修士論文

信頼値モデルによるオンラインショッピングサイトでの投稿判断支援に関する研究

1910252　Xue, Feifang

主指導教員　金井 秀明

北陸先端科学技術大学院大学
先端科学技術研究科
（知識科学）

令和 3 年 3 月

A Study on Decision Support for User Comments on Online Shopping Sites Using Confidence Value Models

Xue, Feifang

School of Knowledge Science,
Japan Advanced Institute of Science and Technology

March 2021

Keywords: latent semantic analysis, identification of fake reviews, level of dislocation

Abstract

With the rapid development of e-commerce, consumers are keen on online shopping and publishing reviews. These reviews provide important reference information for stores, manufacturers, and potential consumers. However, the reviews may be false, and false reviews will influence the reference value to a certain extent. Therefore, it is essential to identify false e-commerce reviews.

This article mainly focuses on e-commerce reviews, focusing on reviewing explicit features, implicit features, and the combinations of both. The following research work has been completed:

(1) The current false reviews recognition is performed through explicit features of reviewer or review text. It does not consider the review text's semantic information, and the analysis result is not accurate enough. Therefore, an electronic business false review recognition method based on implicit semantic analysis is proposed. Based on the analysis of user behavior characteristics, this method adds semantic analysis information of review text, making it possible to identify false reviews at the semantic level of the text and improve the recognition accuracy.

(2) We use the dislocation level method to evaluate the experimental results. Due to the particularity of the experimental method, it is not convenient to use traditional assessment methods. Therefore, the dislocation-level assessment method was used in this paper.

(3) Finally, by providing a reference trust value, the detection proposal system can strengthen consumers' ability to identify fake reviews.

目次

図目次

表目次

第１章　はじめに

1.1 研究背景と目的

1.1.1 研究背景

インターネット技術の発達により，ユーザーとのコミュニケーションを重視するサイトが増えている．ユーザーが様々な方法でコメントを残すことで，自分の体験や意見を伝えることができるようになっている．例えば，日本の楽天市場，中国の淘宝網（world.taobao.com），京東商城（global.jd.com）などでは，オンラインで買い物をした後，顧客はレビューコメントにより，購入した商品についての自分の経験やコメントを書くことができる．オンラインショッピングが懸念されている限り，電子商取引の売り手は顧客からのレビューが大幅に潜在的なバイヤーの行動に影響を与えると考え，顧客のレビューに多くの注意を払う必要がある．一方で，これらのレビューは，電子商取引の販売者に有益なコメントを提供することができ，サービスの質の向上に役立ち，メーカーが製品の全体的な品質を向上させるのにも役立つ．消費者にとって，他者の購入レビューは，買うか買わないかを決めるための重要な情報でもある．例えば，ある製品を買う前に，関連の投稿を読み，ほとんどの投稿がポジティブなものであれば，製品購入の可能性が高くなる．一方，投稿の多くがネガティブなものであれば，製品購入しなくなる可能性がある．

ネット上の口コミ情報に関するアンケート調査によると，商品・サービス購入時に他の消費者のレビューの影響を参考したことがある人は６割に達す．EC サイト（Amazon，楽天市場など）のレビューに対して，他の消費者のレビューを重視する傾向が増加している．55.7%の消費者は製品レビューが購入の意思決定に大きな影響を与えると感じている[1]．また，「購買行動におけるクチコミの影響」に関する調査によると，71.8%の消費者が悪いレビューを見ることで製品に対する以前の態度が変わって非常に気になると考えている[2]．これらの数字は，レビューが消費者に与える影響が大きいことを示唆している．

商品レビューは，電子商取引の売り手や消費者にとって非常に有益なものである．一方，重要な参考となる電子商取引のレビューの信頼性に質疑を持たれる場合がある．商品レビューのすべてが真実であり，信頼できるものであるとは限

らない．企業間の競合による自己利益のために，販売側はやらせ投稿者を雇い，一般の利用客に扮して自社の製品を高く評価し，また競合他社の製品を悪意的に歪曲する場合がある．このような故意に投稿するレビューは「やらせ投稿」と呼ばれ，やらせ投稿により消費者や競合他社の利益を著しく害し，消費者不正確な情報をもとに商品を購入する場合がある．そのため，本当に良い商品が売れず，その出店者が損害を受ける．一方，やらせ投稿によって質の悪い商品が売れ，その出店者が利益を得ることになりかねない．

このようなやらせ投稿の問題に対し，レビュー判定サイト「サクラチェッカー」というものがある[3]．消費者はこのサイトにより購入予定の商品レビューがやらせ投稿かどうかを判定でき，消費者の正しい商品認識の確立に役になると考える．その判定方法は，ある商品の URL を入力して，やらせ投稿の度合いである「サクラ度」を確認することができる（図１）．その商品のレビューに対して，大まかな傾向を分析した上で，日毎のレビュー数の増減を分析して，やらせ投稿の存在確率を判断できる．

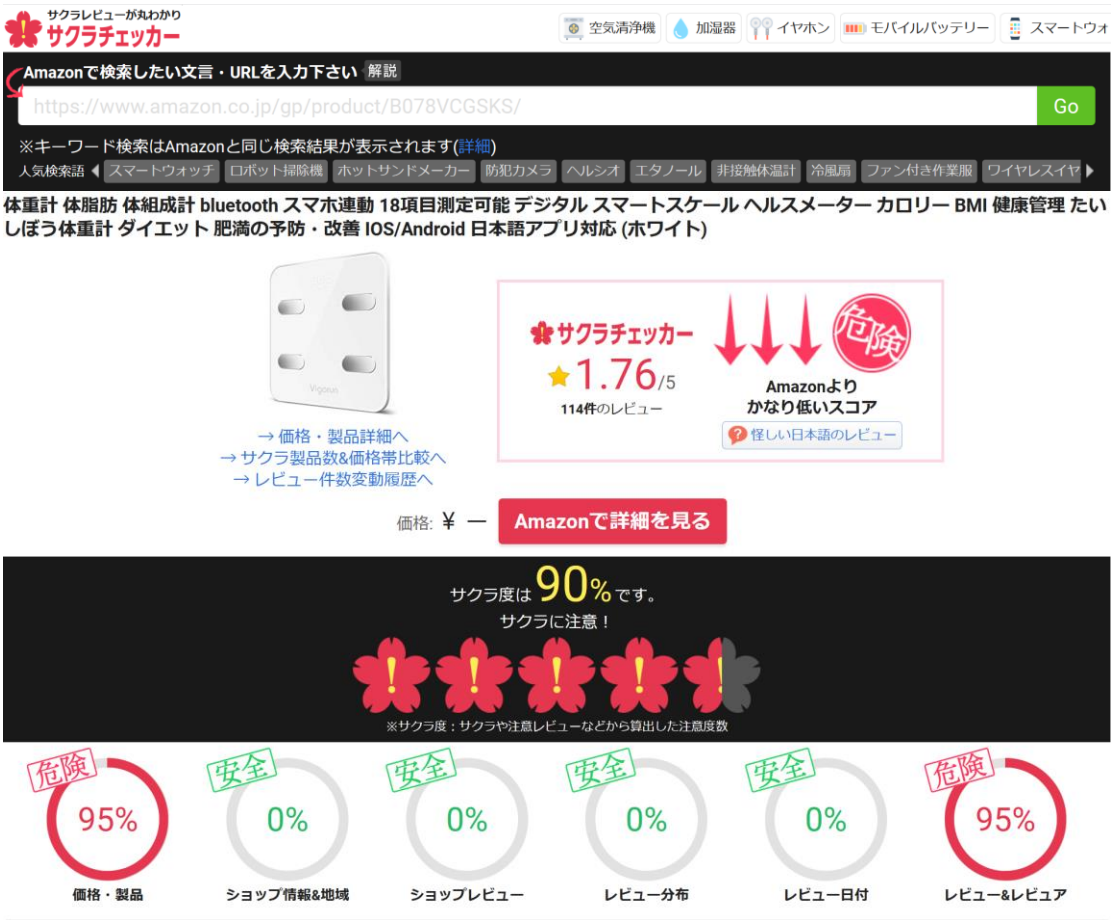

図 1　サクラチェッカーによってある商品のサクラ度の表示画面

「サクラチェッカー」では，各レビューがやらせかどうかは示さずに，商品に関する警報の度合いのみを提示する．大量のやらせ投稿によって消費者の利益を害するが，一般消費者が効果的にやらせ投稿をスクリーニングすることは困難であることが研究で明らかになっている[4]．従って，ネット消費環境を完備するには，やらせ口コミなどの虚偽投稿を選別するシステムの構築が必要である．

1.1.2 研究目的

本研究は，（1）読み手の受ける印象からやらせ投稿を代表できる主な特徴を抽出し，（2）それに基づいた機械学習を用いて，やらせ投稿への「信頼値モデル」を構築することを目的とする．また，構築する信頼値モデルによって，消費者によるやらせ投稿に対する識別能力を強化させたかどうかを検証する．

1.2 本論文の構成

本論文は第 1 章である本章を含めて全 6 章で構成される．

第 2 章では関連研究について取り上げる．

第 3 章ではオンラインショッピングサイトでの投稿の獲得とデータの前処理について解説する．

第 4 章では提案モデルについて記述する．

第 5 章では実験の詳細および得られた結果から考察を行う．

第 6 章では本研究のまとめと今後の展望について述べる．

第 2 章　関連研究

2.1 やらせ投稿の定義

Jindai らは商品レビューには以下の 3 つのタイプを提案している.

- Type1：不真実な投稿(untruthful opinions)：読者を誤解させる虚偽的なレビュー.
- Type2：ブランドに関する投稿(review on brands only)：製品についてではなく，電子商取引の販売者などだけについてのレビュー.
- Type3：無関係の投稿(non-review)：広告，ウェブサイトの URL，およびほかの無関係の内容があり，または意見を含まないレビュー[5,6,7].

Type1 の偽物レビューは人間の読者でも本物のレビューと偽物のレビューの違いを見分けるのが困難で，機械で見分けるのは非常に難しいと言われている. 他の文献では，レビューを真と偽の 2 つに分類しており，レビューがブランドに関するものであるか, 商品に関連するものであるかに関わらず, レビューは真と偽に分類されている[8]．ただし，多くの研究では，Jindai らの分類を用いている[6,7]．定義によって異なる識別方法が使用されており，Type2 と Type3 の識別技術は比較的成熟しており，現在の研究では消費者を誤解させる可能性のある真実ではないレビュー（Type1）に焦点が当てられている[9]．この分野では，一部の研究では「重複レビュー」のアプローチが用いられている[7,10]．本研究では,「消費者を誤解させるような真実のないレビューを特定する」という目的とし，実際に商品を購買せず（または使用せず）投稿したレビューや商品本物と関係なしレポーをやらせ投稿と定義し，Jindai らの分類に従って議論する.

2.1.1 ブランドに関する投稿

Jindai らは, 投稿者とレビューからテキスト特徴を含む有用なメタデータ特徴を抽出し，ブランドと無関係なレビューを検出するために回帰モデルを使用した. サポートベクターマシンやベイズ法を用いた研究もあるが, 回帰モデルの方が相対的に優れていると指摘されている[7].

S. Patil M らは，ブランドに対して投稿したユーザ ID の重複性に基づいて，ブランドに関するレビューを見つけるために決定木を用いている[10]．決定木は次のように生成される．2 つの投稿内容の語順確率をイテレータで計算した閾値と比較し，その確率が閾値よりも大きければ，投稿は「コピー」であることを示した．例えば，R1 と R2 が 2 つのレビューだとし，閾値は 0 から 1 の間にある値である．n=gram モデル式を用いて，レビューR2 の単語列の確率，すなわちレビューR1 の単語列についての P(R2)を求める．閾値を越えると，ユーザー ID は重複で，またはスパマーのより可能性の高いセットを与え，やらせ投稿であると考えた．

$$p(w_n \mid w_{n-1}) = \frac{c(w_{n-1}.w_n)}{c(w_{n-1})}$$

図 2　S. Patil M らにより n=gram モデル

2.1.2 無関係な投稿

Jayanthi S K らは，レビューコンテンツ内の URL リンクを識別するために DBSpamClust を使用した[11]．ネットでリンクとその内容を取得し，各ホストとページについて，特定の特徴を計算し，重要なリンクベースの特徴をいくつか挙げる．これらの特徴を合わない場合，スパムページとしてマークされ，やらせ投稿と判別される．

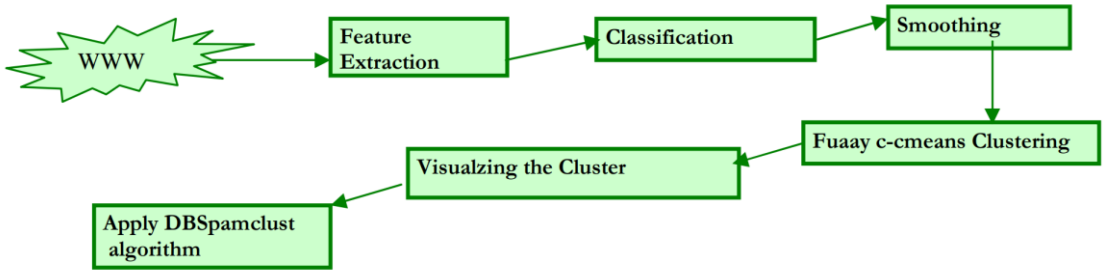

図 3　Jayanthi S K らにより DBSpamClust 方法

また，Sihong Xie らはサポートベクターマシン使用して，語彙，構文，ジャンルの特徴を識別し，無関係なレビューを識別した[9]．Xu X らは製品との関連性が低い投稿，およびウェブページや他のドメインに関連した「フィラー」やキーワードを含むレビューは無関係投稿であると結論づけている[12]．

2.1.3 複製された投稿

Xu X らは, 投稿内容の類似度を比較し, 類似度が一定の設定値を超えると「複製された投稿」にみなす[12]. 伊木らは, 複製されたレビューをやらせ投稿と定義する. レビューを投稿する人の行動に注目し, 文章を単語により区切って 1 つの要素とし, 各投稿の類似要素の比率から類似度を計算する. その類似度を信頼性判断に関する分類指標とした[13].

ブランドに関する投稿, 無関係な投稿, 複製された投稿の他にも, これら 3 つに当てはまらない投稿も虚偽的である可能性がある. これらはより判別が難しく, 多くの研究者は特徴に基づいてやらせ投稿を分析する.

2.2 特徴に基づくやらせ投稿検知

2.1 節で紹介した Type1 の読者を誤解させる虚偽的な投稿に対して, 大量な研究が特徴に基づく投稿分類を行い, 以下のような側面からのがある.

(1) 投稿行動や投稿者の会員レベルなど投稿者の特徴に基づく.

(2) 投稿のデータ機能とテキスト機能を含む投稿特徴に基づく. 投稿のデータ機能にはスコア, 投稿時刻, 日付などが含まれ, 投稿のテキスト機能には, テキストの長さ, 感情語数, または関連語数などが含まれる.

(3) 投稿者特徴と投稿特徴の融合に基づいて, より良い結果が得られる現在の方法を利用している.

(1)については, Lim らはある投稿者の複数回の投稿行動を分析してやらせ投稿を検出した. 投稿データを調査したところ, やらせ投稿者は特定の商品を複数回投稿し, 投稿文の内容や評価が類似していること, また, やらせ投稿者は短期間で対象商品に高い評価や低い評価を与える傾向があることが明らかになった[14].

(2)については, 岡山らはテキストに用いられる表現の特徴に差異があるという考えに基づいて分類を行っている[15]. Youli らは文字レベル, 単語レベルと構文レベルの特徴を抽出した上で, 字母・単語の重複性, 大小文字, スペース, 句読点及び機能性単語の使用頻度を分析した[16]. しかし, 漢字の書式やテキストの意味が英語より変わりやすくて, 本文の様式からやらせレビューを判断す

ることや難しいと指摘されている．

（3）については，Crawford M らはどの言語にも適用される明確的な特徴として，評価者の会員レベル，評価本文の長さ，評価スコアという三つの特徴に基づいて投稿分類を行う[17]．Jingdong W らは投稿の基本的特徴：商品（値段，平均スコア），レビューの長さ，評価者（信用，発信の方式）と，感情の極性（ポジティブ/ネガティブ），テキストの内容類似度やスタイルの類似度を融合し特徴抽出アルゴリズムを設計し，やらせ投稿検出のための分類モデルを構築する[18]．

2.3 本研究の位置づけ

先行文献のように投稿の文書や投稿時刻，投稿者の特徴に基づく投稿分類を行う従来の手法では，単語や書き手である評価者の推定といった「投稿の書き手の視点からの側面」から分析するものが主流である．「投稿者側」の視点から分類を行う手法により，抽出された特徴が研究者たちの主観的観点に基づいた代表性が低く，「閲覧者側」の視点からその抽出した特徴を承認されない可能性がある[19]．そのため，やらせ投稿を代表するものとは難しい．また，これらの研究では，構築したモデルがやらせ投稿の識別に対する消費者への影響があるかどうかも検証されていない．そこで，本研究では，投稿された文章と投稿者の顕在的な特徴に基づき，「消費者が投稿を読むときに受ける印象という視点」からやらせ投稿を代表する明示的特徴を抽出し，多項式回帰を用いて信頼値公式を導入し，投稿への信頼度を計算する．それを暗黙的特徴と融合させ，総合的な「信頼値モデル」が得られる．また，各投稿の信頼値を提示し，消費者によるやらせ投稿に対する識別能力を強化させたかどうかを検証する．

2.4 第 2 章のまとめ

本章では，2.1 節から 2.2 節でこの 2 つの側面から関連研究を述べた．2.3 節で本研究の位置づけを述べた．次章では，オンラインショッピングサイトでの投稿の獲得とデータの前処理について述べる．

第 3 章　提案手法

やらせ投稿の特定プロセスとし，最初に投稿データを取得し，データの前処理を行う．投稿データの前処理としては，非テキストデータのクリーニング，統合，正規化，変換などが含まれる．テキストデータの前処理としては，単語分離，テキストベクトル化，次元削減などがある．そして，やらせ投稿を特定するために，ある特徴を持つ投稿がやらせ投稿であると仮定してモデルを作成する．やらせ投稿識別の問題はテキスト分類の問題として処理ができ，その多くは機械学習アルゴリズムを用いてやらせ投稿を識別し，最終的に識別結果を評価する．全体の流れを図 4 に示す．

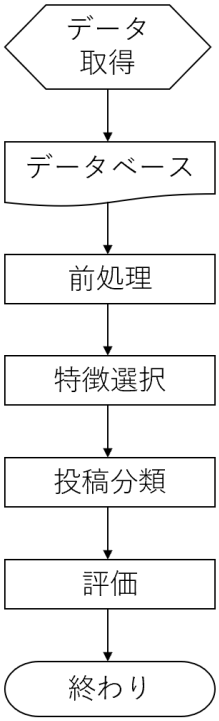

図 4　やらせ投稿識別問題の流れ

3.1 投稿データ取得

特徴はデータに由来するが，実際には，データは，ある特徴を選択するために選択できる大量の特徴を提供することが必要である．すなわち，良い実験結果を得るためには，良い特徴のセットを持っていなければならない．特徴がより重要になるだけでなく，データのバランスなども研究実験では非常に重要になる．やらせ投稿を特定するために使用されるデータセットとしては，公開データセットと非公開データセットの2つのタイプがある．多くの研究では「gold-standard」の公開データを使って，英文のやらせ投稿を特定している．そのデータ数は合計1600件で，そのうち800件が真，800件が偽レビューである．真と偽のレビューのうち400件がよいレビュー，400件が悪いレビューである．一方，他の研究では，特定の商用サイトを独自のクローラでデータを収集し，またパートナーから提供されたものもある．商用サイトとしては，Amazon，楽天市場（日本），淘宝網，京東商城（中国）などがある．

次に，他の研究で利用されているデータセットについて述べ，本研究で用いるについて説明する．

3.1.1 公開データセットに基づく

「gold-standard」の公開データを利用するメリットは，他の研究と実験結果を比較することができ，ラベル付きデータを利用できることである．そのため，多くの学者は「gold-standard」の公開データセットを使ったり，公開データセットに何らかの変更を加えたりして，以前のデータセットとの比較実験を行うことを選択している[20,21,22,23]．しかし，データ数が少なすぎて，テキストでしか特定できないという問題点がある．

3.1.2 非公開データセットに基づく

他の研究では，自分でラベルを付けている非公開のデータセットを用いて実験を行っている．このような実際のデータに基づく実験結果は，他の研究との比較が容易ではなく，公開されているデータセットと比較しても良い結果が得られない可能性がある．プライベートデータセットはデータ数が多い，一方，アノテーションなどの点で問題がある．そのため，やらせ投稿識別の調査を行う際には，投稿データが結果に大きく影響を与える．

岩井らは Amazon のレビューデータセットを使用して，やらせ投稿を特定する[24]．三舩らは，楽天市場の「みんなのレビュー・口コミ」というレビューデータソースに基づいてやらせ投稿の特徴を抽出する[25]．劉らは中文のやらせ投稿を注目し，中国の淘宝網（world.taobao.com）からデータを獲得して投稿識別を行う[26]．

3.1.3 JD.com での投稿の獲得

本研究では，中文のやらせ投稿を対象とし，中国の京東商城（global.jd.com，後文を「JD」と略称する）の投稿データを用いる．JD は日本でいうアマゾン型モールである．自社モール内に JD 名義で商品を販売するモデル（いわゆる仕入れ販売）が主な販売方法である．電子製品の売上が全体の 50%を占めており，総合通販というよりも電子製品専門モールの位置付けで中国国内では認知されている[27]. 2020 年 9 月 24 日では, iPhone 11 という製品のレビュー数は 475 万件に及んでいる．また，JD のデータセットは多くの特徴が含まれており，やらせ投稿に関する特徴を選択するのに有利である．データソースはプライベートデータ型であり，レビューのメタデータとレビューのテキストデータを含む．例えば，JD からクローラされた iPhone 11 の投稿では，主な項目として投稿者の ID，投稿のテキストの内容，投稿時間，購買時間，「この投稿が有用」とクリックされた数，フィードバックの数，評価スコアがある．この実験で使用したデータは，前処理後の 1961 項目を合計したものである．データの一例を図 5 に示す．

ユーザID	評価本文	購買時間	有用と言われ	フィードバック	スコア	評価時間
14247978741	运行速度：是在京东小当家的直播间下单，送了	2020/6/1 2:08:21	78	22	5	2020/6/6 22:13:25
14283691428	双摄像头的拍照清晰，续航能力强，充电速度快	2020/6/12 0:02:27	145	52	5	2020/6/14 18:28:27
14238206677	拍照效果：是在京东小当家的直播间下单，送了	2020/6/1 2:23:51	44	38	5	2020/6/4 20:00:29
14231253907	握起来很有手感，也很大气，白色很好看，照相	2020/6/1 0:11:28	63	14	5	2020/6/3 12:35:56
14242684503	外形外观：白色的 后面是玻璃面 手感好 又好看	2020/6/1 0:00:11	43	13	5	2020/6/5 19:30:28
14250370998	收到手机立即激活，很便捷，原手机所有数据自	2020/6/3 21:45:09	34	22	5	2020/6/7 13:55:50
14675112126	是在京东下单，送了一年延保，可以保修2年，.	2020/8/28 0:40:18	2	6	5	2020/9/23 14:37:51
14549791614	苹果11紫色外观时尚惊艳，超级耐看，LCD高清.	2020/8/18 0:02:39	7	6	5	2020/8/19 23:46:38
14215214453	外形外观：外观很好看 很时尚 比较精致屏幕音	2020/5/6 0:14:15	27	11	5	2020/5/30 20:28:22
14015815509	一直喜欢用苹果手机，外观设计精美，声音柔软	2020/4/7 20:31:50	289	62	5	2020/4/9 22:26:47
14281428347	在Xs，11和11pro中选了很久，最后还是买了11，	2020/6/6 22:10:55	19	16	5	2020/6/14 10:13:55
14338019912	外形外观：外形，没的说，好看，颜色奶黄色屏	2020/6/1 0:01:11	35	18	5	2020/6/24 17:01:28
14517841399	很细买了，准备用个几年。因为之前一直用的小	2020/8/8 10:28:32	4	5	5	2020/8/11 9:26:31
14168077304	待机时间：一天屏幕音效：安卓转来，音效简直	2020/5/6 0:08:36	75	19	5	2020/5/18 19:36:55
13617927026	外形外观：买了黑色款，经典颜色，手感握持度	2019/12/12 1:16:15	208	74	5	2019/12/18 15:59:52
14614582126	本来想等12出来再换的可是手机的手机实在撑不	2020/8/30 17:17:05	8	7	5	2020/9/7 10:16:48
14168618134	外形外观：？？?赶上活动非常便宜的价格入手的	2020/3/28 13:45:21	79	22	5	2020/5/18 21:54:28
14545913807	外观好看漂亮，外形设计精美大气上档次，而且	2020/8/18 9:25:13	2	1	5	2020/8/18 22:24:21
14234556412	这次在京东小当家直播间买的手机，送延保一年	2020/6/1 1:22:25	18	14	5	2020/6/4 0:55:03
14599179437	物流非常快，一天不到就收到了。屏幕音效：屏	2020/9/1 16:11:31	8	5	5	2020/9/2 22:55:13

図 5　クローラされた投稿データの一例

データをクローラする処理とは，プログラムを使って人間のようにウェブページを閲覧し，目的のコンテンツを指定された場所に記録することをシミュレートすることである．クローラのカギは，特定のコンテンツ（投稿の特徴）を抽出し，次のページへのクリックをシミュレートし，アンチクローラを回避することである．プログラムでは，特定のコンテンツを抽出する際に，特定のフィールドにマッチする正規表現を用いる．次のページをシミュレートするには，ページカウント変数をループして，訪問ごとにページカウントに 1 を追加する．本研究では，Python 3 で投稿をクローラする．図 6 にクローラのコードを示す．

```python
from fake_useragent import UserAgent
from bs4 import BeautifulSoup
import time
import numpy as np
import requests
import json
import csv
import io

def commentSave(list_comment):
    file = io.open('data/JDComment_data2.csv','w',encoding="utf-8",newline = '')
    writer = csv.writer(file)
    writer.writerow(['ユーザID','評価本文','購買時間','有用といわれ','フィードバック','スコア','投稿時間','カテゴリ'])
    for i in range(len(list_comment)):
        writer.writerow(list_comment[i])
    file.close()
    print('導入しました！')

def getCommentData(maxPage):
    sig_comment = []
    global list_comment
    cur_page = 0
    while cur_page < maxPage:
        cur_page += 1
        url = 'https://club.jd.com/comment/productPageComments.action?callback=fetchJSON_comment98%s&score=%s&sortType=5&page=%s&pageSize=10&isShadowSku=0&fold=1'
              %(proc,i,cur_page)
        try:
            response = requests.get(url=url, headers=headers)
            time.sleep(np.random.rand()*2)
            jsonData = response.text
            startLoc = jsonData.find('{')
            #print(jsonData[::-1])//文字列逆順
            jsonData = jsonData[startLoc:-2]
            jsonData = json.loads(jsonData)
            pageLen = len(jsonData['comments'])
            print("当前第%s页"%cur_page)
            for j in range(0,pageLen):
                userId = jsonData['comments'][j]['id']#ユーザID
                content = jsonData['comments'][j]['content']#評価本文
                boughtTime = jsonData['comments'][j]['referenceTime']#購買時間
                voteCount = jsonData['comments'][j]['usefulVoteCount']#有用といわれ
                replyCount = jsonData['comments'][j]['replyCount']#フィードバック
                starStep = jsonData['comments'][j]['score']#スコア
                creationTime = jsonData['comments'][j]['creationTime']#投稿時間
                referenceName = jsonData['comments'][j]['referenceName']#カテゴリ
                sig_comment.append(userId)#各行のデータ
                sig_comment.append(content)
                sig_comment.append(boughtTime)
                sig_comment.append(voteCount)
                sig_comment.append(replyCount)
                sig_comment.append(starStep)
                sig_comment.append(creationTime)
                sig_comment.append(referenceName)
                list_comment.append(sig_comment)
                sig_comment = []

        except:
            time.sleep(5)
            cur_page -= 1
            print('ネットがつながらなくて、しばらくお待ちください。')
    return list_comment

if __name__ == "__main__":
    global list_comment
    ua=UserAgent()
    # headers={"User-Agent":ua.random}
    headers = {
    'Accept': '*/*',
    'Content-Type':'application/json; charset=utf-8',
    "User-Agent":'Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10_5_8 rv:2.0; ku-TR)AppleWebKit/531.15.1 (KHTML, like Gecko) Version/5.0 Safari/531.15.1',
    'Referer':"https://item.jd.com/100008348542.html#comment"
    }
    #iPhone 11に対して4種類カテゴリのID番号
    productid = ['&productId=100008348542','&productId=100008348530','&productId=100004770263',
                 '&productId=100008348534','&productId=100008348550']
    list_comment = [[]]
    sig_comment = []
    for proc in productid:#各色
        i = -1
        while i < 7:
            i += 1
            if(i == 6):
                continue
             #第0ページから訪問して、最大のページ数を獲得する。それから、繰り返してクローラする。
            url = 'https://club.jd.com/comment/productPageComments.action?callback=fetchJSON_comment98%s&score=%s&sortType=5&page=%s&pageSize=10&isShadowSku=0&fold=1'
                  %(proc,i,0)
            print(url)
            try:
                response = requests.get(url=url, headers=headers)
                jsonData = response.text
                startLoc = jsonData.find('{')
                jsonData = jsonData[startLoc:-2]
                jsonData = json.loads(jsonData)
                print("最大页数%s"%jsonData['maxPage'])
                getCommentData(jsonData['maxPage'])#各ページ
            except:
                i -= 1
                print("wating---")
                time.sleep(5)
                #commentSave(list_comment)
    print("クローラして終わり、導入し始めます-------")
    commentSave(list_comment)
```

図 6　Python 3 で JD の商品をクローラするコード

3.2 投稿データの前処理

データの前処理は，データを効果的かつ効率的に利用する，つまりモデルにうまくフィットして比較的良い結果を得るために，処理をする前の準備処理である．現実世界のデータは各種のデータがあり，そのまま利用することは難しい．モデルをうまく適用するため，データを洗練する前処理ステップが必要になる．

本研究では，構造化されていないテキストデータが含まれているため，従来のデータ処理方法とは異なる前処理を行う．本論文の投稿データには，テキストデータと非テキストデータの両方が含まれているため，データの種類ごとに別々に前処理を行う必要がある．

3.2.1 非テキストデータの前処理

以下に，本研究の実験中に行われた非テキストデータの前処理工程を示す．

（1）欠落しているデータ

一部の消費者はオンラインで購入した後にスコアを与えず，コメントを書かない消費者もいる．評価スコアは消費者に対して，投稿を読む際に重要な参考である．本研究では，スコアがない場合には中央値を用いる．具体的には，まず投稿データから評価スコアに基づいて投稿データを大きさ順にソートし，ランキングの中間位置にあるスコア値を欠落項目のフィラー値として用いる．コメントが書かれていないデータについては，コメントが書かれていなければ分析の意味がないので無視している．

（2）値の正規化

各属性の値の範囲が異なるため，大きなデータ値が大きな重みを持ち，小さなデータ値の影響を示さないことを避けるために，様々な属性のデータを 0.0～1.0 の間で正規化する．大きなデータが結果にあまり影響を与えないようにするために，データ規定を採用する．本論文で使用した量の入力方法は，ゼロ平均入力である．ゼロ平均入力は，データセットの各値からデータセットの平均を差し引き，それをデータセットの標準偏差で割ることによって求める．

（3）次元の正規化

テキストデータを処理する場合，テキストデータをベクトル化する必要がある．ベクトル化後の次元が非常に大きく（辞書サイズとほぼ同じ）なる．次元が大きい場合，処理の効率やスピードに大きな影響を与える．そのため，次元削減が必要となる．本研究では，特異値分解という手法を用いて次元削減を行う．これをデータ圧縮と呼ぶこともある．

（4）タプル重複

コメントのすべてのデータ項目の重複と，テキストの内容だけの重複の 2 つのケースがある．本研究では，テキストの内容が重複している限り，タプルの重複を考慮する．同じ商品のレビューは違う消費者が書いているが，投稿者が面倒だと感じ，他の消費者が書いた投稿をコピーすることもある．このような投稿については，本研究では重複排除（削除）の方法で対応する．これは，先章に触れ，複製された投稿を特定することである．本稿では，このような特定のやらせ投稿には触れていない．

3.2.2 テキストデータの前処理

本論文では，テキストデータの前処理として，単語分離，ストップワードの除去，テキストベクトル化，学習データのアノテーションを行う．以下に，詳細について述べる．

（1）単語分離

中国語は英語や他の言語と異なり，単語を区別するための単語間のスペースのような明確なマークがない．また，中国語では個々の単語が実物を表すために使われても意味が明確ではないため，一般的に単語は中国語の処理の最小単位である．単語分離とは，「この商品を好きで支持する」という分詞の後に，例えば「この　商品　を　すき　で　支持する」というように，文章やフレーズを単語に分割したものである．文やフレーズを単語に分割することで，文の意味を理解しやすくなり，文を数学的なベクトルに形式化することを避けることができる．また，その単語を使って辞書を構築でき，特定の用途に非常に効果的である．

本論文では，オープンソースの「jieba」という単語分割ツールを用いて単語の分割を行っている．「jieba」を用いる理由の 1 つは Python で操作できる点であ

る．この論文のデータ処理はすべて Python で使用し，全体のプロセスに単語の分離を統合する際に便利である．また，単語のマークアップに対応しており，ストップワードを削除するためにストップワードリストを追加することができる．

投稿テキストの単語を分割した結果を図 7 に示す．

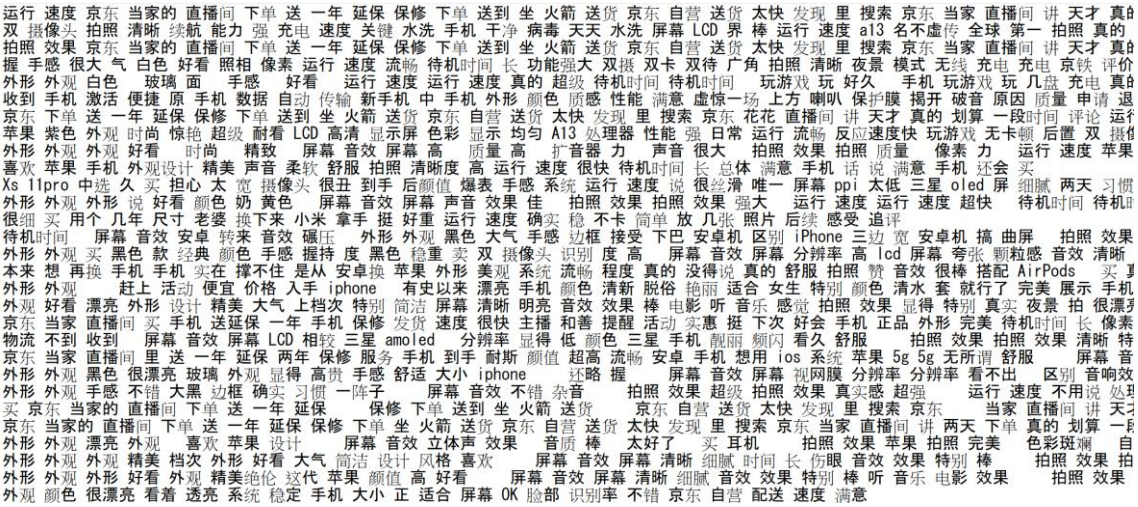

図 7　投稿テキストの単語を分割した結果の一例

（2）ストップワードの除去

ストップワードは，一般的に，コンマや感嘆符など，明確な意味を持たない言葉や記号のことを指す．ストップワードを除去する目的は，記憶容量を節約して処理速度を上げ，データ辞書の次元を小さくすることである．本論文では，単語分離と一緒に扱われ，同じ「jieba」を使用している．単語分離後にその単語がストップワードである場合，分割した結果には追加しない．ストップワードを除去するカギは，より良い，より完全なストップワードのリストを見つけることである．本稿で使用しているのは，ウェブ上で見つかったリスト「stopwords.txt」である．このリストに含まれるストップワードの数は比較的多く，完全性の観点からこのストップワードリストを使用することにした．図 8 にストップワードリストのサンプルを示す．図 9 に単語分離とストップワードの除去のコードを示す．

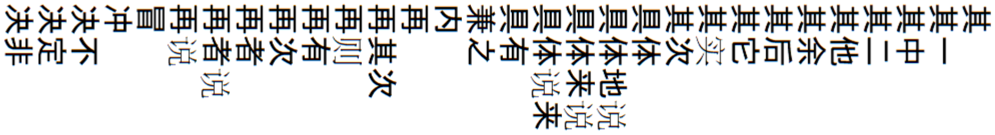

図 8　ストップワードリストの一例

```
#encoding=utf-8

import jieba

#stopwords list
def stopwordslist():
    stopwords = [line.strip() for line in open('stopwords.txt',encoding='UTF-8').readlines()]
    return stopwords

#単語分離
def seg_depart(sentence):
    #各投稿を中国語の単語に基づいて分割する
    sentence_depart = jieba.cut(sentence.strip())
    #ストップワードリストを設定する
    stopwords = stopwordslist()
    #outstrを導入する
    outstr = ''
    #ストップワードを除去する
    for word in sentence_depart:
        if word not in stopwords:
            if word !='\t':
                outstr += word
                outstr += " "
    return outstr

#text path
filename = "JDSpiderComment.txt"
outfilename = "word_cut_stop.txt"
inputs = open(filename,'r',encoding='UTF-8')
outputs = open(outfilename,'w',encoding='UTF-8')

#txtを導出する
for line in inputs:
    line_seg = seg_depart(line)
    outputs.write(line_seg + '\n')
    print("--------------操作中----------------")
outputs.close()
inputs.close()
print("導出しました！")
```

図 9　単語分離とストップワード除去のコード

（3）テキストベクトル化

ベクトル化の目的は，コンピュータの理解と処理を容易にすることである．自然言語処理における人工知能を実現するためには，機械が特定の表現の意味を知るため，各テキストがどのように表現されているかを理解する必要がある．テキストベクトル化は様々な表現の一つである．本論文では,「TF-IDF」を使用して評価本文の中で各単語の重みとキーワードを抽出した．使用しているツールは「sklearn」である．

TF (Term Frequency) の考え方は，ある単語が文書（本論文では一条の投稿）の中でより頻繁に出現する場合，その単語はより大きな重みを持っていると考えられる．しかし，重要でない単語が文書中に複数回出現した場合，TF 法ではその単語に大きな重みを与えることになり，これは不合理であるため，IDF（Inverse Document Frequency）を用いている．IDF の考え方は，ある単語が複数の文書（本論文では複数の投稿）に現れた場合，その単語の重みが低いと考えられるというものである．しかし，重要な言葉が複数の文書の中で繰り返されることもある．以上の 2 点を踏まえて，TF と IDF の長所と短所を組み合わせて，ワードの重みを計算する TF-IDF 法を使用した．

TF-IDF 法により，200 個のキーワードが抽出せれ，図 10 のように示す．TF-IDF を用いて生成した初期行列の行ベクトルは図 11 に示す．

	词	总词频数	词性	词频
0	很/d	1129	d	很
1	好/a	1099	a	好
2	拍照/v	1020	v	拍照
3	很/zg	1016	zg	很
4	速度/n	993	n	速度
5	手机/n	945	n	手机
6	运行/v	914	v	运行
7	效果/n	913	n	效果
8	屏幕/n	866	n	屏幕
9	非常/d	865	d	非常
10	外观/n	699	n	外观
11	苹果/n	693	n	苹果
12	音效/n	674	n	音效
13	不错/a	655	a	不错
14	待机时间/	635	n	待机时间
15	喜欢/v	564	v	喜欢
16	快/a	560	a	快
17	外形/n	554	n	外形
18	都/d	543	d	都
19	特别/d	518	d	特别
20	不/d	446	d	不
21	买/v	406	v	买
22	好看/v	374	v	好看
23	京东/ns	373	ns	京东
24	清晰/a	371	a	清晰
25	还/d	369	d	还
26	手感/n	323	n	手感

	0	1
0	拍照	0.243324
1	待机时间	0.227956
2	音效	0.213625
3	屏幕	0.196834
4	手机	0.145521
5	速度	0.142879
6	效果	0.14091
7	外观	0.140891
8	苹果	0.134644
9	运行	0.134302
10	非常	0.109533
11	外形	0.106614
12	不错	0.104677
13	京东	0.094209
14	喜欢	0.08306
15	手感	0.076017
16	好看	0.073565
17	清晰	0.066335
18	流畅	0.065889
19	特别	0.06339
20	特色	0.049554
21	颜色	0.049143
22	很快	0.038991
23	感觉	0.038317
24	摄像头	0.036155
25	满意	0.035321
26	很棒	0.035061

図 10　TF-IDF による計算した各単語の出現回数と重みの一例

Column2	Column1644	Column1645	Column1646	Column1647	Column1648	Column1649	Column1650
0	小时	小朋友	小灵通	小点	小牛	小白	小米
运行速度：是在京东小当家	0.0	0.0	0.0	0.0	0.0	0.0	0.0
双摄像头的拍照清晰，续航	0.17157682144739003	0.0	0.0	0.0	0.0	0.0	0.0
拍照效果：是在京东小当家	0.0	0.0	0.0	0.0	0.0	0.0	0.0
握起来很有手感，也很大气	0.0	0.0	0.0	0.0	0.0	0.0	0.0
外形外观：白色的 后面是玻	0.0	0.0	0.0	0.0	0.0	0.0	0.0
收到手机立即激活，很便捷	0.0	0.0	0.0	0.0	0.0	0.0	0.0
是在京东下单，送了一年延	0.0	0.0	0.0	0.0	0.0	0.0	0.0
苹果11紫色外观时尚惊艳，	0.0	0.0	0.0	0.0	0.0	0.0	0.0
外形外观：外观很好看 很时	0.0	0.0	0.0	0.0	0.0	0.0	0.0
一直喜欢用苹果手机，外观	0.0	0.0	0.0	0.0	0.0	0.0	0.0
在Xs，11和11pro中选了很	0.0	0.0	0.0	0.0	0.0	0.0	0.0
外形外观：外形，没的说，	0.0	0.0	0.0	0.0	0.0	0.0	0.0
很细买了，准备用个几年。	0.0	0.0	0.0	0.0	0.0	0.0	0.25876163914348516
待机时间：一天 屏幕音效：	0.0	0.0	0.0	0.0	0.0	0.0	0.0
外形外观：买了黑色款，丝	0.0	0.0	0.0	0.0	0.0	0.0	0.0
本来想等12出来再换的可是	0.0	0.0	0.0	0.0	0.0	0.0	0.0
外形外观：？？赶上活动非	0.10227458139123843	0.0	0.0	0.0	0.0	0.0	0.0
外观好看漂亮，外形设计精	0.0	0.0	0.0	0.0	0.0	0.0	0.0
这次在京东小当家直播间买	0.0	0.0	0.0	0.0	0.0	0.0	0.0
物流非常快，一天不到就收	0.0	0.0	0.0	0.0	0.0	0.0	0.0
在京东小当家直播间里下的	0.0	0.0	0.0	0.0	0.0	0.0	0.0
外形外观：黑色很漂亮，玻	0.0	0.0	0.0	0.0	0.0	0.0	0.0
外形外观：手感不错，大黑	0.0950102122727004	0.0	0.0	0.0	0.0	0.0	0.0

図 11　TF-IDF によるテキストをベクトル化した結果の一例

このようにして，テキストデータはベクトルの形で表現される．本論文では，この処理を「sklearn」を使って行っている．「sklearn」は，TF-IDF 法を用いて学習前にテキストをベクトル化することができる統合型機械学習 Python ライブラリです．また，「sklearn」には多数の機械学習モデルが含まれ，直接モデル API を呼び出して訓練し，その後テストに適用することもできる．テキストのベクトル化が完了すると，分類器を訓練するようなデータとして使用することができる．テキストベクトル化のコードは図 1 2 に示す．

```
import pandas as pd
import numpy as np
from sklearn.feature_extraction.text import CountVectorizer,TfidfTransformer
import jieba
from pylab import *
mpl.rcParams['font.sans-serif'] = ['SimHei']
plt.rcParams['font.sans-serif'] = ['SimHei']
plt.rcParams['axes.unicode_minus'] = False


def data_to_cu(x_sorce):
    contentlist = []
    for j in x_sorce:
        text = re.sub("[\s+\.\!\/_,$%^*(+\"\']+|[+——！，。？、~@#￥%……&*（）]+|[\d+]|[a-zA-Z+]|⊙|∀|：", "", j)
        current = []
        for word in jieba.cut(text):
            if word not in stopword and word not in current and word != ' ':
                current.append(word)
        content = ' '.join(current)
        contentlist.append(content)
    return contentlist

file = open("stopwords.txt",encoding='utf8')
stopword = [line.lstrip().rstrip() for line in file.readlines()]
file.close()
df = pd.read_csv('JDSpiderComment1.txt',error_bad_lines=False,header=None)
df['len'] = df[0].apply(lambda x : get_count(x))

x_sorce = df[0].astype(str).values
contentlist = data_to_cu(x_sorce)
counter = CountVectorizer()
counts = counter.fit_transform(contentlist)
tfidfer = TfidfTransformer()
x = tfidfer.fit_transform(counts)
word=counter.get_feature_names()
df1 = pd.DataFrame(x.toarray())
df1.columns = word
df__1 = df.join(df1)
df__1.to_csv('結果TFIDF.csv')
```

図 12　テキストベクトル化のコード

（4）学習データのアノテーション

分類器を訓練する上でより重要なことは，訓練データのラベル付けである．学習データにラベルを付ける目的は，分類器にデータサンプルがどのような種類のものであるか，あるいはどのくらいの測定値であるかを伝えることである．例えば，投稿を真か偽かのラベル付けは，一般的に真なら 1，偽なら 0 で行われる．本論文では，一般的な教師あり学習アプローチとは異なり，データセットにラベルを付けて分類器を訓練する．データのラベル付けの際に，訓練機にサンプルがどのクラス（真または偽）に属しているかを教えるのではなく，そのサンプルの尺度が何であるかを教えることを設定する．本論文のネットショッピングサイトでの投稿の場合，投稿が真か偽かという明確な証拠がないため，単純にサンプルが真か偽かを訓練機に伝えるのではなく，サンプルの信頼程度を訓練機に伝える訓練を行う．アノテーションされた訓練データを図 13 に示す．

Percent	Delay	Length	Keywords	Useful	Feedback	Score
125	5.836851852	71.6790362	74	20.96095366	12.38419216	5
65	2.768055556	3.969388245	26	42.11691695	33.49091292	5
110	3.733773148	71.6790362	74	10.22509168	23.6411099	5
120	2.516990741	67.27233186	76	16.22454396	6.755733288	5
40	4.812696759	5.438289691	28	9.909331037	6.05217593	5
40	3.674085648	2.255669892	35	7.067485221	12.38419216	5
125	26.58163194	66.29306423	74	3.036855459	1.127274418	5
45	1.988877315	1.906217537	44	1.458052227	1.127274418	5
55	24.84313657	4.948655875	30	4.857160697	4.645061212	5
100	2.079826389	16.94468435	17	87.58645001	40.52648651	5
50	7.502083333	21.00193633	40	2.331075527	8.162848006	5
60	23.70853009	4.109569705	36	7.383245867	9.569962723	5
20	2.95693287	15.72059981	18	2.405334166	0.42371706	5
120	12.81133102	43.76990874	72	20.01367172	10.27352008	5
130	6.613622685	40.09765512	51	62.00983767	48.96917481	5
30	7.708136574	15.96541672	21	1.142291581	1.830831777	5
165	51.33966435	55.76593721	75	21.2767143	12.38419216	5
30	0.541064815	7.396824951	33	3.036855459	2.390512375	5
50	2.98099537	17.57449962	48	2.015314881	6.755733288	5

図 13　アノテーションされた訓練データの一例

ここで，Delay は投稿を遅延させる日数，あるいは購買時間と投稿時間の時間差を表す．Length は投稿テキストの長さ，Keywords は投稿テキストの中で商品に関連するキーワード（感情を表す単語や商品に関する属性ワードなど）の数，Useful は「この投稿が有用」とクリックされた数，Feedback は他の消費者がこの投稿を見る時にフィードバックを書く数，Score は投稿者がこの商品に対する評価スコアを表す．Percent はネットショッピングする際に投稿を注目する 30 名の消費者から，手動で採点した投稿の信頼度スコアを表す．

3.3 第 3 章のまとめ

本章では，3.1 節では訓練用の投稿データの獲得について，3.2 節では投稿データ（非テキストデータとテキストデータ）をそれぞれ前処理することの詳細について述べた．次章では，前処理されたデータを適用して信頼値モデルを構築することについて述べる．

第 4 章　信頼値モデルの提案

4.1 やらせ投稿特徴の選択

特徴選択の目的は，分類器を訓練するためにモデルに適したデータを選択することである．特徴の選択には 2 つの主要な出発点がある．1 つ目は，特徴をベクトル化する際に，特にテキストデータの場合は次元を下げることである．この操作を行わないと特徴のベクトルが辞書サイズに達し，分類器の訓練が困難になる．2 つ目は，データを絞り込むことである．特徴が多すぎるとモデルの複雑さが増加し，訓練するときにオーバーフィットし易くなり，訓練された分類器を新しいサンプルに適用することは難しくなる．そのため，特徴の選択が必要になる．本稿で使用する特徴選択の視点は，明示的特徴と暗黙的特徴である．

4.1.1 明示的特徴の選択

明示的特徴とは，投稿での評価スコアのように，明示的に表現できる特徴やデータ項目のことである．識別された特徴を直接選択して訓練に使用できる．一般的に使用されている明示的特徴選択法には，情報利得，利得率，ジニ係数などがある．本論文では，回帰分析による選択した明示的な特徴は以下の通りである．

（1）投稿の遅延時間（購買時間と投稿時間の時間差）：
投稿の遅延時間が短いほど信頼性が低い．商品購入後に，配達と使用することが時間をかかるため，遅延時間が長いほど投稿者が投稿に要する時間が長く，投稿の信頼性が高くなる．

（2）関連性（キーワード数）：
投稿の商品に関する単語や感情を表す単語の数が少ないほど，投稿の信頼性は低くなる．実際の消費者によって書かれた投稿は，通常，製品の特定の側面に関する内容や使用した感情を表す内容などがある．やらせ投稿は，多くの場合，関連性がほとんどなく，または全く関係なし内容を書かれている．

（3）投稿テキストの長さ：

投稿テキストの長さが短いほど，その投稿の信頼値は低くなる．投稿者は，不真面目で早く投稿しようとしているので，文章の長さが短くなると考える．現在のコピーされたやらせ投稿を検出する技術は高度化され[23]，長く貼り付けられた投稿は検出される可能性が高い．そのため，やらせ投稿を隠して素早く書くために，投稿者は比較的短いコメントを書くことが多い．

（4）有用といわれた数：

有用といわれた数が少ないほど，投稿の信頼性は低くなる．この数は，ほかの消費者が商品投稿を見るとき参考になる．「この投稿が有用」とクリックされた数であり，大きいほど，多くの人がこの投稿が信頼できると判断することを表す．

以上のような特徴を選択した方法は，まず，第二章のように前処理したすべての項目を特徴として使用する．投稿の遅延時間 D(Delay)，評価スコア S(Score) は投稿者に関する明示的特徴である．投稿のテキストの長さ L（Length），キーワード数 K（Keywords）は投稿本文に関する明示的特徴である．「この投稿が有用」とクリックされた数 U（Useful），フィードバックの数 F（Feedback）は投稿を見る消費者に関する明示的特徴である．以下のように，6 つの特徴を融合させ，信頼値 P を求める．

$$P(r) = \lambda_1 \times D(r) + \lambda_2 \times S(r) + \lambda_3 \times L(r) + \lambda_4 \times K(r) + \lambda_5 \times U(r) + \lambda_6 \times F(r) + b$$

ここで，r は投稿である．D，S，L，U，F の値については，実際の数字に基づいている．D∈(0,80)，L∈(1,500)，U∈(0,800)，F∈(0,800)である．より大きなデータでは小さいデータが見えないという役割を回避するために，L の値を正規化して(1,100)になり，U の値を正規化して(0,200)になり，F の値を正規化して(0,200)になる．C は TF-IDF による結果に応じて設定する．

次に，上記の 6 つの明示的な特徴を理解しており，ネットショッピングする際に投稿を注目する 30 人に投稿を読んで，その信頼値を評価してもらい，P の値を付け（P の範囲が(0,300)に設定すること），最後に各項目と信頼値との相関関係を行う．そして，D，L，K，U が有意で保留した．S と F は線形関係がな

し，排除した．

明示的特徴を最終に，投稿の遅延時間 D（Delay），投稿のテキストの長さ L（Length），キーワード数 K（Keywords），「この投稿が有用」とクリックされた数 U（Useful）という 4 つの項目を決定する．

4.1.2 暗黙的特徴の選択

明示的特徴選択とは異なり，暗黙的特徴の選択は，特徴を明示的に述べることができず，解釈の幅が狭くなる．暗黙的特徴選択は，通常，データの次元を低減する過程で実現される．本論文では，特異値分解（Singular Value Decomposition, SVD）という方法を主に用いている．

特異値分解（SVD）は，最初が情報検索に使われていた．これは行列の別の表現であり，その目的は行列 M を 3 つの行列として表現することである．

$$M = U\Sigma V^*$$

M を階数 r の m 行 n 列の行列とする． U は m 行 m 列のユニタリ行列， Σ は半正定値 m 行 n 列の対角行列，V*は n 行 n 列のユニタリ行列 V の随伴行列（複素共役かつ転置行列）．Σ の対角線上の要素 Σ_i，ここで Σ_i は M の特異値である．SVD のコードは図 1 4 に示す．

```
from numpy import *
from numpy import linalg as la

my Mat = mat(weight)

def SVDapp():
#SVD 特異値分解の部分
    U,Sigma,VT = la.svd(myMat) #TF-IDFによる形成した行列が三つの行列に分ける。
    my Mat1=mat(U)
    my Mat2=mat(Sigma)
    my Mat3=mat(VT)

    #三つの行列の行と列の数を見る
    print "U : ",myMat1.shape
    print "Sigma : ",myMat2.shape
    print "VT : ",myMat3.shape
    Sig2 = Sigma**2

    for LineNum in range(len(Sig2)): #特異値がいくつになるとすべての項目の90%以上が表せるかについての計算
        if sum(Sig2[:LineNum])>sum(Sig2)*0.9:
            break

    print "元の90%が到達して、行列を保留する行と列の数: ",LineNum
    print 'SVD 特異値分解が完成しました'
    return U,Sigma,VT,LineNum

U,Sigma,VT,LineNum = SVDapp()
```

図 14　特異値分解（SVD）のコード

初期行列の行ベクトルは TF-IDF を用いて生成され，各行は投稿テキストに対応し，異なる列は異なるキーワードに対応する．初期行列$A_{m\times n}$は，SVD を用いて$U_{m\times m}$，$S_{1\times m}$，$V_{n\times n}$に分解され，ここで，$U_{m\times m}$の行は投稿テキスト，列はキーワードに対応し，$V_{n\times n}$の行は投稿，列はキーワードに対応する．$U_{m\times m}$行列を使用し，最も信頼性の高い投稿テキストは計算され，手動で識別され，その後，その投稿に対する他の投稿の類似度が計算される．類似度を計算することに基づいて信頼度の指標を補正している．最後に，明示的特徴と暗黙的特徴を組み合わせて平均化し，投稿の信頼性を得る．暗黙的特徴の選択によって，人間の主観や特徴の不完全な問題は回避される．

4.2 モデル訓練

データと特徴量が決まれば，次はデータ訓練の決定であり，データの訓練方法は実験データにある程度基づいて行われる．投稿の真偽についての決定的な証拠がほとんどなく，すなわち，投稿データにラベルを付ける明白な特徴がないため，やらせ投稿を特定するために用いられる方法は様々である．主に機械学習に関連するアルゴリズムや言語モデルに関連する方法が用いられている．やらせ投稿の特徴によって識別する可能がある．そのため，これらの特徴は教師あり学習法に使用機会を与え，それゆえ，やらせ投稿を特定するために教師あり機械学習アルゴリズムを使用する多くの学者がいる．Jindal N らは，投稿，店舗，投稿者で構成される特徴セットに対してロジスティック回帰を使用し，やらせ投稿を特定した．ロジスティック回帰の他に，サポートベクターマシンやベイズ分類も用いたが，どちらもロジスティック回帰ほどの効果はなかった[6]．そして，本論文では，やらせ投稿を検出する方法として，回帰分析を用いて信頼値という尺度を適合させる．

統計学で，回帰分析とは，複数の要素や変数の関係性を分析するために用いられる手法である．この論文では，データがサンプル $x=(x_1;x_2;\cdots;x_d)$ を記述する d 個の特徴を持っていることを考えると，重回帰を使用する．ここで x_i は i 番目の特徴量の x のメジャーである．重回帰モデル（linear model）の目的は，複数の変数間の関係を特徴づけるために使用される関数，すなわち

$$f(x) = w_1x_1 + w_2x_2 + \ldots + w_dx_d + b$$

ベクトル形式では，次のように書くことができます.

$$f(x) = w^T x + b$$

ここで，w=(w1;w2;…;wd)．w と b が学習後に，モデルが決定できる.

回帰分析は，現実の状況をうまく描写するためだけでなく，視覚的な顕在化を容易にし，受け入れやすく理解しやすいようにするために，現実に広く使われている．本論文では，手動でラベル付けされたデータに対して重回帰モデルを学習するために最小二乗法を用いている．最小二乗の考え方は，ラベル付けされた学習データを未知の関数でフィットさせ，その誤差の二乗和を最小化することでフィットの効果を高めることで，関数内の位置変数を求め，フィットの関数式を決定するというものである．関数が決定された後，未知の変数に対応する関数の値を予測することができ，異なる状況を組み合わせることで所望の結果を得ることができる．信頼値公式を取得した後，未知のレビューの信頼値を予測するために使用することができる．4.1 節で説明していた明示的特徴が非常に明らかな 300 件の投稿を選択し，信頼値公式を計算する．信頼値 P が以下のようになる.

$$P(r) = 1.002 \times D(r) + 0.337 \times L(r) + 1.172 \times K(r) + 0.716 \times U(r) - 10.402$$

この公式を用い，すべての投稿の信頼値を予測し，最終的に信頼値をランキングする．信頼値フィットの実験結果を図 15 に示す．ここで，横軸は特徴値，縦軸は信頼値であり，赤線は 4.1 節で言った 30 人が採点した信頼値の線，青線が公式を用いて算出した信頼値の線を表す．信頼値フィットのコードは図 16 に示す.

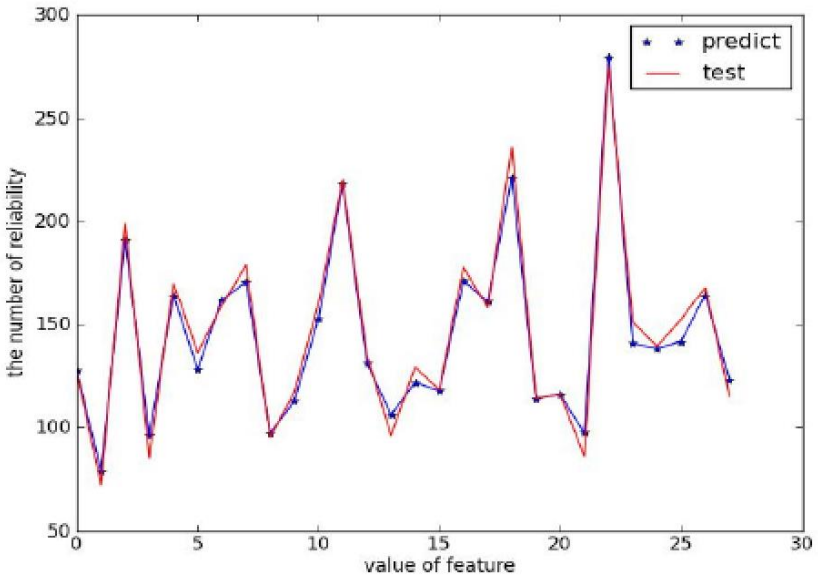

図 15　信頼性フィットの実験結果

```
#coding=utf-8

import pandas as pd
from sklearn.cross_validation import train_test_split
from sklearn.linear_model import LinearRegression
import matplotlib.pyplot as plt
from sklearn import metrics
import numpy as np

sum_mean = 0
data = pd.read_csv("JDcommentTest.csv")
feature_cols = ['Delay','Length','Keywords','Useful']

X = data[feature_cols]
y = data['Percent']
X_train,X_test,y_train,y_test = train_test_split(X,y,random_state=1)

linreg = LinearRegression()
model = linreg.fit(X_train,y_train)

print "モデル：",model
print "切片：",linreg.intercept_
print "係数：",linreg.coef_
print "切片、係数：",zip(feature_cols,linreg.coef_)
    y_pred = linreg.predict(X_test)
print "予測：",y_pred

for i in range(len(y_pred)):
    sum_mean+=(y_pred[i]-y_test.values[i]**2)
    sum_erri = np.sqrt(sum_mean/50)
    print "RMSE by hand:",sum_mean

plt.figure()
plt.plot(range(len(y_pred)),y_pred,'*',label="predict")
plt.plot(range(len(y_pred)),y_pred,'b')
plt.plot(range(len(y_pred)),y_test,'r',label="test")
plt.legend(loc="upper right")
plt.xlabel("value of feature")
plt.ylabel('the number of reliability')
plt.show()
```

図 16　回帰分析による信頼性フィットのコード

4.3 モデル評価

4.3.1 転位レベル評価方法

実験方法（順位付け）の特殊性から，本稿では手動採点を当てはめた上で，実験結果を順位付けして比較するというアプローチをとっている．手動スコアリングと比較しない理由は，やらせ投稿には明らかな特定の特徴がなく，データをラベル化した際に信頼値公式に適合するように比較的明らかな特徴を持つ投稿がいくつか選択されただけである．これらの比較的明らかな特徴については 4.1 節で説明していた．比較的多くのデータを持つため，全てのデータを対象に実験を行った．転位レベルをカウントすることをコメント数と組み合わせることで，精度を計算している．本研究では最大の転位レベルが 8 であるため，転位レベルが 8 に分かれている．

本論文における転位レベルの意味は，手作業で採点して投稿の信用度をフィッティングした後で順番付けの信頼度位置と，暗黙的特徴に基づいて定まった類似度を乗算した後の信頼度位置の変化の大きさである．類似度を乗算する理由は，フィットした信頼度測定値を暗黙的特徴から識別した信頼度類似度に縛り付けることができ，すなわち，各投稿の信頼度とその投稿テキストの類似度を乗算することで，フィットした信頼度比較が暗黙的特徴識別の精度の判定を容易にするためである．類似度は，手動判定基準に基づいて最も信頼性の高い投稿を算出し，他の投稿との類似度を求めることで算出される．

転位レベルが 1 レベルであれば，手動採点の信頼性位置と SVD 処理後の信頼性位置が 1 位変化したことを意味にする．例えば，手動採点後の投稿の信頼性位置が 10 番目，SVD 処理後の信頼性位置が 11 番目である場合，このような転位レベルを 1 レベルと呼び，転位レベルが大きいほど，その差は大きくなり，信用度位置の変化も大きくなると示す．統計の時では，異なる転位レベルに対応する投稿数を別々にカウントし，レベル 1 以上の転位レベルにはレベル 2 とレベル 2 以降の転位レベルが含まれ，同様にレベル 2 以上の転位レベルにはレベル 3 とレベル 3 以降の転位レベルが含まれている，というようにした．本研究では，実験データを 4 つのグループに分け，まず各グループの転位レベルに対応する投稿数を個別に計算し，正確率を求め，最終的に 4 つのグループの平均正確率を求め，全体の結果の評価を記述することができる．コードは図 17 に示す．

```
def Data_Count():
    P0 = []
    for i in range(len(D)):


P0.append(1.32*int(D[i])+1.1*int(M[i])+0.83*int(L[i])+1.29*int(T[i])-19.69)
    print "明示的特徴をフィットした信頼値が一番大きい値と、その行の数:",max(P0),P0.index(max(P0))
    k1 = [];k2 = [];k3 = []
    for i in range(len(P0)):
        k1.append([i,P0[i],S[i],R[i]]) #k1 信頼値と類似性のマッピング、投稿内容も貼り付け
    k2 = sorted(k1,key=lambda x:x[1],reverse=True) #k2 フィッティングした後で順番付けの信頼度位置の表
    km = []
    for m in range(len(P0)):
        km.append([k2[m][0],k2[m][1]*k2[m][2],k2[m][2],k2[m][3]])
    k3 = sorted(km,key=lambda x:x[1],reverse=True) #k3 類似度を乗算した後の順番付け、投稿内容も貼り付け
        #for j in range(len(P0)):
            #print k1[j],k2[j],k3[j] #k1,k2,k3 を同じ列で導出し、転位の状況が見える

    v1=0
    v2=0
    v3=0
    v4=0
    v5=0
    v6=0
    v7=0
    v8=0
    v9=0
    k22=[];k33=[]
    for p in range(1961):
        k22.append(k2[p][0])
        k33.append(k3[p][0])

    for n in range(1961):
        try:
            f=abs(k33.index(k22[n])-n)
        except Exception,e:
            pass
        if f==0:
            pass
        elif f==1:
            v1=v1+1
        elif f==2:
            v2=v2+1
        elif f==3:
            v3=v3+1
        elif f==4:
            v4=v4+1
        elif f==5:
            v5=v5+1
        elif f==6:
            v6=v6+1
        elif f==7:
            v7=v7+1
        elif f==8:
            v8=v8+1
        else:
            v9=v9+1

    print "レベル1の数: ",v1
    print "レベル2の数: ",v2
    print "レベル3の数: ",v3
    print "レベル4の数: ",v4
    print "レベル5の数: ",v5
    print "レベル6の数: ",v6
    print "レベル7の数: ",v7
    print "レベル8の数: ",v8
    print "レベル9の数: ",v9 #v1,v2,v3,v4,v5,v6,v7,v8,v9 は信用度位置の変化を表す。 1-8 和 8 以上の数。
    print "レベル1とそれ以上の数: ",v1+v2+v3+v4+v5+v6+v7+v8+v9,"正確率: ",1-(v1+v2+v3+v4+v5+v6+v7+v8+v9)/1961.0
    print "レベル2とそれ以上の数: ",v2+v3+v4+v5+v6+v7+v8+v9,"正確率: ",1-(v2+v3+v4+v5+v6+v7+v8+v9)/1961.0
    print "レベル3とそれ以上の数: ",v3+v4+v5+v6+v7+v8+v9,"正確率: ",1-(v3+v4+v5+v6+v7+v8+v9)/1961.0
    print "レベル4とそれ以上の数: ",v4+v5+v6+v7+v8+v9,"正確率: ",1-(v4+v5+v6+v7+v8+v9)/1961.0
    print "レベル5とそれ以上の数: ",v5+v6+v7+v8+v9,"正確率: ",1-(v5+v6+v7+v8+v9)/1961.0
    print "レベル6とそれ以上の数: ",v6+v7+v8+v9,"正確率: ",1-(v6+v7+v8+v9)/1961.0
    print "レベル7とそれ以上の数: ",v7+v8+v9,"正確率: ",1-(v7+v8+v9)/1961.0
    print "レベル8とそれ以上の数: ",v8+v9,"正確率: ",1-(v8+v9)/1961.0

Data_Count()
```

図 17　転位レベル評価方法の応用コード

4.3.2 転位レベル評価結果

最初から 500 件の投稿を処理した後の転位レベルに対応する投稿数を表 1 に示す.

表 1　最初 500 件の投稿に対して各転位レベルに対応する投稿数

レベル 1 とそれ以上の数	レベル 2 とそれ以上の数	レベル 3 とそれ以上の数	レベル 4 とそれ以上の数	レベル 5 とそれ以上の数	レベル 6 とそれ以上の数	レベル 7 とそれ以上の数	レベル 8 とそれ以上の数
290	154	71	43	22	13	10	5

上の表は，転位レベルとそれに対応する投稿数を示しており，その後，下記のような計算式に従って正確率を計算する.

$$R_d = 1 - \frac{\sum_{i=d}^{8} V_i}{N}$$

ここで，N は投稿数，d は転位レベルである．各転位レベルの正確率は，上記の式に従って算出される．表 2 のように示す.

表 2　最初 500 件の投稿に対して各転位レベルに対応する正確率

転位レベル	レベル 1	レベル 2	レベル 3	レベル 4	レベル 5	レベル 6	レベル 7	レベル 8
正確率	0.421	0.692	0.858	0.914	0.956	0.968	0.98	0.99

投稿数が最初から 1000 件になった場合の結果を表 3，表 4 に示す.

表 3　最初 1000 件の投稿に対して各転位レベルに対応する投稿数

レベル 1 とそれ以上の数	レベル 2 とそれ以上の数	レベル 3 とそれ以上の数	レベル 4 とそれ以上の数	レベル 5 とそれ以上の数	レベル 6 とそれ以上の数	レベル 7 とそれ以上の数	レベル 8 とそれ以上の数
696	451	271	177	115	71	45	27

表 4　最初 1000 件の投稿に対して各転位レベルに対応する正確率

転位レベル	レベル 1	レベル 2	レベル 3	レベル 4	レベル 5	レベル 6	レベル 7	レベル 8
正確率	0.304	0.549	0.729	0.823	0.885	0.929	0.955	0.973

投稿数が最初から 1500 件になった場合の結果を表 5，表 6 に示す.

表 5　最初 1500 件の投稿に対して各転位レベルに対応する投稿数

レベル 1 とそれ以上の数	レベル 2 とそれ以上の数	レベル 3 とそれ以上の数	レベル 4 とそれ以上の数	レベル 5 とそれ以上の数	レベル 6 とそれ以上の数	レベル 7 とそれ以上の数	レベル 8 とそれ以上の数
1111	721	458	302	199	123	80	50

表 6　最初 1500 件の投稿に対して各転位レベルに対応する正確率

転位レベル	レベル 1	レベル 2	レベル 3	レベル 4	レベル 5	レベル 6	レベル 7	レベル 8
正確率	0.259	0.519	0.695	0.799	0.867	0.918	0.947	0.967

投稿数が最初から 1961 件(すべての投稿データ)になった場合の結果を表 7, 表 8 に示す.

表 7　1961 件の投稿に対して各転位レベルに対応する投稿数

レベル 1 とそれ以上の数	レベル 2 とそれ以上の数	レベル 3 とそれ以上の数	レベル 4 とそれ以上の数	レベル 5 とそれ以上の数	レベル 6 とそれ以上の数	レベル 7 とそれ以上の数	レベル 8 とそれ以上の数
1429	910	568	365	239	151	99	56

表 8　1961 件の投稿に対して各転位レベルに対応する正確率

転位レベル	レベル 1	レベル 2	レベル 3	レベル 4	レベル 5	レベル 6	レベル 7	レベル 8
正確率	0.271	0.536	0.710	0.814	0.878	0.922	0.950	0.971

最初から 500，1000，1500，1961 件の投稿の平均正確率を表 9 に示す.

表 9　4 つのグループの平均正確率

転位レベル	レベル 1	レベル 2	レベル 3	レベル 4	レベル 5	レベル 6	レベル 7	レベル 8
平均正確率	0.314	0.574	0.748	0.838	0.896	0.936	0.958	0.975

前の表で，異なる転位レベルの正確率が見える．転位レベル 1 以上は，レベル 2 からレベル 8 までの各レベルに対応する投稿数を含み，他も同様である.

転位レベルが低いほど（レベル 1 からレベル 8 までどんどん高くなる）レベルに対応する投稿数が多いので，精度も低く，高いほど相対的に精度が高くなる．これは，レベルが低いほど，人間とアルゴリズムの両方からより多くの精度が求められることを意味するが，人間もアルゴリズムも，データ処理する時にラベリングやランキングの計算の際には非常に正確ではない.

実際には，人や方法の耐障害性を考慮すると，非常に厳しい要求があればレベル 1 の誤差を考慮することができ，精度の低いものにはレベルを上げることができると考える．そのため，モデル評価では，許容できる誤差レベルに基づく．例えば，レベル 3 以上の誤差を許容できる場合，今回の評価結果はレベル 4 で，正確率は 0.838 となっている．正確率が高いということは，明示的特徴量に基づいて信頼値をフィッティングする効果と，類似度を計算するために選択したサンプルが比較的よく選択されていることを示しており，その結果，類似処理後に明示的特徴があまり変化しない状況になり，明示的特徴と暗黙的特徴の整合性を示す．これも後の実験パートの基礎となる．

4.4 第 4 章のまとめ

本章では，4.1 節では信頼値モデルを構築するためにやらせ投稿特徴（明示的特徴と暗黙的特徴）の選択について，4.2 節ではデータ訓練とモデル構築の詳細について，4.3 節では構築した信頼値モデルの評価方法と結果について述べた．次章では，実験の詳細および考察について述べる．

第５章　実験

5.1 実験内容

本研究では，「信頼値」により投稿判断を行う．提案した信頼値モデルの有効性の検証方法としては，提案モデルを利用しない場合（ショッピングサイトで見える特徴だけ提示する）場合，利用する（信頼値モデルに関する特徴と一緒に提示する）場合の投稿の信頼性判断に対するアンケートを行う．そして，アンケートの結果から提案モデルの有効性の評価を行う．

5.1.1 実験用投稿データ

本実験は，モデル構築するときに使わなく，京東商城（global.jd.com）の「HUAWEI P40」という携帯電話に対してクローラした投稿データ（100 行）を用いる．各投稿が投稿のテキスト，投稿時間，購買時間，「この投稿が有用」とクリックされた数，フィードバックの数，評価スコアなど全 8 項目が取得する．図 18 にクローラした投稿データの一例を示す．

	ユーザID	テキスト	有用といわれ	フィードバック	スコア	購買時間	投稿時間	商品カテゴリ	
2	15029276266	在刷存在感吗？无货订货等待中，刚买完一天就有活动立减200	0	1	1	2020/12/9 22:36	2020/12/11 13:32	华为 HUAWEI P40 麒麟990 5G SoC芯片 5000万	
3	15024191136	我买贵了100元，产品是很不错，只是这个平台很很不够意思而	0	1	1	2020/12/7 19:30	2020/12/10 11:11	华为 HUAWEI P40 麒麟990 5G SoC芯片 5000万	
4	15020770364	外形外观：靓屏幕音效：中音甜，低音准，高音稳，无得顶拍照	0	1	5	2020/12/7 9:54	2020/12/9 15:17	华为 HUAWEI P40 麒麟990 5G SoC芯片 5000万	
5	15011233337	京东第一个差评，昨晚10点过看，说23点10分前下单，7号到，	0	1	1	2020/12/6 23:07	2020/12/7 13:03	华为 HUAWEI P40 麒麟990 5G SoC芯片 5000万	
6	15014603453	第二次购买，这次是帮朋友买的，自己用的好用才推荐给朋友，	0	1	5	2020/12/6 22:30	2020/12/8 8:53	华为 HUAWEI P40 麒麟990 5G SoC芯片 5000万	
7	15022164496	外形外观：手机很流畅，外形美观，给老婆买的，很喜欢！	0	1	5	2020/12/6 17:29	2020/12/9 20:27	华为 HUAWEI P40 麒麟990 5G SoC芯片 5000万	
8	15011413261	昨天晚上刚买完今天就有一百块钱优惠券了，还不能申请价保，	0	1	3	2020/12/6 16:26	2020/12/7 13:46	华为 HUAWEI P40 麒麟990 5G SoC芯片 5000万	
9	15019812533	刚收到货，非常兴奋，期待使用效果。送了手机壳和耳塞，没有	0	0	5	2020/12/6 12:25	2020/12/9 11:34	华为 HUAWEI P40 麒麟990 5G SoC芯片 5000万	
10	15018211019	性能优良，包装完好，期望后续服务更周到！	0	1	5	2020/12/6 10:28	2020/12/8 22:08	华为 HUAWEI P40 麒麟990 5G SoC芯片 5000万	
11	15009572899	到货非常快，头天下单，第二天就到了，很满意。	0	1	5	2020/12/5 21:39	2020/12/7 0:24	华为 HUAWEI P40 麒麟990 5G SoC芯片 5000万	
12	15014304120	拍照效果：帮屏幕音效：3D环绕外形外观：漂亮运行速度：常	0	1	5	2020/12/4 1:59	2020/12/8 6:48	华为 HUAWEI P40 麒麟990 5G SoC芯片 5000万	
13	14999097277	大小合适，金属质感，手感也很好。拍照效果特别棒	0	1	5	2020/12/3 22:59	2020/12/4 14:38	华为 HUAWEI P40 麒麟990 5G SoC芯片 5000万	
14	14998267703	我还没有确认收货自动就收货了	0	1	5	2020/12/3 20:10	2020/12/4 11:09	华为 HUAWEI P40 麒麟990 5G SoC芯片 5000万	
15	15006677249	华为p40，拍照效果好，运行速度快，支持华为。好评。	0	1	5	2020/12/3 19:56	2020/12/6 12:24	华为 HUAWEI P40 麒麟990 5G SoC芯片 5000万	
16	14999169835	不愧是华为P系旗舰手机，本想128足够…256更好????	0	1	5	2020/12/3 11:28	2020/12/4 14:55	华为 HUAWEI P40 麒麟990 5G SoC芯片 5000万	
17	15016491891	外形外观：很不错，以前的荣耀10不用了，就买这个了，整体和	0	1	5	2020/12/3 10:03	2020/12/8 15:46	华为 HUAWEI P40 麒麟990 5G SoC芯片 5000万	
18	15021281201	其他特色：小屏好操作运行速度：目前还可以	0	1	5	2020/12/3 8:08	2020/12/9 17:18	华为 HUAWEI P40 麒麟990 5G SoC芯片 5000万	
19	15013358918	本来想买M40的，实在是抢不到，就干脆买了P40，用了一天，	0	0	5	2020/12/3 2:00	2020/12/7 20:57	华为 HUAWEI P40 麒麟990 5G SoC芯片 5000万	
20	14998533553	还没打开	0	0	5	2020/12/2 22:09	2020/12/4 12:13	华为 HUAWEI P40 麒麟990 5G SoC芯片 5000万	
21	14995186231	外形外观：简洁屏幕音效：出色拍照效果：惊艳运行速度：快待	0	1	5	2020/12/2 12:05	2020/12/3 15:42	华为 HUAWEI P40 麒麟990 5G SoC芯片 5000万	
22	15008762464	外形外观：漂亮大气！高大上！屏幕音效：音响效果杠杠的，屏	0	1	5	2020/12/2 7:47	2020/12/6 20:37	华为 HUAWEI P40 麒麟990 5G SoC芯片 5000万	
23	14993735634	非常不错，给对象买的礼物。她很喜欢	0	1	5	2020/12/1 23:24	2020/12/3 10:04	华为 HUAWEI P40 麒麟990 5G SoC芯片 5000万	
24	15029157255	待机时间：同一款手机，人家的手机打打游戏一天下来了，我这	0	1	1	2020/12/1 22:20	2020/12/11 13:04	华为 HUAWEI P40 麒麟990 5G SoC芯片 5000万	
25	15002906686	外形外观：挺好看拍照效果：清晰运行速度：还可以，挺快的	0	1	5	2020/12/1 20:36	2020/12/5 13:52	华为 HUAWEI P40 麒麟990 5G SoC芯片 5000万	
26	15017785802	外形外观：外观漂亮是我喜欢的屏幕音效：真的不错音效挺好拍	0	1	5	2020/12/1 18:13	2020/12/8 20:36	华为 HUAWEI P40 麒麟990 5G SoC芯片 5000万	
27	14990114651	京东物流真心棒，包装完好。新机手感好，反应灵敏，不用担心	0	1	5	2020/12/1 17:05	2020/12/2 12:59	华为 HUAWEI P40 麒麟990 5G SoC芯片 5000万	
28	15012562866	上京东买正品支持国货！	0	1	5	2020/12/1 15:15	2020/12/7 18:02	华为 HUAWEI P40 麒麟990 5G SoC芯片 5000万	

図 18　HUAWEI P40 の投稿データの一例

クローラした投稿データに対して，購買時間と投稿時間の遅延時間を算出し，3.2 節のように非テキストとテキストデータを処理し，明示的特徴の投稿の遅延時間（Delay），投稿のテキストの長さ（Length），キーワード数（Keywords），

「この投稿が有用」とクリックされた数（Useful）を抽出し，信頼値公式による各投稿の信頼程度を計算する．そして，4.1 節のように SVD を利用して信頼度の指標を補正した後，明示的特徴と暗黙的特徴を組み合わせて平均化し，各投稿の信頼値（Percent）を得る．最後に，その中から各段階の信頼値を 1－2 件（全部 15 件）選択し，実験用投稿データにする．図 19 に前処理して選択した実験用投稿データを示す．

投稿ID	テキスト	Percent	Length	Keywords	Delay	Useful	购买时间	评价时间
1	外形外观：靓屏幕音效：中音甜，低音准，高音稳，无得顶	16.80716	5.26	14	2.224016	9.50	2020/12/7 9:54	2020/12/9 15:17
2	运行速度非常快，因为不玩游戏，感觉非常流畅很好。30倍	67.98159	9.20	19	2.474468	56.61	2020/4/21 10:20	2020/4/23 21:44
3	外形外观：还好，就是摄像头太大，手机裸机放桌子上不稳	85.45287	68.26	64	2.431701	26.74	2020/4/10 1:59	2020/4/12 12:21
4	屏幕音效：颜色真实度很高，清晰，显示效果很出色，不论	47.54156	18.93	38	1.138299	8.22	2020/4/13 20:46	2020/4/15 0:05
5	外形外观：长宽比太大，短一点就协调了；巨大的药丸挖孔	96.04329	99.33	118	11.52609	5.68	2020/3/31 10:17	2020/4/11 22:54
6	做工精致，华为手机做的真的不错，外观很漂亮，细节也很	47.2983	9.68	36	7.25441	6.95	2020/9/12 15:37	2020/9/19 21:44
7	不值这个价。华为的手机是靠吹出来的，拍照很一般，还没	35.45519	2.56	14	22.198	8.86	2020/9/21 11:08	2020/10/13 15:53
8	说真心话，P40实机跟渲染图还是有很大差异，没有惊艳，	78.81418	1.71	15	0.941296	97.93	2020/4/8 20:39	2020/4/9 19:15
9	京东第一个差评，昨晚10点过看，说23点10分前下单，7号	38.46408	22.25	29	0.580637	9.50	2020/12/6 23:07	2020/12/7 13:03
10	非常好！速度很快！操作方便简洁！华为???	12.36841	14.52	5	5.209942	9.50	2020/11/27 15:56	2020/12/2 20:59
11	泄露我买手机的信息，还有人专门电话推销！可恶	22.07648	14.28	6	14.26296	8.86	2020/11/11 12:32	2020/11/25 18:51
12	外形外观：外观上做工紧致，，，	27.53263	15.94	4	21.03492	9.50	2020/11/11 11:36	2020/12/2 12:26
13	这手机上手感觉好小啊，	3.986087	16.89	2	4.097778	3.14	2020/7/21 11:57	2020/7/25 14:17
14	塑料后壳质感很差，膜会割手，屏幕看着发虚，一点都不清	32.25974	9.06	15	0.608021	29.92	2020/4/9 20:59	2020/4/10 11:35
15	比4g还卡	7.24291	18.31	0	10.59787	9.57	2020/4/16 5:48	2020/4/26 20:09

図 19　実験用投稿データ

投稿の信頼性評価を行うために，信頼値（Percent）が「全然信頼できない」から「非常に信頼できる」までの 5 段階の評価尺度に変更させる．信頼値の設定範囲が(0,100)のため，(0,20)が「全然信頼できない」，(21,40)が「あまり信頼できない」，(41,60)が「どちらともいえない」，(61,80)が「比較的に信頼できる」，(81,100)が「非常に信頼できる」と設定する．そして，実験用投稿の信頼度は図 20 の通りに整理した．

投稿ID	テキスト	Percent		参考信頼度
1	外形外观：靓屏幕音效：中音甜，低音准，高音稳，无得顶	16.80716	全然信頼できない	1
2	运行速度非常快，因为不玩游戏，感觉非常流畅很好。30倍	67.98159	比較的に信頼できる	4
3	外形外观：还好，就是摄像头太大，手机裸机放桌子上不稳	85.45287	非常に信頼できる	5
4	屏幕音效：颜色真实度很高，清晰，显示效果很出色，不论	47.54156	どちらともいえない	3
5	外形外观：长宽比太大，短一点就协调了；巨大的药丸挖孔	96.04329	非常に信頼できる	5
6	做工精致，华为手机做的真的不错，外观很漂亮，细节也很	47.2983	どちらともいえない	3
7	不值这个价。华为的手机是靠吹出来的，拍照很一般，还没	35.45519	あまり信頼できない	2
8	说真心话，P40实机跟渲染图还是有很大差异，没有惊艳，	78.81418	比較的に信頼できる	4
9	京东第一个差评，昨晚10点过看，说23点10分前下单，7号	38.46408	あまり信頼できない	2
10	非常好！速度很快！操作方便简洁！华为???	12.36841	全然信頼できない	1
11	泄露我买手机的信息，还有人专门电话推销！可恶	22.07648	あまり信頼できない	2
12	外形外观：外观上做工紧致，，，	27.53263	あまり信頼できない	2
13	这手机上手感觉好小啊，	3.986087	全然信頼できない	1
14	塑料后壳质感很差，膜会割手，屏幕看着发虚，一点都不清	32.25974	あまり信頼できない	2
15	比4g还卡	7.24291	全然信頼できない	1

図 20　実験用投稿の信頼度整理

5.1.2 被験者

実験参加者は，北陸先端科学技術大学院大学の中国留学生 10 人であり，全員が平均毎月ネットで 3－5 回ショッピングする経験があり，10 人中 6 名がショッピングサイトで携帯電話を購入する経験があった.

5.1.3 実験の流れ

① 実験に選択した商品（「HUAWEI P40」という携帯電話）に関する基本情報を紹介する.

② ショッピングするときに見える特徴を含む投稿画面を展示させ，被験者がそれを読んで信頼程度を評価する．図 21 に被験者が展示させる投稿画面の一例を示す.

③ アンケート 1 を記入する.

④ 参考信頼値画面を展示させ，信頼程度評価の変更ができる．図 22 に被験者が展示させる参考信頼値画面の一例を示す.

⑤ ステップ②③④を 15 回繰り返す.

⑥ アンケート 2 を記入する.

商品のカテゴリ： 华为 HUAWEI P40 麒麟990 5G SoC芯片 5000万超感知徕卡三摄 30倍数字变焦 8GB+128GB亮黑色全网通5G手机

投稿3

投稿本文	点赞数	回复数	评分	投稿時間
外形外观：还好，就是摄像头太大，手机裸机放桌子上不稳，必须得套上壳子。 屏幕音效：还好，正常效果，没惊喜也没有太差。 拍照效果：就这个还行，如果拍照都不行这款手机真没救了。 运行速度：还好，新机子都挺快的，看后续表现。 待机时间：这个得差评，感觉手机电量是虚的，放一晚上就没电了，手机完全遭不住，昨天试了五个手机一起直播，就这个手机耗电最快。 其他特色：没有特色，挺失望的，用过苹果全系列，去年买过nova5P当备用机，感觉还不错，今年就想着P系列好歹也是旗舰机就试一试，昨天手机充电线都不敢拔掉，得一直插着，昨晚满电睡觉，早晨起床只剩一半，而且手机发烫，只要屏幕亮着就发烫。真配不上这个价格。 哦对，我是用另一个手机写的，新买的P40就屏幕放那儿亮着没动，就写评价的这几分钟，手机电量成功共53%降到50%，厉害了！！！	57	27	1	2020/4/12 12:21

図 21　ステップ②に展示させる投稿画面の一例

図 22　ステップ④に展示させる参考信頼値画面の一例

5.1.4 アンケート内容

投稿画面を展示させ，信頼程度評価に関するアンケート 1 の質問項目は以下のとおりである．これは，投稿 1 件ずつ，15 件すべてに対して行う．

(1)　投稿が信頼できるかどうか判定してください

1)　全然信頼できない
2)　あまり信頼できない
3)　どちらともいえない
4)　比較的に信頼できる
5)　非常に信頼できる

(2)　何を理由に判定しましたか（複数の特徴を選択できる）

a)　投稿本文：長さから
b)　投稿本文：読むから
c)　投稿時間
d)　「有用」が言われる数
e)　フィードバックの数
f)　スコア

また，参考信頼値画面を展示させ，参考信頼値と各項目が被験者に対する影響に関するアンケート 2 の質問項目は以下のとおりである．これは，15 件の投稿評価が終わりに実施した．

(1)　参考信頼度を使ったことで，投稿の信頼性を判断することに有用性を感じましたか？

a)　はい
b)　いいえ
c)　どちらともいえない

(2) モデルを使うとき，信頼性判断に参考する項目は何ですか？（複数の特徴を選択できる）

a) 投稿内容：長さから
b) 投稿内容：読むから
c) 投稿時間
d) 購買時間
e) 「有用」が言われる数
f) フィードバックの数
g) スコア
h) 購買時間と投稿時間の時間差
i) 投稿内容のキーワード数
j) 信頼値

(3) ある商品を購入するとき，意識的に見る情報はほかに何がありますか？

5.2 実験結果と考察

5.2.1 アンケート結果

表 10，表 11，表 12 に投稿の信頼性判断に関するアンケート結果を示す．表 10 は各投稿の信頼度評価に関するアンケート 1 の回答結果（参考信頼度を展示しないときの結果）である．行は各質問項目に，列は各投稿に対応する．各セルの値は，各投稿に対する各回答を行った被験者の人数である．

表 10 に注目し，参考信頼値を提示しないとき，各被験者が投稿に対して信頼度への判断が基本的に似ている．判断理由としては，「投稿本文を読んでから」という投稿テキストの内容を参考にした被験者が多く，1 つの投稿あたり平均 9 人が参考にしていた．また，1 つの投稿あたり「有用といわれる数」は平均 6 人，「投稿本文の長さ」は平均 4 人，「フィードバック数」は平均 2 人，「スコア」は平均 2 人が参考にしていた．「投稿時間」は投稿信頼度判断するときにあまり参考にしていなかった．このように，ユーザが投稿への信頼度を判断するとき，テキスト内容だけ主観的に判断を行うことがわかる．

表 10　各投稿の信頼度評価に関するアンケート 1 の結果

投稿 ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	投稿が信頼できるかどうか判定してください														
全然信頼できない	7	0	0	0	0	0	1	0	0	7	1	4	6	3	6
あまり信頼できない	3	0	0	3	0	2	7	1	1	2	6	6	3	5	2
どちらともいえない	0	0	0	6	0	5	1	2	1	1	3	0	1	1	2
比較的に信頼できる	0	6	3	1	2	3	1	4	8	0	0	0	0	1	0
非常に信頼できる	0	4	7	0	8	0	0	3	0	0	0	0	0	0	0
	何を理由に判定しましたか														
投稿本文：長さから	4	5	6	1	7	1	1	1	6	4	2	5	5	1	2
投稿本文：読むから	8	8	8	8	10	10	8	9	8	10	10	10	10	10	9
投稿時間	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0
「有用」が言われる	1	8	7	5	5	2	3	9	1	1	1	1	6	8	5
フィードバック数	1	5	3	0	0	0	2	7	1	0	2	0	6	5	0
スコア	2	0	2	1	3	3	3	3	2	2	2	2	2	3	3

次に，表 11 は各被験者が各投稿に対して信頼性判断を行った回答結果（参考信頼度を提示しないときの結果）であり，参考信頼度の欄は 5.1 節で整理した各投稿の参考信頼度の評価段階の値である．表 12 は参考信頼度を展示した後，各被験者が各投稿に対して信頼性判断の変更を行った回答結果である．色付けセルは被験者に参考信頼度を展示した後，変更したところである．参考信頼度を展示するとき，被験者が投稿への信頼度判断を変更したところである．つまり，参考信頼度が投稿の信頼程度への判断に影響を与えたと考えられる．

表 11　各投稿に対して信頼性判断の結果

投稿 ID	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
参考信頼度	**1**	**4**	**5**	**3**	**5**	**3**	**2**	**4**	**2**	**1**	**2**	**2**	**1**	**2**	**1**
被験者 1	2	4	5	3	5	4	1	4	2	3	1	2	2	2	1
被験者 2	1	5	4	3	5	4	2	4	4	1	2	2	3	1	2
被験者 3	1	5	5	2	5	3	4	3	4	1	2	1	1	1	1
被験者 4	1	4	5	3	4	3	3	5	4	2	3	2	1	2	3
被験者 5	1	5	4	3	5	3	2	2	4	1	2	1	2	2	1
被験者 6	1	4	5	4	5	3	2	5	4	1	3	1	1	2	1
被験者 7	2	4	5	3	4	2	2	4	4	1	2	2	1	3	1
被験者 8	1	5	4	2	5	4	2	4	3	2	2	2	2	4	2
被験者 9	1	4	5	2	5	3	2	5	4	1	2	2	1	1	1
被験者 10	2	4	5	3	5	2	2	3	4	1	3	1	1	2	3

表 12　参考信頼度が展示させた後信頼性判断を変更した結果

投稿 ID	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
参考信頼度	**1**	**4**	**5**	**3**	**5**	**3**	**2**	**4**	**2**	**1**	**2**	**2**	**1**	**2**	**1**
被験者 1	**1**	4	5	3	5	**3**	1	4	2	**1**	1	2	2	2	1
被験者 2	1	**4**	**5**	3	5	**3**	2	4	**2**	1	2	2	**1**	1	2
被験者 3	1	**4**	5	**3**	5	3	**2**	3	**2**	1	2	**2**	1	1	1
被験者 4	1	4	5	3	**5**	3	3	**4**	**2**	**1**	**2**	2	1	2	**1**
被験者 5	1	5	4	3	5	3	2	2	**2**	1	2	1	2	2	1
被験者 6	1	4	5	**3**	5	3	2	**4**	**2**	1	**2**	1	1	2	1
被験者 7	2	4	5	3	4	2	2	4	**2**	1	2	2	1	3	1
被験者 8	1	5	4	**3**	5	**3**	2	4	3	**1**	2	2	2	**2**	2
被験者 9	1	4	5	2	5	3	2	5	**2**	1	2	2	1	1	1
被験者 10	2	4	5	3	5	**3**	2	3	**2**	1	3	1	1	2	**1**

表 13 は，アンケート 2 の質問項目「参考信頼値を使ったことで，投稿の信頼性を判断することに有用性を感じましたか？」に対する回答結果である．被験者 10 人中 9 人（90%）は「はい」と回答し，参考信頼度を使ったことで，投稿の信頼性を判断することに有用性を感じたことがある．また，信頼値モデルを利用する場合，「信頼性判断に参考する項目は何ですか？」に対する回答が図 23 に示す．被験者 10 人中 9 名が「参考信頼値」に影響されていると感じ，信頼値判断が第一番の参考項目になった．「購買時間と投稿時間の時間差」が 8 人，投稿内容を読む感じが 7 人選択され，第二，三番の参考項目になり，表 10 と比べると，提案モデルが展示させる内容（ネットショッピングするときに見られない内容）がもっと参考になると確認できる．以上より，提案モデルを利用することで，ユーザに対して，投稿の信頼性判断に意識的な影響を与えるとともに，有効な信頼性の判断支援ができたと考えられる．

表 13　参考信頼度の有用性に関する質問項目の結果

	参考信頼度を使ったことで，投稿の信頼性を判断することに有用性を感じましたか？
はい	9(90%)
いいえ	0(0%)
どちらともいえない	1(10%)

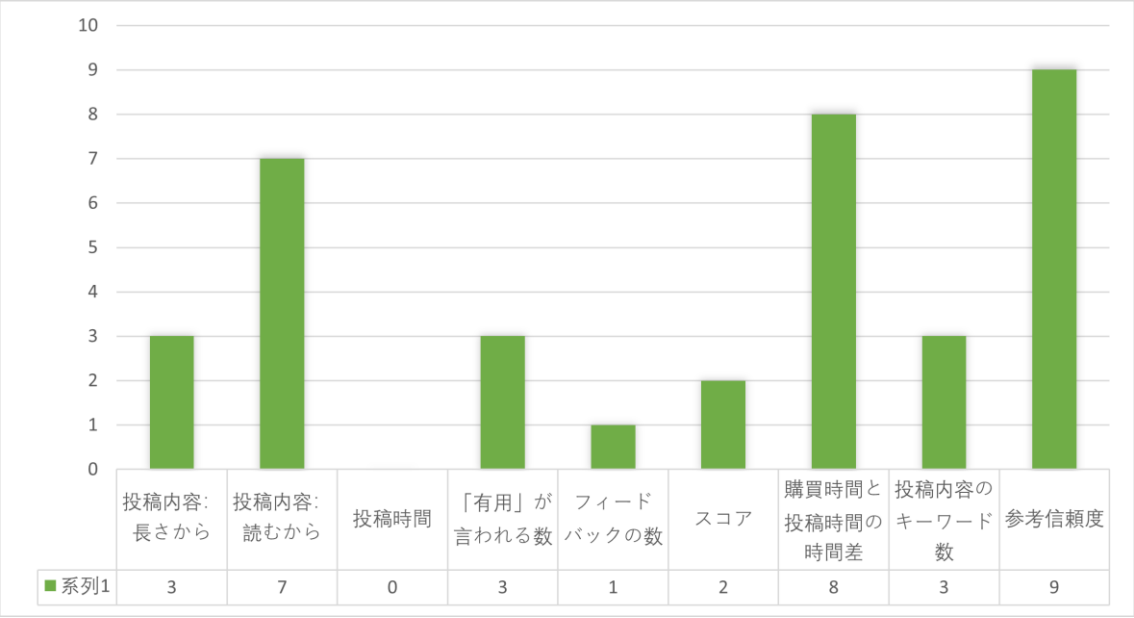

図 23　信頼性判断に参考する項目に関する質問の結果

5.2.2 考察

表 12 により，投稿 9 への信頼度判断は提案モデルを利用するときに，大きな変更があることを確認できる．そして，投稿 9 について考察する．表 10 から見ると，被験者 10 人中 6 人は投稿の長さを判断参考になり，投稿本文が長くてスコアが低い投稿がよく信頼できるといわれた．しかし，投稿 9 のテキストの内容から考察し，「JD にはプライスプロテクション[1]がない」という内容が現実とは違い，「天猫旗艦店に購買しよう」というほかの購買店舗を宣伝する内容があり，まだ購買時間と投稿時間の時間差が半日しかないため，信頼程度が低く，競合他社の意図的な悪評の可能性があると考えられる．そして，参考信頼値と遅延時間を見られた後，8 人の被験者が投稿 9 の信頼度の「比較的に信頼できる」を「あまり信頼できない」と変更した．それより，時間差と参考信頼値が被験者に大きな影響を与えることを見られ，特に提案モデルは有用だと考えられる．

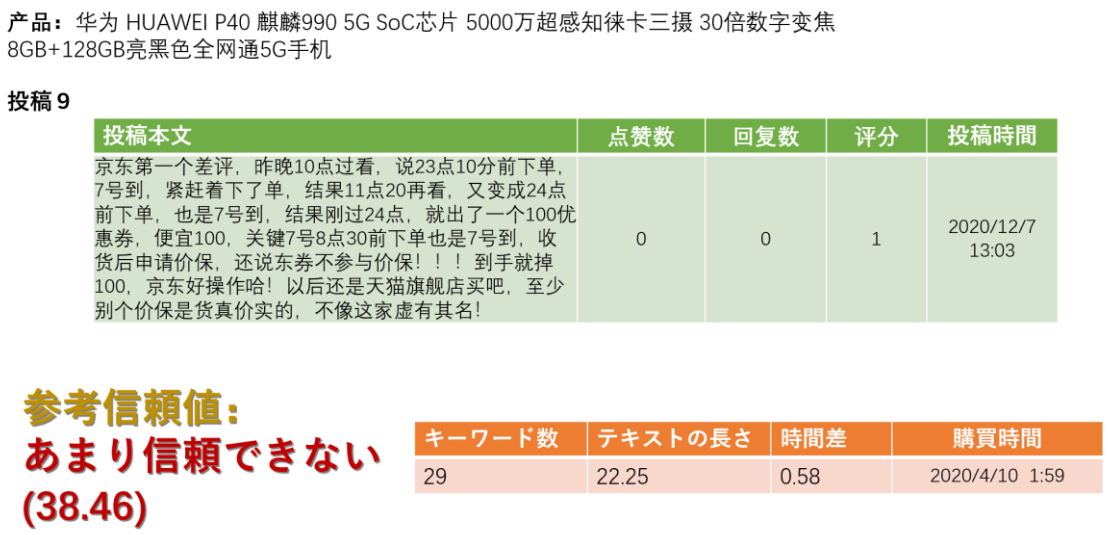

図 24　投稿 9 の内容と参考信頼値の展示画面

[1] プライスプロテクションとは，商品購入後，同一商品の価格が値下げされた場合，プライスプロテクションの申請をすることで，差額相当分のクーポンをプレゼントすることをいう．

最後に，実験の最後に行ったアンケート 2 の質問項目「ある商品を購入するとき，意識的に見る情報はほかに何がありますか？」に対する回答結果について考察する．商品投稿を見るとき，投稿内容だけでなく，写真付き投稿やビデオ付き投稿に対してよく気が付くことが 8 人いわれた．例えば，販売店内の写真を直接的利用したり，解像度がすごく低い写真を利用したりすることがやらせ投稿と考えられる．7 人が販売店舗の総合的評価スコアや販売量，同一店舗内ほかの商品の投稿真実性に対しても注目するといわれた．1 つの店舗で販売量がすごく高い商品が 1 個しかない場合，やらせ投稿の可能性が高いと考えられる．また，5 人が追加評価をよく参考になるといわれた．商品を長い時間使用する後，使用感などの追加は信頼性が高いと考えられる．そのほか，3 人が極端的な好評や悪評，2 人がバイヤーからの質問欄[2]，1 人が店舗内各投稿者の会員レベルに対して注目するといわれた．この結果から，消費者が商品購買を考えるとき，提案モデルにある投稿者ひとりの特徴と投稿テキストの特徴だけでなく，店舗評価や写真などの特徴にも重視されていることがわかる．

5.3 第 5 章のまとめ

本章では，5.1 節では実験内容について，5.2 節では実験結果と考察について述べた．次章ではまとめと今後の展望について記す．

[2] 中国の EC サイトで，1 つの商品に対して未購入の人が質問され，購入した人だけ回答できる機能である．

第 6 章　おわりに

6.1 本研究のまとめ

本研究では，ネットショッピングするときに投稿を読む消費者の受ける印象からやらせ投稿を代表できる主な特徴を抽出し，それに基づいた機械学習を用いて，やらせ投稿への「信頼値モデル」を構築した．また，提案モデルによって，消費者によるやらせ投稿に対する識別能力を強化させたかどうかを検証した．

本研究では，京東商城（global.jd.com）でクロールされたデータを利用し，アノテーションも手動で実装した．特徴選択の部分では，投稿者に基づくか投稿テキストに基づくかにかかわらず，真の投稿と偽の投稿の違いを明らかな特徴で識別することができず，偽の投稿をより良く効果的に検出するためには，投稿者に基づくアプローチと投稿テキストに基づくアプローチの両者をより効果的に組み合わせることがより良いアプローチである．また，回帰分析の教師あり学習アプローチを，特異値分解アプローチを用いて投稿テキストの暗黙的特徴を分析することと融合させ，明示的特徴のみを用いてやらせ投稿を判断するという欠点を克服した．モデル評価結果からすると，このアプローチが有効であることが示されている．

実験結果より，ユーザ自分自身で投稿への信頼値判断と提案モデルを利用するとき判断の変更から，提案モデルを利用することで，ユーザに対して投稿の信頼性判断に意識的な影響を与えるとともに，有効な信頼性の判断支援を行うことができると示唆できた．ビッグデータ時代に，大量のデータにおけるやらせ投稿の認識における一助になることに期待したい．

6.2 今後の展望

投稿テキストの解析については，投稿テキストを処理する際に，依存性構文を用いてやらせ投稿のキーワード辞書を作成することで，より深い探究を行うことも可能である．また，やらせ投稿を識別するための特徴パターンの組み合わせは多数あり，識別方法も多種多様である．やらせ投稿をよりよく識別するために，今後の作業では，より多い種類の特徴パターンを探りながら，より多くの異なるタイプの識別方法を試すことに焦点を当てていくと考える．

謝辞

本研究を遂行するにあたり，多くの方々に多大なご支援をいただきました．この場をお借りして御礼申し上げます．

指導先生である金井准教授には，ご多用の中でも筆者のために時間を割いてくださり，研究に関して様々なご指導，ご鞭撻を賜りました．また，本研究を進めるにあたり，貴重なご意見やアドバイスを頂きました，内平直志教授，西本一志教授，由井薗隆也准教授に心より感謝致します．

研究活動ならびに学生生活において，研究室の皆様や友人にも大変お世話になりました．良い研究室メンバー，友人に恵まれたことで，非常に充実した大学院生活を送ることができました．深く感謝いたします．

最後に，ご多用の中，各実験に被験者として協力いただきました北陸先端科学技術大学院大学博士前期課程の皆様にも感謝申し上げます．

参考文献

[1] 入手先: https://myel.myvoice.jp/products/detail.php?product_id=26007

[2] 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(オンライン).
入手先:https://research.nttcoms.com/database/data/001436/

[3] 入手先: https://sakura-checker.jp/

[4] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate [EB/OL], [2017-10-20]. https://arxiv.org/abs/1409.0473.

[5] Jindal N, Liu B. Review spam detection[C]// International Conference on World Wide Web. 2007:1189-1190.

[6] Jindal N, Liu B, Lim E P. Finding unusual review patterns using unexpected rules[C]// ACM Conference on Information and Knowledge Management, CIKM2010, Toronto, Ontario, Canada, October. DBLP, 2010:1549-1552.

[7] Jindal N, Liu B. Opinion spam and analysis[C]// International Conference on Web Search and Data Mining. ACM, 2008:219-230.

[8] Li X, Yan X. A Novel Chinese Text Mining Method for E-Commerce Review Spam Detection[M]// Web-Age Information Management. Springer International Publishing, 2016.

[9] Sihong Xie, Guan Wang, Shuyang Lin, Philip S. Yu“Review spam detection viatime series pattern discovery”, ACM Proceedings of the 21st international conference companion on World Wide Web,pp.635-636,2012.

[10] S. Patil M, M. Bagade A. Online Review Spam Detection using Language Modeland Feature Selection[J]. International Journal of Computer Applications, 2012,59(7):33-36.

[11] Jayanthi S K, Sasikala S. Link Spam Detection based on DBSpamClust with Fuzzy C-means Clustering[J]. International Journal of Next-Generation Networks, 2010,2(4):1-10.

[12] Xu X, Han T, Xu Z, et al. Design and Implementation of Chinese Spam Review Detection System[M]// Chinese Lexical Semantics. Springer Berlin Heidelberg, 2013.

[13] 伊木惇, 亀井清華, 藤田聡. レビューを対象とした信頼性判断支援システムの提案. 情報処理学会論文誌, 55(11), pp. 2461-2475, 2014.

[14] Lim E P, Nguyen V A, Jindal N, et al. Detecting product review spammers using rating behaviors[C]. ACM International Conference on Information and Knowledge Management. ACM, 2010:939-948.

[15] 岡山光平, 石川博, 廣田雅春. フェイクニュース分類器を用いたロコミサイトのレビューの分析. 第 10 回データ工学と情報マネジメントに関するフォーラム, 2018.

[16] Fang Youli, et al.: "Detecting Fake Reviews Based on Review-Rating Consistency and Multi-dimensional Time Series." International Conference on Algorithms and Architectures for Parallel Processing. Springer, Cham, 2018.
[17] Crawford M, Khoshgoftaar TM, Prusa JD, Richter AN, Al Najada H. Survey of review spam detection using machine learning techniques. Journal of Big Data. 2015 Oct 5;2(1):1.
[18] Jingdong Wang, Haitao Kan, Fanqi Meng, et al. Fake Review Detection Based on Multiple Feature Fusion and Rolling Collaborative Training. IEEE Access, Digital Object Identifier, 2020.
[19] 安藤まや, 関根聡. レビューには何が書かれていて、読み手は何を読んでいるのか? 言語処理学会第 20 回年次大会発表論文集, 2014.
[20] Ott M, Choi Y, Cardie C, et al. Finding deceptive opinion spam by any stretch of the imagination[J]. Computer Science, 2011, 1:309-319.
[21] C. Harris. Detecting deceptive opinion spam using human computation. In Workshops at AAAI on Artificial Intelligence, 2012.
[22] Li J, Ott M, Cardie C, et al. Towards a General Rule for Identifying Deceptive Opinion Spam[C]// Meeting of the Association for Computational Linguistics.2014:1566-1576.
[23] Chen C, Zhao H, Yang Y. Deceptive Opinion Spam Detection Using Deep Level Linguistic Features[M]//Natural Language Processing and Chinese Computing. Springer International Publishing, 2015:465-474.
[24] 岩井秀成, 池田 郁, 土方嘉徳, 西田正吾: レビュー文を対象としたあらすじ分類手法の提案, 電子情報通信学会論文誌, pp. 1222-1234, 2013.
[25] 三舩正暁, 金明哲 (n.d.) , ネットショッピングにおけるスパムレビューの特徴分析, 日本計算機統計学会 第 30 回大会 p. 9- 12, 2016.
[26] 劉 真真, 桝井文人, プタシンスキ ミハウ, 中文スパムレビュー検出のためのショッピングサイトレビュー分析, 人工知能学会全国大会 第 30 回大会, 2016.
[27] 入手先: https://netshop.impress.co.jp/node/5215