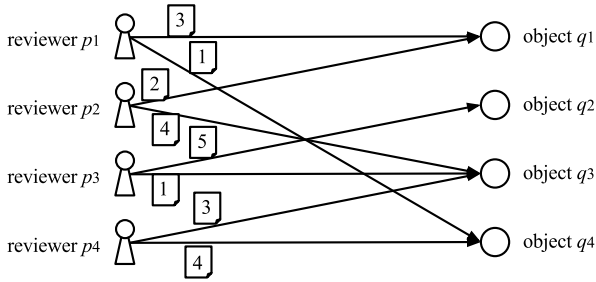


図 1: 二部グラフの例

た, e を頂点に付加されたラベルを取得する関数とする. この時, 本稿で用いる二部グラフは $G = (V_R, V_O, E, e, \{a_{p_1}, a_{p_2}, \dots, a_{p_{|V_R|}}\}, \{s_{q_1}, s_{q_2}, \dots, s_{q_{|V_O|}}\})$ と書ける.

例 3.1 図 1 は 4 人の評価者 (reviewer) と 4 つの評価対象 (object) からなる二部グラフの例を示したものである. この例では, 例えば評価者 p_1 は評価対象 q_1 と q_4 に対して, それぞれ 3 と 1 という評価値を与えている. それゆえ, 評価者 p_1 は q_1 と q_4 それぞれ結びつけられておりそのラベルは $e(p_1, q_1) = 3$ 及び $e(p_1, q_4) = 1$ となっている.

次節では, 特異度及び評価の要約を計算する方法について議論する.

4. Repeated Improvement Analysis

本稿では, 前節で定義した二部グラフ上で動作する *Repeated improvement analysis (RIA)* と名付けたアルゴリズムを導入する. 一般的に, 初期レビューすなわち評価数が少ない段階において, 特異な評価者の発見や特異な評価値を除いた適切な要約を求めることは難しい. 我々は, それぞれの評価者による過去の評価の影響や多くの評価を集めた評価対象の影響を二部グラフを用いて伝播させる. そうして, 特異な評価者や適切な要約を計算する. この影響の伝播を行うために, 我々は評価者と評価対象の間で *Repeated improvement* の原則 [1] を利用する.

この原則を利用するために, 特異度と評価の要約の間に次のことを仮定する.

(1) 特異な評価者は個々の評価対象についてその評価の要約とは異なる評価値を与えるだろう.

(2) 評価の要約は特異でない通常の評価者の評価値と近い値であるべきである.

この仮定を元に, 評価者 p が評価対象 q に与えた評価値の特異度 a_{pq} を求め, 評価者 p が評価した対象すべてに対する特異度を集約することで評価者 p の特異度 a_p を求める. 評価値に対する特異度 a_{pq} は評価値 $e(p, q)$ と評価の要約 s_q を用いて求める. 一方, 評価対象 q に対する評価の要約は, この対象に対するすべての評価値を評価者の特異度による重みをつけて集約することで求める.

まとめると, 特異度は評価の要約を用いて計算され, 評価の要約は特異度を用いて計算される. したがって, これらの計算

を繰り返し行い最終的な値を求めることになる. 言い換えれば, *Repeated improvement* の原則である. この原則は, どの評価者が特異なのか一度の計算で判断できない本問題のような場合にうまく機能する.

繰り返しの計算によって, 二部グラフにおけるある頂点の影響を他の頂点へ広く伝播させることができる. そのため, 少しの評価対象しか評価していないような評価者の特異度や少しの評価しか集めていない評価対象に対する要約を, それ以外の評価者及び評価対象に関する評価値と比較することで求めることができる.

4.1 特異度の計算

始めに評価者 p による評価対象 q への評価値 $e(p, q)$ の特異度 a_{pq} の計算方法について述べる. 本稿では, この評価値の特異度を単純に評価値 $e(p, q)$ と要約 s_q との類似度によって測る.

定義 4.1 評価者 p による評価対象 q への評価値 $e(p, q)$ の特異度 a_{pq} は, 評価対象 q への評価の要約 s_q との差

$$a_{pq} = |e(p, q) - s_q|. \quad (1)$$

とする.

それゆえ, 小さい類似度であるほど高い特異度となる.

また, 信頼できる評価の要約は特異度の計算に対してより強い影響を及ぼすべきである. そこで, 評価の要約に対して信頼度を導入する. この信頼度は評価対象に対する評価の総数とその分散を用いて定義する. 信頼度の定義に進む前に, 次の二つの前提を確認する.

(1) 多くの評価値から計算された要約は, 少ない評価値から計算された要約よりも信頼できる.

(2) ばらつきの大きい評価値から計算された要約は, ばらつきの小さい評価値から計算された要約よりも信頼できない. これらを基にすると, 二部グラフにおいて次数の高い評価対象及び, 評価値の分散が小さい評価対象の信頼度は高くすべきである. 一方で, ある評価対象が得た評価は十分に大きな評価からサンプリングされたものであると考えることができる. 実際, 早期レビューは長期間に行われるレビューの一部である. そこで, 分散の計算には不変分散を用いる.

定義 4.2 評価対象 q の要約 s_q に対する信頼度 c_q は,

$$c_q = \begin{cases} 0.5 & \text{if } N_q = 1, \\ \frac{1}{\sigma_q^2 + 1} \times \log(N_q) & \text{otherwise} \end{cases} \quad (2)$$

である. ただし, N_q は評価対象 q への評価数を表し, σ_q^2 は下記により求まる不変分散である.

$$\sigma_q^2 = \frac{1}{N_q - 1} \sum_{p:(p,q) \in E} \left(e(p, q) - \sum_{p:(p,q) \in E} \frac{e(p, q)}{N_q} \right)^2$$

最後に, これらを合わせて評価者の特異度を定義する.

定義 4.3 評価者 p の特異度 a_p を

$$a_p = \sum_{q:(p,q) \in E} \frac{c_q}{\sum_{q:(p,q) \in E} c_q} a_{pq}. \quad (3)$$

と定める.

4.2 評価値の要約

評価値の要約を考える場合, 平均値は妥当な選択である. しかし, 我々は個々の評価値を与える評価者について特異度を考えている. 特異な評価者による評価値は要約結果を歪ませることに繋がる. そのため, 我々は特異度とシグモイド関数を用いて求めた重みを用いた重み付き平均を採用することにした. この評価者 p に対する重み $w_\alpha(p)$ は次のようになる.

$$w_\alpha(p) = \frac{1}{1 + \exp(\alpha \frac{a_p - \mu_a}{\sigma_a})} \quad (4)$$

ここで,

$$\mu_a = \sum_{p \in V_R} \frac{a_p}{|V_R|}, \quad \sigma_a^2 = \sum_{p \in V_R} \frac{(a_p - \mu_a)^2}{|V_R|}$$

である. 即ち, μ_a は特異度の平均値であり σ_a^2 は特異度の分散である. α はゲインと呼ぶパラメータである. 小さな α は, ほとんどの重みを 0.5 周辺に集める. 言い換えれば, 要約の計算における特異度の影響を下げることを意味する. 逆に, 大きな α は重みを広い範囲に分布させ要約の計算において特異度をより考慮することを意味する. その結果, 重みが 1 に近い一部の評価値のみを要約計算に用いることになる. ゲイン α による特異度及び要約計算への影響は 5. 節でも議論する.

最後に, 評価値の要約を定義する.

定義 4.4 評価対象 q に対する評価の要約 s_q を

$$s_q = \sum_{p:(p,q) \in E} \frac{w_\alpha(p)}{\sum_{p':(p',q) \in E} w_\alpha(p')} e(p, q). \quad (5)$$

と定める.

4.3 アルゴリズム

アルゴリズム 1 は RIA の手続きを示している. このアルゴリズムは 3. 節にて定義した二部グラフ G と 4.2 節にて定義したゲインを受け取る. 始め, 各評価者 p の特異度 a_p は $1/N_p$ と初期化される. ここで, N_p は評価者 p が行った評価数である. すなわち, 多くの評価を行っている評価者の特異度は小さくなる.

アルゴリズムの前半では, それぞれの評価対象 q に対して要約 s_q の信頼度 c_q を計算する. この初期化処理の後, 特異度 a_p と要約 s_q を更新幅が無視できるようになるまで繰り返し計算する. 繰り返しが終わると各評価者に対する特異度と各評価対象に対する要約を出力する.

4.4 様々な評価形式に対する拡張

ここまで, 評価値は実数の値をとるものとして議論してきた. 一方, 現実のレビューサイトを見ると種々の評価形式が利用されている. 例えば, 複数の軸に対して正・負の値をつける場合

Algorithm 1 Repeated Improvement Analysis

Require: 二部グラフ $G = (V_R, V_O, E, e, \{a_{p_1} a_{p_2}, \dots, a_{p_{|V_R|}}\}, \{s_{q_1}, s_{q_2}, \dots, s_{q_{|V_O|}}\})$

Require: ゲイン α

/* 初期化 */

$1 \leq i \leq |V_R|$ に対して $a_i = 1/N_i$

for each j in $1 \leq j \leq |V_O|$ **do**

信頼度 c_j を式 (2) により計算

end for

repeat

/* 要約 s_j の更新 */

重み $w_\alpha(a_p)$ を式 (4) により計算

for each j in $1 \leq j \leq |V_O|$ **do**

要約 s_j を式 (5) により更新

end for

/* 特異度 a_i の更新 */

for each i in $1 \leq i \leq |V_S|$ **do**

特異度 a_{ij} を式 (1) により求め a_i を式 (3) により更新

end for

until a_i と s_j の更新幅が無視できるようになるまで繰り返す

return $\{a_{p_1}, a_{p_2}, \dots, a_{p_{|V_R|}}\}, \{s_{q_1}, s_{q_2}, \dots, s_{q_{|V_O|}}\}$

表 1: データセットに含まれるレビュー数及び評価者数

2004 年 12 月 31 日までのレビュー数	1555315
2005 年 1 月 1 日以降のレビュー数	613265
2004 年 12 月 31 日までの評価者数	730667

もある. この場合, 実数値ではなくベクトルを用いて表現する方が都合が良い場合もある.

我々の提案手法は実数による評価値のみを対象としたものではない. 言い換えれば, ここまで議論してきた提案手法は容易に一般化することができる. 我々は評価値の特異度を a_{pq} を求めているが, この値を式 (1) に示すように評価者 p による評価値 $e(p, q)$ と評価対象 q の要約 s_q との類似度で評価している. したがって, 評価値と要約との距離

$$a_{pq} = \text{distance}(e(p, q), s_q)$$

が定義さえできれば, 提案手法を他の評価形式に適用できる. 例えば, 評価値がベクトルで表現される場合 distance としてユークリッド距離の利用が考えられる. また, 評価値が確率分布として与えられる場合 distance として KL ダイバージェンスを利用することもできる.

5. 評価実験

5.1 データセット

提案手法による解析結果を評価するために実験を行った. 実験には Amazon.com におけるレビューデータ [3] のうち書籍に関するレビューデータを用いた. このデータセットには, 1996 年 5 月 31 日から 2006 年 5 月 29 日までのレビューが含まれている. また, レビューは 1 から 5 の五段階の整数によって与えられている. 本実験では, 2004 年 12 月 31 日までのレビューデータを用いて, レビュー数の少ない書籍に対する 2006 年 5

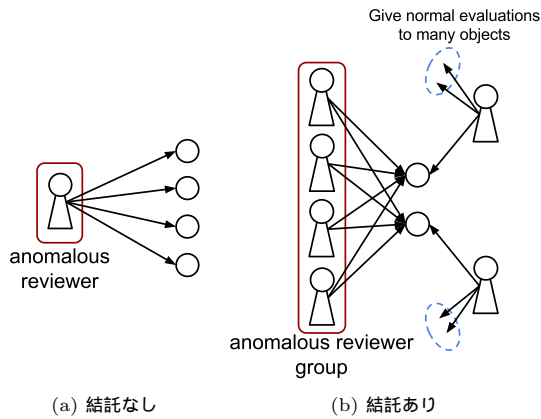


図 2: 特異な評価者の種類

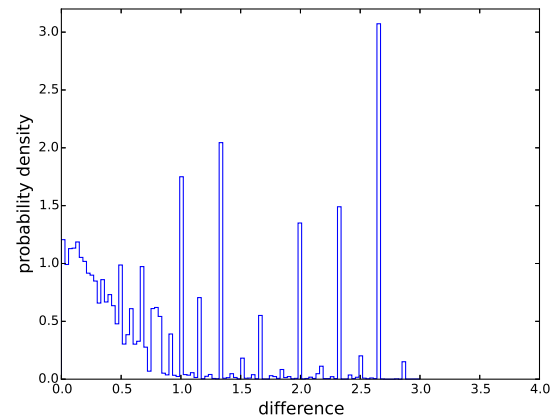


図 3: 単純平均との比較

月 31 日時点での評価結果を推測する課題を設定した。すなわち、個々の評価対象に対する正解評価値として 2006 年 5 月 31 日までに与えられたレビューの単純平均を設定し、2004 年 12 月 31 日までのレビューデータを用いてこれを推測する。データセットに含まれるレビュー数及び評価者数を表 1 に示す。

本実験では、図 2 に示すように二種類の特異な評価者を考える。一種類目の特異な評価者は図 2a のように結託を行わず個別に特異な評価を与える評価者である。図は、4 つの評価対象を評価している特異な評価者の例を示したものである。もう一種類目の特異な評価者は図 2b のように複数人で結託してある評価対象に特異な評価を与える評価者である。図の例では、二人の評価者のみが評価を行っている評価対象に対して、四人の特異な評価者が結託して特異な評価を行う場合を示している。このように、通常の評価者よりも多くの人数で結託し特異な評価を行う場合、正しい評価値を求めることは難しい。しかし、Social spammers や Crowd turfing workers の発見には、このような場合に対処することが求められる。

一般的に、多くの商業サイトではスパムレビューは管理者によって削除される。我々が利用する Amazon.com のデータセットでもスパムレビューの多くは削除済みである可能性が高い。そこで、データセットに含まれる特異な評価者は少ないと仮定し、削除される前の特異な評価者を想定して人工的な評価者を追加する。本実験では、結託しない特異な評価者を全評価者数の約 2.5% である 18300 人追加し、また結託する特異な評価者も同数すなわち 18300 人追加する。

結託しない特異な評価者による評価は、次のように生成した。まず、Amazon.com データセットに含まれる個々の評価者がいくつもの評価を行っているか調べ、評価数の分布を作成する。そして特異な評価者一人が行う評価数をこの分布に従う乱数により決定する。個々の特異な評価者は、この決められた数分だけランダムに選択した評価対象に対して特異な評価を与える。特異な評価は、2004 年 12 月 31 日時点での特異な評価を除いた平均評価値が 3.5 未満であれば 5.0 を、3.5 以上であれば 1.0 とした。また、結託する特異な評価者による評価は、次のように生成した。結託評価は、評価数が 2 である二つの書籍に 4 人の特異な評価者グループが行うこととした。全部で 18300 人

の結託する特異な評価者を追加するため、全部で 4575 組のグループが作成されることになる。結託評価者が結託して与える評価値は結託しない特異な評価者による評価値と同様にした。すなわち、2004 年 12 月 31 日時点での特異な評価を除いた平均評価値が 3.5 未満であれば 5.0 を、3.5 以上であれば 1.0 とした。なお、すべての特異な評価者は 2004 年 12 月 31 日より過去の時点で評価を行うものとした。

図 3 は、特異な評価者が評価を行った評価対象について、2004 年 12 月 31 日までに与えられた評価の単純平均と 2006 年 5 月 29 日時点での平均評価値との誤差の分布を求めたものである。分布図は総面積が 1.0 となるように正規化している。なお、対象となるデータセットは 5 段階評価を採用しているため、誤差の最大は 4.0 である。図によれば、誤差 2.0 以上の評価対象が一定数存在することが見て取れる。5 段階評価において 2.0 程度の差があれば、現実的に購買意欲に大きな影響を与えと言え。これらの改善が本実験での課題である。

最後に、本実験での比較手法について説明する。本実験では、提案手法（以降では RIA と呼ぶ）に加えて次の二種類を比較手法とした。一つ目は、Repeated improvement を用いずに特異度を求める手法である [4]。以降では、本手法を ONE と呼ぶ。二つ目は、Repeated improvement は用いるが、評価値の信頼度や特異度の重みを考慮しない手法である [7]。以降では、本手法を MRA と呼ぶことにする。また、提案手法のパラメータ α については、0.25, 0.5, 1.0 及び 2.0 の 4 種類を設定した。

5.2 特異な評価者の検出

始めに、特異な評価者の検出結果について説明する。すなわち、それぞれの手法で求めた評価者の特異度が、通常の評価者については小さく、特異な評価者については大きな値となっているかについて調べる。図 4 は、各手法によって求まる特異度の分布を評価者の種類別にプロットしたものである。なお、特異度は 0.0 から 1.0 の間にスケールしてあり、分布図はそれぞれ総面積が 1.0 となるように正規化してある。特異な評価者の種類は 5.1 節で述べたように、通常の評価者 (normal)、結託しない特異な評価者 (independence)、そして結託する特異な評価者 (collusion) の三種類である。二種類の特異な評価者に

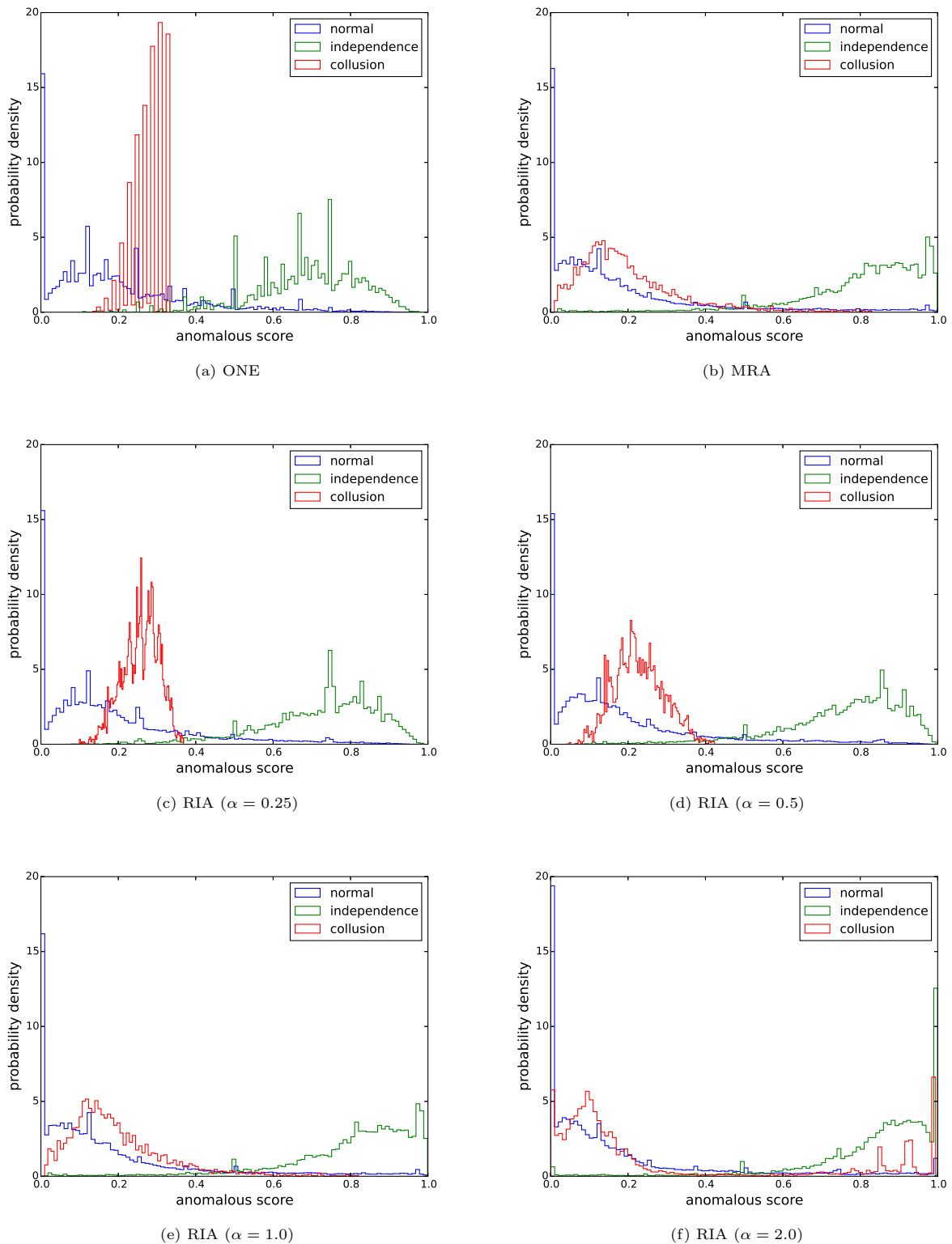
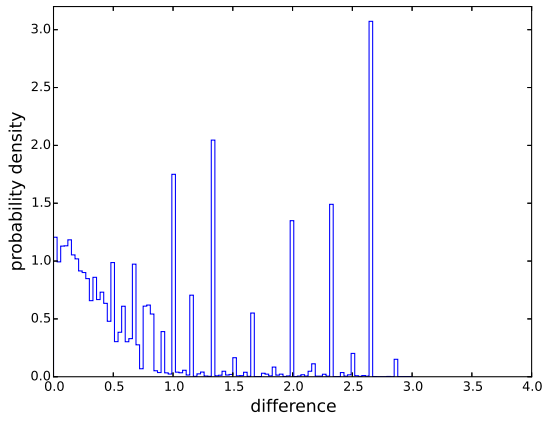


図 4: 評価者別特異度の分布 (normal: 通常の評価者, independence: 結託しない特異な評価者, collusion: 結託する特異な評価者)

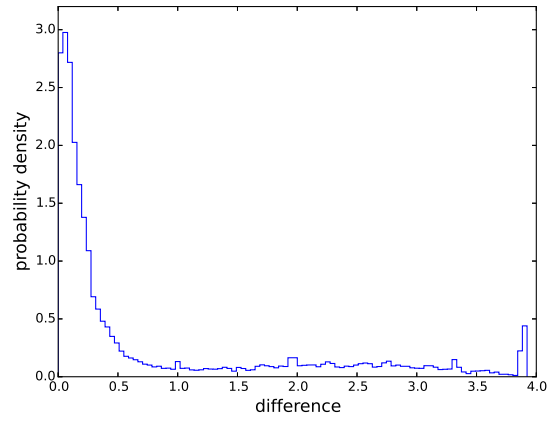
対して、特異度が大きくなっている方が良い結果である。

結託しない評価者 (independence) について、特異度の分布はどの手法でも 0.5 以上に集まっている。従って、どの手法でも結託しない評価者については特異な評価者であると検出できている。中でも、MRA と RIA ($\alpha = 1.0, 2.0$) では特異度の分布がより大きな値となっている。提案手法においては、ゲイ

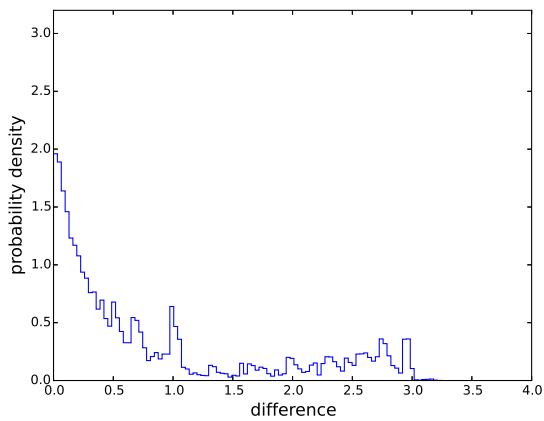
ン α は特異度の影響をコントロールするパラメータであり、大きな値を設定すればより影響を高めることを意味する。結託しない評価者においては、他の評価者と比較することで特異な評価者であると判別しやすい。そのため、ゲインを大きくし影響を高めることで、特異な評価者はより大きな特異度を得るという効果が得られていると考えられる。



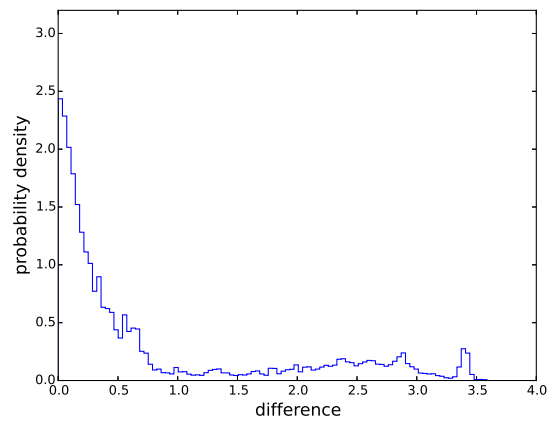
(a) ONE



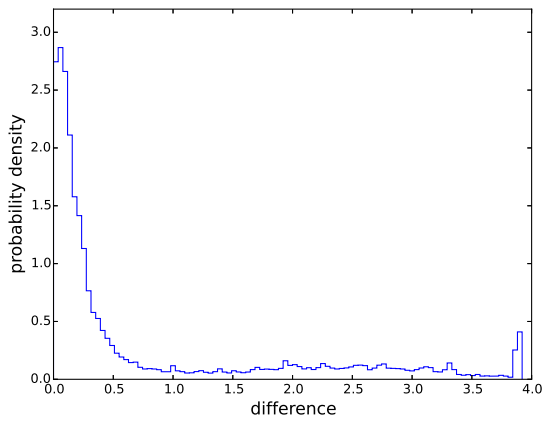
(b) MRA



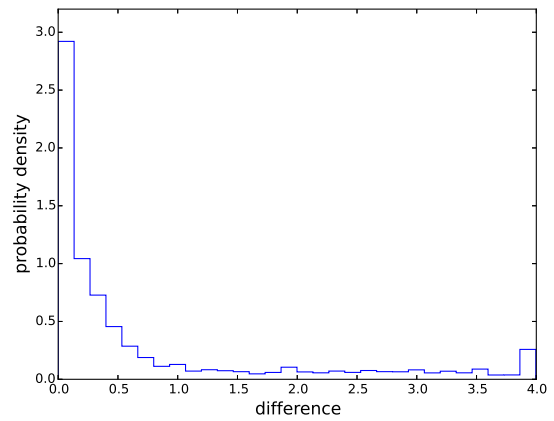
(c) RIA ($\alpha = 0.25$)



(d) RIA ($\alpha = 0.5$)



(e) RIA ($\alpha = 1.0$)



(f) RIA ($\alpha = 2.0$)

図 5: 予測誤差の分布

一方、一般的に結託する評価者の発見は難しく、検出結果もそのことを示している。特に、今回追加した結託する特異な評価者は、各書籍について見れば、特異な評価者グループの方が多数となるように追加しているため通常的评价者の方が特異な評価者となる。そのため、どの手法でも結託する評価者の特異度は十分大きな値とはならなかった。これらの中で最も良い結

果と言えるのは、図 4c に示す $\alpha = 0.25$ の提案手法である。なぜなら、この場合が最も通常的评价者と特異な評価者の分布において重なっている部分が少ないためである。通常的评价者と特異な評価者の特異度の上に境界線を設定できれば良いため、 $\alpha = 0.25$ の提案手法が最もエラーの少ない結果と言える。実際、後で述べる評価値の予測においても $\alpha = 0.25$ の提案手法

が最も良い結果となっている。

提案手法におけるゲイン α による影響をまとめる。図 4c, 4d, 4e, 4f によると, α が大きくなれば一部の特異な評価者に対する特異度はより大きく判定される一方, 一部の特異な評価者に対する特異度はより小さく判定される傾向にある。これは, 大きなゲインによって式 (4) の重みを広い範囲に分布させている効果が出ていると考えられる。一方でゲイン値を大きくしすぎると, 重みの値が極端に大きいまたは小さい場合に二極化してしまう。そのため, 同じ重みを持つ評価者が増え評価者ごとの差異を扱えなくなる。その結果, 特異な評価者が決定できない曖昧な集団が形成されることになる。

5.3 評価値の予測

次に, 長期間経過後評価の予測結果について説明する。図 5 は比較手法及び提案手法を用いて予測を行なった場合の結果を示したものである。この図は, 図 3 同様に横軸に長期間平均との誤差をとった分布図であり, 総面積が 1.0 のなるように正規化してある。

Repeated improvement を用いない図 5a の ONE は, 図 3 の単純平均とあまり変わらない結果となった。図 5b, 5e 及び 5f の MRA と RIA ($\alpha = 1.0, 2.0$) は予測誤差が 1.0 以下が大半となっている。その一方で, 予測誤差が 3.0 以上の場合も発生している。これは, 結託する特異な評価者を通常の評価者と後検出し, また通常の評価者を逆に特異な評価者と判断してしまった結果によるものである。図 5c に示す, 特異な評価者の検出結果が最も良かった $\alpha = 0.25$ の提案手法では, 予測誤差 1.0 前後の書籍が増える一方で, 予測誤差が 3.0 以上となる場合はほとんど発生していない。そのため, 全体として $\alpha = 0.25$ の提案手法が評価値の予測でも最も良い結果であると考えられる。

6. まとめと今後の課題

本稿では, 二部グラフを用いて評価者と評価対象の関係をモデル化し, Repeated improvement によって特異な評価者を検出し, 個々の評価対象に対して評価の要約を求める手法を提案した。提案手法は Repeated improvement を用いているため, 評価者と評価対象それぞれの影響をグラフ上に伝播させる。この影響の伝播によって, ある評価対象に対する評価が少ない, 初期レビュー段階であっても, その他の評価対象に向けられた評価値を用いて, 特異な評価者の検出と要約を求めることができる。また, 提案手法は実数値の評価値のみだけでなく, 差が定義できる様々な指標に対して適用することが可能である。

今後の課題として, 幾つかの問題について検討する必要がある。一つは, 評価対象の評価が分かれた場合の取り扱いである。現在, 各評価対象の評価はある値に収束することを仮定している。しかし, 一般的には評価が分かれるような対象も存在するため, このような評価の取り扱いが課題である。また, 長い期間通常の評価者のように振る舞い, あるタイミングで特異な評価者となるような悪意ある評価者の取り扱いも課題である。

文 献

[1] Easley, D., Kleinberg, J.: Networks, Crowds, and Markets: Reasoning About a Highly Connected World. Cambridge

University Press (2010)

[2] Liu, J., Cao, Y., Lin, C., Huang, Y., Zhou, M.: Low-Quality Product Review Detection in Opinion Summarization. In: Proc. of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Prague, Czech Republic, Association for Computational Linguistics (2007) 334–342

[3] Jindal, N., Liu, B.: Opinion spam and analysis. In: Proc. of the 2008 International Conference on Web Search and Data Mining, Palo Alto, California, USA, ACM Press (February 2008) 219–230

[4] Lim, E.P., Nguyen, V.A., Jindal, N., Liu, B., Lauw, H.W.: Detecting Product Review Spammers using Rating Behaviors. In: Proc. of the 19th ACM International Conference on Information and Knowledge Management, Toronto, ON, Canada, ACM Press (October 2010) 939–948

[5] Mukherjee, A., Liu, B., Wang, J., Glance, N.S., Jindal, N.: Detecting group review spam. In: In Proc. of the 20th International Conference on World Wide Web (Companion Volume). (2011) 93–94

[6] Lauw, H.W., Lim, E., Wang, K.: Summarizing review scores of "unequal" reviewers. In: Proceedings of the Seventh SIAM International Conference on Data Mining, April 26–28, 2007, Minneapolis, Minnesota, USA. (2007) 539–544

[7] Tawaramoto, K., Kawamoto, J., Asano, Y., Yoshikawa, M.: A Bipartite Graph Model and Mutually Reinforcing Analysis for Review Sites. In: Proc. of the 22nd International Conference on Database and Expert Systems Applications, Toulouse, France, Springer (2011) 341–348

[8] Wang, G., Xie, S., Liu, B., Yu, P.S.: Review Graph Based Online Store Review Spammer Detection. In: Proc. of the 11th IEEE International Conference on Data Mining, Vancouver, BC, Canada, IEEE Computer Society (December 2011) 1242–1247

[9] Easley, D., Kleinberg, J.: Networks, Crowds, and Markets: Reasoning About a Highly Connected World. Cambridge University Press

[10] Aggarwal, C.C., Yu, P.S.: Outlier detection for high dimensional data. In: In Proc. of the 2001 ACM SIGMOD International Conference on Management of Data. (2001) 37–46

[11] Chandola, V., Banerjee, A., Kumar, V.: Anomaly Detection: A Survey. ACM Computing Surveys **41**(3) (July 2009) 1–58

[12] Wang, X., Davidson, I.: Discovering Contexts and Contextual Outliers Using Random Walks in Graphs. In: Proc. of the Ninth IEEE International Conference on Data Mining, Miami, FL, USA, IEEE Computer Society (December 2009) 1034–1039

[13] Noble, C.C., Cook, D.J.: Graph-based anomaly detection. In: Proc. of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. (2003) 631–636

[14] Chakrabarti, D.: Autopart: Parameter-free graph partitioning and outlier detection. In: Proc. of the 8th European Conference on Principles and Practice of Knowledge Discovery in Databases. (2004) 112–124

[15] Sun, J., Qu, H., Chakrabarti, D., Faloutsos, C.: Neighborhood Formation and Anomaly Detection in Bipartite Graphs. In: Proc. of the Fifth IEEE International Conference on Data Mining, Houston, TX, USA, IEEE Computer Society (November 2005) 418–425