

レビュアの信頼性評価における手法比較と属性比較

Comparing methods and attributes
for evaluating reviewers' credibility

田中 祐也[▼] 中村 伸子[◆]
土方 嘉徳[▲] 西田 正吾^{*}

Yuya TANAKA Nobuko NAKAMURA
Yoshinori HIJIKATA Shogo NISHIDA

近年、レビューサイトに書き込みを行うユーザ(レビュア)が増加している。しかし、レビュアの中には価値ある情報を提供する信頼できるレビュアだけでなく、あまり価値のない情報しか提供しない信頼できないレビュアも存在している。我々は信頼できるレビュアを特定することによって良質なレビューを特定することができると考える。そこで本研究では、レビュアの信頼性評価における手法間の比較を行い、レビュアの信頼性を評価する際に有効な手法を特定することを目指す。また、レビュアの信頼性評価における有効な属性についても合わせて検討を行う。

In recent years, reviews have been numerous with increasing the number of reviewers (users who comment on items). Some reviewers provide valuable information, other ones provide information which many people already know. We think that we can identify highly helpful reviews identifying credible reviewers. In this study, we aim to identify the most effective method for measuring the reviewers' credibility comparing several existing methods. We also study effective attributes for measuring the reviewers' credibility.

1. はじめに

近年、レビューサイトに書き込みを行うユーザ(レビュア)が増え、商品やサービスに対するレビュー数が増加している。しかし、レビューサイトが普及した結果、レビュアが多様化し、有益な情報を発信する信頼できるレビュアだけでなく、商品購入を検討するにあたって参考にならない情報しか発信しない信頼できないレビュアも書き込みを行うようになった。我々は、信頼できるレビュアを特定することで、購入の意思決定の判断材料となるような有用性の高いレビューを特定することができると考える。

これまでの研究から、レビュアの信頼性評価に用いることができる方法は2つある。1つは、注目アイテムへのレビューの内容をもとにレビューの質を評価し、レビューの評価値をレビュアの評価値とみなす方法である(以下、「内容に基づく手法」と記す)。もう1つは、レビュアが過去にレビューを行った履歴をもとにレビュアを直接評価する方法である(以下、「履歴に基づく手法」と記す)。

▼ 大阪大学 tanaka@nishilab.sys.es.osaka-u.ac.jp
 ◆ 大阪大学 nakamura@nishilab.sys.es.osaka-u.ac.jp
 ▲ 正会員 大阪大学 hijikata@sys.es.osaka-u.ac.jp
 * 大阪大学 nishida@sys.es.osaka-u.ac.jp

これまで、これらの研究は個別に行われてきたため、それぞれの手法間の比較は行われてこなかった。また、内容に基づく手法に関する研究は、レビューを評価することを目的としているため、注目アイテムへのレビューしか用いていない。しかし、注目アイテムへのレビューの内容だけでは、レビュアの書くレビューの質のバラつきを考慮していないと言える。我々はレビュアが過去に執筆したレビューの内容を用いることによって、レビュアの信頼性をより一般的に評価できるのではないかと考える。

そこで本研究では、上記2つの手法に、レビュアが過去に執筆したレビューの内容に基づく手法(以下、「過去の内容に基づく手法」と記す。)を加えた3つの手法の比較を行う。これまでに行われてきたレビュアの信頼性評価に関する研究と同様、本研究においても英語のレビューを対象とする。これら3つの手法を比較することによって、有用なレビューを特定するにあたり、どの手法を選択すればよいかを決定するための知見を得ることを目指す。また、それぞれの手法において、関連研究で提案されている属性間の比較は行われてこなかった。そのため、数ある属性の中でどの属性がレビュアの質と関連が高いのかは明らかになっていない。本研究では、各手法におけるそれぞれの関連研究で提案されている属性間の比較も合わせて行う。

2. 関連研究

2.1 評判情報に関する研究

評判情報に関する研究として、評判情報の意見の分類を行う研究と評判情報の意見の要約を行う研究がある。前者の研究として、Turney[3]、Pangら[4]、Daveら[5]による研究がある。彼らはレビューを対象として、そのレビューが肯定的であるか否定的であるかの分類を行っている。Turneyはレビュー文中に含まれている形容詞と副詞の意味の極性を用いて、レビューの分類を行っている。Pangらは機械学習(ナイーブベイズ、最大エントロピー法、SVM)を用いて分類を行い、その効果を調査している。Daveらは、ナイーブベイズを用いて、レビューを文単位で肯定または否定に分類している。

後者の研究として、Huら[6]、Hijikataら[7]による研究がある。Huらは、レビューから意見文を抽出し、特徴ごとに要約する方法を提示している。Hijikataらは、ネットオークションにおける出品者に対する評価コメントを、ネットオークション上の社会的関係を用いて要約する手法を提案している。

2.2 内容に基づく信頼性評価手法

注目アイテムへのレビュー内容を基にレビューの質を評価する研究として、Kimら[8]、Zhangら[9]、Liuら[10]によるものがある。彼らはあらかじめ、良質なレビューの影響因子について仮説を立て、影響因子を定量的に測るために属性候補をいくつか列挙している。そして、各研究で用いたレビューの有用性に関する正解データ(Amazon.comにおいてhelpfulと投票された割合や人手による評価値)と属性候補との相関を分析し、列挙した属性の有用性を検証している。調査はAmazon.comを対象に行われている。

Kimら[8]は、良質なレビューの影響因子として有用性の高さを挙げ、3つの関連研究の中で最も多くの属性候補を列挙している。各属性の値からレビューの有用性を予測する回帰モデルをSVR(Support Vector Regression)[12]で学習

している。

Zhangら[9]は、レビューが事実と意見の両方から構成されているかどうかがレビューの質に影響するとしている。事実という観点から“商品紹介文との類似度”的属性を、意見という観点から“各品詞の割合”と“意見性のある単語の数”的属性を用いている。各属性の値からレビューの有用性を予測する回帰モデルをSVRおよびSLR (Simple linear regression: 単回帰分析) [13]で学習している。

Liuら[10]は、良質なレビューの影響因子として、情報量の多さと読みやすさ、意見の有無を挙げている。情報量の多さを測るために“単語の数”，“単語の頻度”，“文の数”的属性を、レビューの読みやすさを測るために“段落分けの数”的属性を、意見を含んでいるかどうかを測るために“意見性のある単語を含んだ文の数”的属性を用いている。SVM (Support Vector Machine) [14]を用いてデータセットをhigh-qualityとlow-qualityに分類している。

2.3 履歴に基づく信頼性評価手法

レビュー履歴とともにレビューを評価する研究として、Riggsら[15]とChenら[16]によるものがある。彼らは、良質なレビューを提供する信頼性の高いレビューについて、あらかじめ仮定を行っている。そして、その仮定を反映させた属性候補をいくつか列挙し、列挙した属性候補の有効性を検証している。調査はEpinions.com¹を対象に行われている。

Riggsら[15]は、アイテムに対してレビュー群が付けた評価値の平均に対し、その値に近い評価値を早期に付けていたレビューがより信頼できるレビューを提供すると仮定している。Riggsらは、上記の仮定を反映した属性として“平均評価値と評価値との差”²と“投稿の時系列上の番号”を用いている。ここで、“投稿の時系列上の番号”とは、あるアイテムに対するレビューを時系列で並べた際の順序を指す。さらに、注目レビューが投稿した“レビュー数”とそのレビューがレビューした“アイテムの被レビュー数”的2つの属性の有効性についても検証している。

Chenら[16]は、あるレビューが執筆したレビューがそのアイテムが属するカテゴリにおいて高い評価を得ていれば、同一カテゴリ内におけるそのレビューは信頼できると仮定している。そのため彼らは、レビューが執筆したレビューに対する第三者からの“被評価値”をカテゴリレベルで集約し、各カテゴリにおけるレビューの“被評価値”であるReputation値を算出している。

2.4 我々の見解

2.1節では、評判情報の意見分類や意見要約の研究を紹介したが、これらは評判情報の信頼性を直接評価するものではない。レビューの信頼性評価の関連研究では、内容に基づく手法と履歴に基づく手法の関連研究において、それぞれ用いられた属性候補および組み合わせ方法は異なっている。そのため、現段階では各手法における属性の中でどの属性が最も有効であるかについて議論することはできない。また、内容に基づく手法と履歴に基づく手法では、関連研究で用いられているデータセットが異なるため、これらの手法について単純に比較することもできない。

そこで、本研究では共通のデータセットを用いて手法間の

比較を行う。そして、これまでの研究で用いられた属性を全て用いて属性間の比較も行う。

3. 実験の設計

3.1 実験の目的

実験の目的の1つ目は、関連研究で用いられた属性と、本研究で提案する過去の内容に基づく手法で用いる属性のうち、レビューの信頼性に対する評価値を算出する際に用いるべき有効な属性はどれかを検討することである。そこで、各属性の属性値と、レビューの評価値の正解値との相関を計測する。具体的な正解値については後述する。

実験の目的の2つ目は、レビューの信頼性に対する評価値を算出する際に、内容に基づく手法、履歴に基づく手法、過去の内容に基づく手法のうち、どの手法を用いるべきかを検討することである。なお、過去の内容に基づく手法では、カテゴリ間でレビューの書き方が異なると考え、同一カテゴリにおいて過去に執筆したレビューを対象とする。まず、それぞれの手法において、使用する属性の属性値からレビューの評価値を予測するモデルを学習する。そして、学習したモデルによる予測値と実際の評価値との相関を計測することで有効な手法について検討する。

3.2 使用する属性

本研究で用いる属性(表1)の属性値の算出方法について述べる。本研究では、関連研究で用いられた属性を全て用い、関連研究と同じ方法で属性値の算出を行う。なお、下記手法で算出された属性値は、同一属性の属性値のうち最大となる値で割ることによって正規化する。

内容に注目した属性

内容に基づく手法における各属性の属性値の算出方法について述べる。“単語数”は対象レビューを構成する全単語数を、“商品名数”は対象レビュー中で対象商品名が使われた回数を用いる。“単語頻度”は対象レビューを構成する全単語に対してtf-idf値を算出し、その平均値を用いる。df(出現文書数)値を算出する際の文書集合として、対象商品が属するカテゴリ内の商品に対する全レビュー(cate)と対象商品に対する全レビュー(item)の2種類を設定した。また、単語として1文字(unigram)と隣接2文字(bigram)の2種類を用いる。

各品詞の割合を算出するためにレビューを品詞解析し、当該品詞と判定された単語数を全単語数で割る。本研究では単語の品詞分析ツールとして、Apple Pie Parser³という英語構文解析システムを用いる。wh-句としてwhat, where, when, which, why, who, whose, howを用いる。“意見性のある単語の数”は意見語辞書中に出てくる単語数を用いる。本研究では、意見語辞書として、ハーバード大学が提供するGeneral-Inquire Dictionariesという辞書を用いる⁴。“商品特徴語の数”は商品特徴語辞書に出現する語の個数，“商品特徴語の頻度”は商品特徴語辞書に出現する語の回数を用いる。商品特徴語辞書は、米レビューサイトEpinions.comにおける特性を利用して自動で辞書を作成したKimら[9]の方法に従い、作成した。

“段落数”および“段落の平均長”は改行タグ数および各改行タグ間の単語数の平均値を用い、“文の数”および“文の平

¹ <http://www1.epinions.com>

² 注目レビューが評価した各アイテムにおいて全レビューが付けた星の数の平均値とレビュー自身の星の数との差を算出し、全アイテムに対し総和をとっている。

³ <http://nlp.cs.nyu.edu/app/>

⁴ http://www.wjh.harvard.edu/_inquirer/

表 1 本研究で用いた属性
Table 1 Attributes used in this work

属性番号	レビュー内容に注目した属性	属性番号	
1	単語数	32	平均星の数との差
2	商品名数	33	カテゴリ別レビュー数
3	単語頻度(uni-cate)	過去のレビュー内容に注目した属性	
4	単語頻度(uni-item)	34	単語数
5	単語頻度(bi-cate)	35	商品名数
6	単語頻度(bi-item)	36	単語頻度(uni-cate)
7	固有名詞の割合	37	単語頻度(uni-item)
8	名詞の割合	38	単語頻度(bi-cate)
9	感嘆詞の割合	39	単語頻度(bi-item)
10	動詞の割合	40	固有名詞の割合
11	数詞、序詞の割合	41	名詞の割合
12	形容詞の割合	42	感嘆詞の割合
13	副詞の割合	43	動詞の割合
14	形容詞比較級の割合	44	数詞、序詞の割合
15	形容詞最上級の割合	45	形容詞の割合
16	wh-句の割合	46	副詞の割合
17	意見性のある単語の数	47	形容詞比較級の割合
18	商品特徴語の数	48	形容詞最上級の割合
19	商品特徴語の頻度	49	wh-句の割合
20	段落数	50	意見性のある単語の数
21	段落の平均長	51	商品特徴語の数
22	文の数	52	商品特徴語の頻度
23	文の平均長	53	段落数
24	商品特徴語を含む文の数	54	段落の平均長
25	否定的意見文の割合	55	文の数
26	肯定的意見文の割合	56	文の平均長
27	商品紹介文との類似度	57	商品特徴語を含む文の数
28	レビューが付けた星の数	58	否定的意見文の割合
属性番号	レビュー履歴に注目した属性	59	肯定的意見文の割合
29	全レビュー数	60	商品紹介文との類似度
30	被レビュー数	61	レビューが付けた星の数
31	投稿の時系列上の番号		

均長”は句点記号の数および各句点記号間の単語数の平均値を用いる。また，“商品特徴語を含む文の数”は前述の商品特徴語辞書中に出現する語句が含まれる文の数を，“否定的意見文の割合”は、前述の意見語辞書において否定的な意味合いで用いられる語とされている単語が含まれる文の割合を用いる。“肯定的意見文の割合”は、前述の意見語辞書において肯定的な意味合いで用いられる語とされている単語が含まれる文の割合を用いる。“商品紹介文との類似度”はAmazon.comの編集者によって書かれた商品紹介文とのコサイン類似度を用い，“レビューが付けた星の数”は、レビューが対象商品に対して付けた星の数を用いる。

履歴に注目した属性

履歴に基づく手法の属性の算出方法について述べる。履歴に基づく手法の属性についても、同一カテゴリのレビューの履歴を用いることとする。“全レビュー数”は注目レビューが過去にレビューを行った商品数を用いる。“被レビュー数”的属性値は以下の式で算出する。この式は、あまりレビューが行われていない商品に対してもレビューを行っているレビューが高く評価されるという考えに基づいている。なお、“被レビュー数の平均値”とは、注目レビューが過去にレビューを行った商品の被レビュー数の平均値のことであり、“被レビュー数の平均値の最大値”とは、全レビューの被レビュー数の平均値を算出し、その中で最大の値のことである。

$$\text{レビュー } i \text{ の被レビュー数の属性値} = \frac{\text{被レビュー数の平均値の最大値} - \text{被レビュー数の平均値}}{\text{被レビュー数の平均値の最大値}}$$

“投稿の時系列上の番号”は、あるアイテムに対する全レビ

ューを最新のものから時系列順に並べた際の番号を総レビュー数で割ったものを用いる。また，“平均星の数との差”は、注目レビューがレビューした全商品に対して、他レビューが付けた星の数の平均値と注目レビューが付けた星の数との差の絶対値を算出し、それらの総計を注目レビューがレビューした商品数で割り、その値を5から引いたものを用いた。またChenらの研究におけるレビュー評価方法から，“注目カテゴリにおけるレビュー数”という属性を新たに設定し、注目レビューが属するカテゴリにおいて、注目レビューがレビューを行った商品数を用いた。

過去の内容に注目した属性

過去の内容に基づく手法の属性の算出方法について述べる。この手法の属性は、内容に基づく手法で用いた属性と同じものを用い、属性値の算出も同様に行う。注目レビューが同一カテゴリにおいて執筆した過去の全レビューにおいて、それぞれ属性値を算出し、それらを足し合わせる。足し合わせた結果を注目レビューが執筆したレビュー数で割ったものを属性値とする。

3.3 モデルの学習と評価方法

それぞれの手法において、各属性の属性値から正解値であるレビューの評価値を予測するモデルを学習する。モデルの学習は、Kimら[8]およびZhangら[9]と同様に、あるレビューにおける複数の属性値とレビューの評価値の関係を表す式をSVM回帰 (SVM Regression) [11]によって求めることで行う。データを学習用データとテスト用データに分割し、*SVM^{light}* というツールを用いて、学習用データからモデルの学習を行う⁵。そして、得られたモデルを用いて、テスト用データからレビューの評価値を予測し、実際の評価値との相関を計測することでモデルの評価を行う。また、相関を計測する際には、ピアソンの積率相関係数 (Pearson product-moment correlation coefficient) を用いる。

4.. 実験の実施

我々は、実際の商用Webサイトにおけるレビューを用いて、3章で設計した実験を実施した。本章ではその実施結果について述べる。まず、データセットとレビューの評価値について述べた後、実験の結果と考察について述べる。

4.1 データセット

我々はAmazon.comで取り扱われている商品に対してレビューを行ったレビューを対象として、データセットを収集した。具体的には、Amazon.comにおける4つのカテゴリ、ミステリー映画(Mystery), ロックミュージック(Rock), MP3プレーヤー(MP3), デジタルカメラ(DCamera), に属する全商品に対する全てのレビューを収集した。ここで、対象カテゴリを選択基準としてレビューの書き方の違いに注目した。商品の内容に価値がある商品のカテゴリとしてMysteryとRockを、商品本体に価値がある商品のカテゴリとしてMP3とDCameraを選択した。収集したレビューのうち、閲覧者による投票数が10件以上のレビューを実験用のデータセットとした。表2に各カテゴリにおける商品数(2009年12月時点)および全レビュー数、閲覧者による投票数が10件以上のレビュー数(対象レビュー数)を示す。

⁵ <http://svmlight.joachims.org/>

表2 各カテゴリごとのデータ数

Table 2 The number of data in each category

	Mystery	Rock	MP3	DCamera	計
商品数	1,972	3,717	686	1,660	8,035
全レビュー数	232,531	209,124	121,125	111,527	674,307
対象レビュー数	7,101	10,282	13,617	23,123	54,123

4.2 レビュアの評価値

本研究では、レビュアの評価値の正解値として、Amazon.comでの投票結果（helpfulと投票された割合）を用いる。レビュアが同一カテゴリにおいて執筆した全てのレビューに対する投票結果による評価値を用いる。これは、過去の内容に基づく手法において同一カテゴリのレビューを用いている理由と同様、レビュアの書くレビューはカテゴリによって異なると考えるからである。

$$\text{レビュア } i \text{ の評価値} = \frac{\text{helpfulの総数}}{\text{helpfulの総数} + \text{nohelpfulの総数}}$$

4.3 実験結果と考察

4.3.1 全属性の有効性に関する実験

各手法で使用する属性の中で有効な属性について検討を行う。表1で示した全61属性に対して属性値とレビュアの評価値との相関係数を計測し、有効な属性を特定する。

4つのカテゴリの全てのレビューに対して、属性値とレビュアの評価値の正解値との相関係数を計測した結果の上位15属性を表3に示す。相関係数の絶対値が0.2以上のとき相関が認められる[1]ので、表3において相関係数が0.2以上の14個の属性に注目する。まず、星の数に関する属性の相関が高いことから、注目レビュアの商品に対する星の数による評価がそのレビュア自身の信頼性評価に影響していることが分かる。そして、商品特徴語に関する属性が相関が認められる属性に含まれていることから、レビュー文中で商品特徴語について言及しているかどうかが注目レビュアの信頼性評価に影響していることが分かる。さらに、“文の数”、“単語数”的属性においても相関が見られることから、レビューの分量も信頼性評価に影響していると言える。

次に、カテゴリごとの相違について検討を行う。表4に、各カテゴリにおける上位10属性を示す。表4より、どのカテゴリにおいても共通して星の数に関する属性と商品特徴語に関する属性群が上位となっていることがわかる。商品の内容に価値がある商品(Mystery, Rock)と商品本体に価値がある商品(MP3, DCamera)では評価される基準が異なると考えていたが、実験結果を見ると、商品の内容に価値がある商品と商品本体に価値がある商品の差は明確ではなかった。しかしながら、カテゴリ間で有効な属性に差があることは確認できる。Mystery, Rock, DCameraでは、星の数に関する属性が上位に位置するのに対し、MP3では商品特徴語に関する属性群が上位に位置している。

4.3.2 各手法の有効性に関する実験

手法間の有効性の比較実験について述べる。まず、各手法とも4.3.1節の実験における上位5属性を用いて予測モデルを学習し、3つの手法の比較を行う。これは履歴に基づく手法の属性数が5個であることによる。各手法において、算出された予測値と正解値の相関を求め、3手法の相関係数の比較を行う。

表3 正解値と各属性値との相関係数の上位15個(4つのカテゴリの全てのレビュー)

Table 3 Top 15 attributes in correlations between actual helpfulness and attribute value (all categories)

順位	属性番号	属性の名前	相関係数
1	61	星の数(過去)	0.5743
2	28	星の数	0.5541
3	32	平均星の数との差	0.4925
4	51	商品特徴語の数(過去)	0.4041
5	18	商品特徴語の数	0.3971
6	52	商品特徴語の頻度(過去)	0.3478
7	19	商品特徴語の頻度	0.3387
8	34	単語数(過去)	0.3134
9	1	単語数	0.2967
10	57	商品特徴語を含む文の数(過去)	0.2694
11	55	文の数(過去)	0.2626
12	24	商品特徴語を含む文の数	0.2582
13	22	文の数	0.2571
14	31	投稿の時系列上の順位	0.2515
15	54	段落の長さ(過去)	0.1787

次に、内容に基づく手法と過去の内容に基づく手法において、4.3.1節の実験で相関係数が高かった順に属性を1つずつ増やしながら予測モデルの学習を行い、有効な属性の組み合わせを発見する。これにより、内容に基づく手法と過去の内容に基づく手法において、上位5属性よりも属性数を増やした場合に、結果がどこまで良くなるかを確認する。得られた有効な属性の組み合わせからモデルを学習し、上位5属性の比較実験と同様に3手法の相関係数の比較を行う。

上位5属性における手法間の比較実験

まず、上位5属性における手法間の有効性の比較実験の結果を示す。各手法の予測値と正解値の相関を図1に示す。Allは4つの全てのカテゴリにおける結果を表している。

図1から、どのカテゴリにおいても、過去の内容に基づく手法が最も高い相関を得ていることがわかる。All, Mystery, Rock, DCameraでは、過去の内容に基づく手法が、他の2手法に比べて有意であることが確認できた。なお、検定には相関係数の同等性の検定[2]を用いている(All, Mysteryは有意水準1%, Rock, DCameraは有意水準5%)。

MP3では、過去の内容に基づく手法が最も高い相関を得ているが、内容に基づく手法と比べて有意差は見られなかった。MP3は、同じカテゴリの商品を何度も買うということが少なく、同一カテゴリにおいてレビューを複数回書くということが少ない。実際、表5に示すようにレビューの平均レビュー数はMP3が最も少なくなっている。このことが有意差が得られなかった原因であると考えられる。

有効な属性の組み合わせの発見と手法間の比較実験

内容に基づく手法と過去の内容に基づく手法の4つのカテゴリにおける、予測モデルの学習に用いた属性数に対する精度（正解値と予測値との相関係数）の推移を図2に示す。

表 4 正解値と各属性値との相関係数の上位 10 個(カテゴリ別)

Table 4 Top 10 attributes in correlations between actual helpfulness and attribute value (in each category)

順位	Mystery	Rock	MP3	DCamera
1	星の数(過去)	星の数(過去)	商品特徴語の数(過去)	星の数(過去)
2	星の数	星の数	商品特徴語の数	星の数
3	平均星の数との差	平均星の数との差	商品紹介文との類似度(過去)	平均星の数との差
4	単語数(過去)	商品特徴語の数(過去)	商品紹介文との類似度	商品特徴語の数(過去)
5	投稿の時系列上の番号	商品特徴語の数	商品特徴語の頻度(過去)	商品特徴語の数
6	文の数(過去)	商品特徴語の頻度(過去)	単語数(過去)	商品特徴語の頻度(過去)
7	商品特徴語を含む文の数(過去)	商品特徴語の頻度	商品特徴語の頻度	商品特徴語の頻度
8	商品紹介文との類似度(過去)	単語数(過去)	単語数	単語数(過去)
9	単語数	単語数	星の数(過去)	単語数
10	商品特徴語の数(過去)	投稿の時系列上の番号	星の数	商品特徴語を含む文の数(過去)

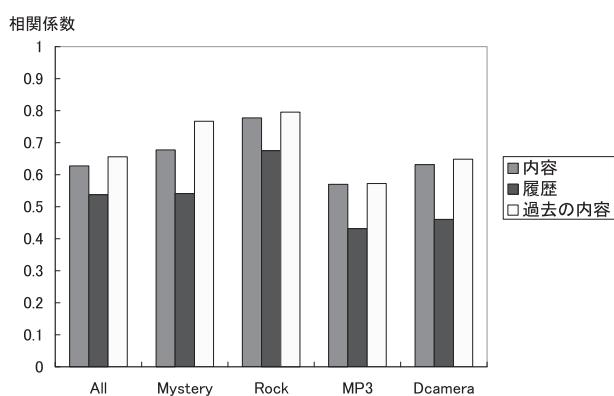


図 1 各手法における予測値と正解値の相関(上位 5 属性)

Fig. 1 Correlations between the predicted values and the actual helpfulness (Top 5 attributes)

表 5 各カテゴリにおけるレビューの平均レビュー数

Table 5 The average number of reviews given by a reviewer

Mystery	Rock	MP3	DCamera
3.282	1.731	1.335	1.592

図2より、どちらの手法も、同じように精度が推移していることがわかる。属性数を2個にしたときに相関係数が急激に上がり、その後下がる所もあるが、属性数10個辺りまでは徐々に相関係数が上昇している。その後も上昇傾向にあるが、計算コストも考慮すると、上位10個までの属性の使用が妥当であると考える。よって、我々は属性の有効性の比較実験において相関の高かった上位10個の属性を用いて予測モデルの学習を行う。

内容に基づく手法と過去の内容に基づく手法については上位10属性、履歴に基づく手法では全5属性を用いてモデルを学習する。各手法の予測値と正解値との相関係数の結果を図3に示す。

図3から、上位5属性の場合と同様に、どのカテゴリにおいても過去の内容に基づく手法が最も高い相関を得ていることがわかる。また、All, Mystery, Rock, DCameraについては、過去の内容に基づく手法が他の手法に比べて有意であ

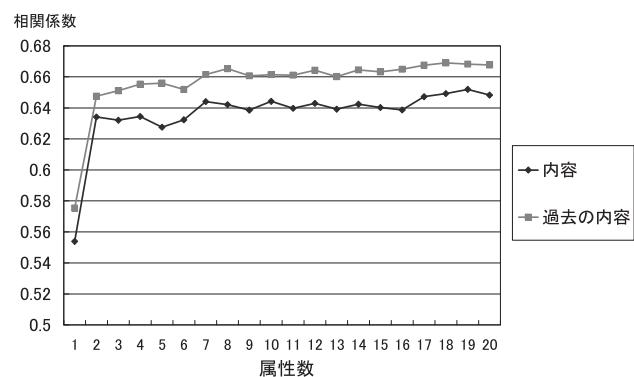


図 2 予測モデルの学習に用いた属性数の変化に対する予測精度の推移

Fig. 2 A transition of the prediction accuracies against the number of attributes

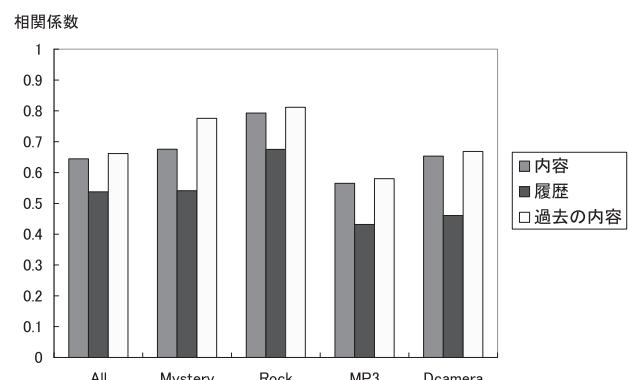


図 3 各手法における予測値と正解値の相関(上位 10 属性)

Fig. 3 Correlations between the predicted values and the actual helpfulness (Top 10 attributes)

ることが確認できた(All, Mystery, Rockは有意水準1%, DCameraは有意水準5%). MP3については、上位5属性の場合と同様、有意差は確認できなかった。しかしながら、有意差は得られなかつたものの、過去のレビューを用いた場合の方が高い相関が得られているため、レビューが過去に執筆し

たレビューを用いることによって信頼性評価の結果がよくなる可能性があると言える。以上のことより、レビューの信頼性評価を行う際、過去の内容に基づく手法を用いることが最も有効であると言える。

5. おわりに

本研究では、レビューの信頼性を評価する際にどのような手法を用いればよいかを決定するための知見を得ることを目的とし、2つの実験を行った。具体的には、レビューの信頼性評価における属性の有効性に関する実験と3つの手法（内容に基づく手法、履歴に基づく手法、過去の内容に基づく手法）の有効性に関する実験を行った。

実験結果から、レビューの信頼性評価における属性の中では、星の数に関する属性と商品の特徴に関する属性が有効であることを特定した。また、信頼性評価の3つの手法の中では、過去の内容に基づく手法を用いることが最も有効であることを特定することができた。今後の課題は、各手法を組み合わせたハイブリッド手法について検討し、レビューの信頼性評価において最も有効な属性の組み合わせを検討することである。

[文献]

- [1] 石村貞夫, すぐわかる統計解析, 東京図書, (1993).
- [2] 肥田野直, 濑谷正敏, 大川信明, 遠藤健治. 心理・社会・教育系のための統計入門, 培風館, (1995).
- [3] Turney, P.D., 'Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews', *Proc. of the 40th Annual Meeting on Association for Computational Linguistics*, pp. 417-424, (2001).
- [4] Pang, B., Lee, L. and Vaithyanathan, S., 'Thumbs up? Sentiment Classification using Machine Learning Techniques', *Proc. of ACL EMNLP'02*, pp. 79-86, (2002).
- [5] Dave, K., Lawrence, S. and Pennock, D.M., 'Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews', *Proc. of WWW'03*, pp.519-528, (2003).
- [6] Hu, M. and Liu, B., 'Mining and Summarizing Customer Reviews', *Proc. of ACM KDD'04*, pp. 168-177, (2004).
- [7] Hijikata, Y., Ohno, H., Kusumura, Y. and Nishida, S., 'Social Summarization of Text Feedback for Online Auctions and Interactive Presentation of the Summary', *Proc. of ACM IUI'06*, pp. 242-249, (2006)
- [8] Kim, S., Pantel, P., Chklovski, T. and Pennacchiotti, M., 'Automatically Assessing Review Helpfulness', *Proc. of ACL EMNLP'06*, pp. 423-430, (2006).
- [9] Zhang, Z. and Varadarajan, B., 'Utility scoring of product reviews', *Proc. of ACM CIKM'06*, pp. 51-57, (2006).
- [10] Liu, J. et al., 'Low-Quality Product Review Detection in Opinion Summarization', *Proc. of ACL EMNLP-CoNLL'07*, pp. 334-342, (2007).
- [11] Joachims, T., 'Making large scale SVM learning practical', in *Advances in Kernel Methods: Support Vector Learning*, Scholkopf, B., Burges, C.J.C., Smola, A.J. (eds), MIT Press.Cambridge, MA, (1999).
- [12] Chang, C.-C. and Lin, C.J., 'LIBSVM: a library for support vector machines', <http://www.csie.ntu.edu.tw/~cjlin/libsvm>, (2001).

[13] Witten, I.H. and Frank, E., 'Data Mining: Practical machine learning tools and techniques', Morgan Kaufmann, San Francisco, (2005).

[14] Vapnik, V.N., 'The Nature of Statistical Learning Theory', Springer, (1995).

[15] Riggs, T. and Wilensky, R., 'An Algorithm for Automated Rating of Reviewers', *Proc. of ACM/IEEE-CS JCDL'01*, pp. 381-387, (2001).

[16] Chen, M. and Singh, J.P., 'Computing and using reputations for internet ratings', *Proc. of ACM EC'01*, pp.154-162,(2001)

田中 祐也 Yuya TANAKA

2010年、大阪大学基礎工学部システム科学科卒業。現在、同大学大学院修士課程在学中。情報信頼性に関する研究と情報フィルタリングの研究に従事。

中村 伸子 Nobuko NAKAMURA

2007年、大阪府立大学工学部卒業。2009年、大阪大学大学院基礎工学研究科修士課程修了。同年、パナソニック電工入社。在学中は情報信頼性に関する研究に従事。

土方 嘉徳 Yoshinori HIJIKATA

1996年、大阪大学基礎工学部システム工学科卒業。1998年、同大学大学院修士課程修了。同年、日本アイ・ビー・エム(株)東京基礎研究所入社。2002年より、大阪大学大学院基礎工学研究科システム創成専攻助手。2009年より、同准教授。知的Web技術、情報推薦、テキストマイニングの研究に従事。情報処理学会、電子情報通信学会、ほか会員。

西田 正吾 Shogo NISHIDA

1974年、東京大学工学部電子工学科卒業。1976年、同大学大学院修士課程修了。同年、三菱電機(株)入社。同社中央研究所システム基礎研究部研究員、グループマネージャーを経て、1995年、大阪大学基礎工学部システム工学科教授。ヒューマンインターフェース技術、メディア技術の研究に従事。情報処理学会、IEEEほか会員。