

ポジティブ・ネガティブ分析を備えた ユーザレビュー集約管理システムの提案

小林亮^{†1} 鈴木浩^{†1} 服部哲^{†1} 速水治夫^{†1}

ネットショッピングの普及とともに、ユーザレビューを商品購入の参考にする利用者が増加している。しかし、近年は EC サイトのほか、価格比較サイトやブログ、SNS など複数のサイトにユーザレビューが分散し、閲覧する際にサイト間を横断検索しなくてはならないという問題が発生している。また、EC サイトのユーザレビューでは、星印などによる点数評価とレビュー文章が掲載されるが、このような評価点数と文章だけでは、商品に対してポジティブな評価のみを閲覧したいときや、反対にネガティブな評価のみを閲覧したいときには不便である。そこで、これらの問題を解決するために、レビューのポジティブ・ネガティブ分析を備えたユーザレビュー集約管理システムを提案する。

A proposal of the unified management system of the user review using analysis of positive and negative

TASUKU KOBAYASHI^{†1} HIROSHI SUZUKI^{†1}
AKIRA HATTORI^{†1} HARUO HAYAMI^{†1}

The users who refer to a user review are also increasing in number with the spread of online shopping. However, decentralization of the user review to two or more websites has occurred in recent years. As for the user review of a major company shopping site, mark and a review text are published in many cases. However, it is inconvenient only by mark and the text to peruse positive or negative evaluation to goods. Then, in order to solve these problems, the user review unified management system equipped with positive and negative analysis of a review.

1. はじめに

ネットショッピングの普及とともに、ユーザレビューを参考にするユーザも増加している。富士通総研が 2009 年にネットショッピングを利用した 10 歳から 69 歳の男女 2,000 人に行った調査では、1 年間での PC でのネットショッピング平均利用回数は約 14 回という結果が出ている¹⁾。さらに、そのうち約 4 割のネットショッピングユーザが、ユーザレビューを参考にしたと回答している。また同年、インターネット調査会社のマクロミルがネットショッピング利用経験のある 20 歳から 59 歳の男女 516 人に行った調査では、約 8 割のネットショッピングユーザがユーザレビューを参考にし、約 6 割のネットショッピングユーザが価格比較サイトやユーザレビュー投稿サイトを参考にしていると回答している²⁾。さらに、2010 年に経済産業省が行った調査では、商品選択時に参考にする情報として、ユーザレビューがテレビ、メーカーの公式ウェブサイトで次いで第 3 位に位置しているという結果とともに、実店舗で購入する場合でも、インターネット上のユーザレビューを参考にするという結果が出ている³⁾。

これらの結果から、ユーザレビューがネットショッピングユーザの商品購入の意思決定に大きく関わっていると考えた。

2. ユーザレビューについて

多くのネットショッピングサイト（以下、EC サイトという）にはユーザレビューを投稿できる機能が備わっている。ユーザレビューは購入者が商品に対する感想や意見を述べたものであり、大手 EC サイトでは、星印などによる点数評価とレビュー文章を投稿する事が多い。さらに近年は購入者が撮影した商品写真や、実際に使用している様子を撮影した動画など、文章以外の方法でレビューを投稿できるサイトも存在する。

また、EC サイトの増加に伴って、ユーザレビューの投稿に特化したユーザレビュー集約サイトや、複数のサイト間の価格を比較する価格比較サイトが人気を得ており、価格比較サイトの中には掲示板機能やユーザレビュー投稿機能を備えたサイトも存在する。

さらに、製品の公式ウェブサイトや情報ポータルサイトで専門家や開発者によるレビューが掲載されることもあるほか、個人のブログやウェブサイト、Twitter や Facebook といった SNS 上でユーザレビューを公開するユーザも多い。ブログやウェブサイトでは、文字数や画像の数に制限が無いことから、より詳しいユーザレビューを公開することができる。

このように現在では EC サイトを含めた様々なサイト上でユーザレビューが公開されている。

なお、本論文中でいう大手 EC サイトの定義は（社）日本通信販売協会が 2008 年に発表した「インターネット通販

^{†1} 神奈川工科大学大学院工学研究科情報工学専攻
Course of Information and Computer Sciences, Graduate School of
Kanagawa Institute of Technology

利用者実態調査」の結果を基にしている⁴⁾。

3. ユーザレビューの問題点

現在のユーザレビューには次のような問題点が存在する
と考える。

3.1 レビューの分散

前章で述べたように、現在では大手 EC サイトを含めた
様々なサイト上でユーザレビューの投稿と閲覧をすること
ができるが、結果として、メジャーな商品ほどレビューが
乱立し、反対にマイナーな商品のレビューを探す際は、様々
なサイトを横断して検索する必要がある。

3.2 ポジティブな評価・ネガティブな評価の混在

現在、多くの EC サイトのユーザレビューでは、レビュ
ー文章と併せて星印などによる点数評価が行われており、
この点数をもとにソートできる機能が備わっていることが
多い。しかし、この点数評価とレビュー文章だけでは、商
品に対してポジティブな評価のみを閲覧したいときや、反
対にネガティブな評価のみを閲覧したいときには不便であ
る。

例えば、提案システムでユーザレビューの収集を行った
3つの EC サイト (Amazon.co.jp, 楽天市場, Yahoo!ショッ
ピング) での点数評価は 5 点満点での評価だが、点数評価
が 5 点であっても商品に対してネガティブな評価が含まれ
ている場合があり、反対に点数評価が 1 点であっても、商
品に対してポジティブな評価が含まれている場合もある。

商品に対してポジティブな評価は、メーカーの公式ウェ
ブサイトや、カタログなどを見れば容易に得ることができ
るが、商品に対してネガティブな評価というのは、EC サ
イトやブログなどに掲載されたユーザレビューからしか得
ることができない。そのため、大量のユーザレビューから
ポジティブな評価とネガティブな評価を抽出することは重
要であり、効率よくユーザレビューを閲覧するためには不
可欠な作業であると考えられる。

4. 問題解決へのアプローチ

本論文では、前章で述べた問題点に対して、次のような
解決策を提案する。

4.1 大手 EC サイトからのユーザレビュー収集

指定した商品のユーザレビューを 3 つの EC サイト
(Amazon.co.jp, 楽天市場, Yahoo!ショッピング) から自
動的に収集する。収集したレビューは、評価点数ごと、取
得先サイトごと等、条件を指定して閲覧することができる。

また、一度ユーザレビューを取得した商品は、その後一
定期間ごとに自動的にレビュー収集処理が行われ、各 EC
サイトの最新のユーザレビューに更新される。

さらに、個人のブログや SNS サイトなど、参考になるユ
ーザレビューが記載されたサイトをユーザ同士でまとめあ
う情報投稿機能、Twitter のハッシュタグを利用した Twitter

からの関連ツイート収集機能も備えた。

4.2 ユーザレビューのポジティブ・ネガティブ分析

収集したユーザレビューを句点や各種記号をもとに一
文単位に分解し、形態素解析処理およびナイーブベイズ分
類器による分類分けを行い、ポジティブな意見が含まれて
いる部分と、ネガティブな意見が含まれている部分に分け、
その割合と分析結果を提示する。

また、ユーザレビューの投稿者ごとに過去に投稿したユ
ーザレビューを収集し、同様の処理を行うことで、ポジテ
ィブとネガティブどちらの評価を投稿する割合が多いユー
ザなのか分析を行うことができる。

さらに、ユーザレビューの分類分けにナイーブベイズ分
類器を用いることで、学習データの修正や追加が容易にな
り、分析結果に誤りがあった場合には即時に学習データの
修正を行える。

5. 関連研究

5.1 評判情報検索のための意見文収集に関する研究

インターネット上のユーザレビューを集約する研究と
して、清水らの「評判情報検索のための意見文収集に関す
る研究」がある⁵⁾。

この研究では Google や Yahoo! といった既存の検索エン
ジンを活用し、指定した商品に対するユーザレビューが記載
された Web ページを収集・提示する「評判情報収集システ
ム」を試作している。

大まかな処理の流れとしては、入力された商品名に対し
て、検索エンジン用の検索クエリ文字列を生成し、検索し
て得られた Web ページの中からユーザレビューが含まれ
るであろうページのみを機械的に選別して提示する。

この方法では、個人のブログやマイナーな EC サイトの
ユーザレビューを収集できるといった利点があるが、ユー
ザレビューに関係のないページが提示されてしまうという
欠点もある。

本論文では、ユーザレビューの自動取得とユーザ同士に
よるまとめ機能を合わせることで、大量のユーザレビュー
を効率よく収集・閲覧できるシステムを提案する。

5.2 テキストマイニングを用いた宿泊レビューからの注 目情報抽出方法

ユーザレビューに対して文章解析処理を行い、特定の情
報を抽出する研究として、辻井らの「テキストマイニング
を用いた宿泊レビューからの注目情報抽出方法」がある⁶⁾。

この研究では、インターネット宿泊サイトに投稿された
宿泊レビューを肯定意見と否定意見に分類し、数値化を行
っている。

ユーザレビューは表記の揺れ (「スタッフ」と「係員」)
や、意図の解釈の違い (「騒音」+「大きい」, 「テレビ」+
「大きい」), 口語調の書き込みが散在することなど、自然
言語処理を行う際に問題になる点があり、これらの問題点

に対して、類義辞書や、評価属性辞書、口語調の単語をまとめた辞書を使用して、解析の精度を向上させている。

本論文では、ナイーブベイズ分類器を使用し、商品ジャンルにとらわれない汎用性と、ユーザが学習データの訂正を行える利便性を兼ね備えた、ユーザレビュー分類システムを提案する。

5.3 高精度日本語解析エンジン「なずき」

文章から感情を分析する研究として、株式会社 NTT データが開発・販売を行っている高精度日本語解析エンジン「なずき」がある⁷⁾。

「なずき」は単語の品詞や読み、アクセント等を管理する辞書と、単語同士の意味のつながりを管理する知識を用いて非常に高精度な解析処理を行える日本語解析エンジンである。

文章がどのような話題について書かれたものなのか解析する「話題抽出」や、その話題に対して文章の書き手がどのような感情を持っているのか判別する「感性抽出」といった機能を備えており、主にアンケートやブログの解析、インターネット広告配信のための Web ページ分析といった、マーケティング用途向けに販売されている。

本論文では、EC サイトに掲載されたユーザレビューの解析に重点を置き、レビュー独自の文章表現への対応や、レビュー投稿者の過去の投稿内容を利用した、ポジティブ・ネガティブ分析を提案する。

6. 提案システム概要

6.1 システム概要

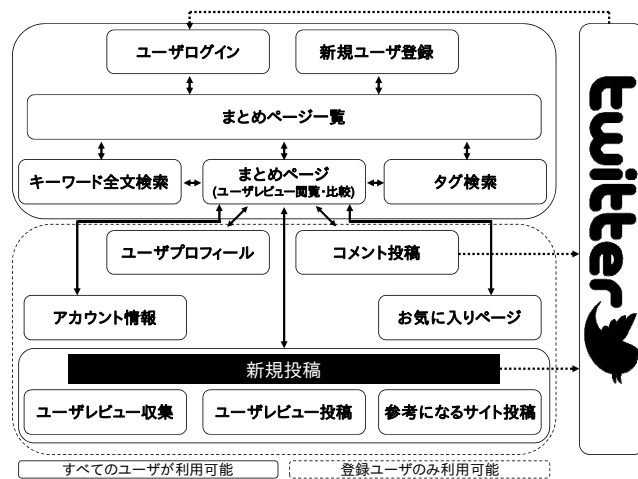


図 1 画面遷移図
 Figure 1 Screen transition diagram.

提案システムはシステム未登録ユーザを含めたすべてのユーザが利用可能な機能と、システムにユーザ登録を行ったユーザのみが利用できる機能に分かれる (図 1)。

大手 EC サイトから収集したユーザレビューや、システム登録ユーザが投稿したユーザレビュー、参考になるサイ

ト情報、コメントは商品ごとに「まとめページ」にまとめられ、一つのページで閲覧することができる。

提案システムでは、商品一つにつきユーザレビューの収集、ユーザレビューの投稿、参考になるサイト投稿のいずれかが行われたときにまとめページ 1 ページが自動生成され、その後はユーザ同士でそのページに追記を行うことで情報を増やす。

6.2 システム構成

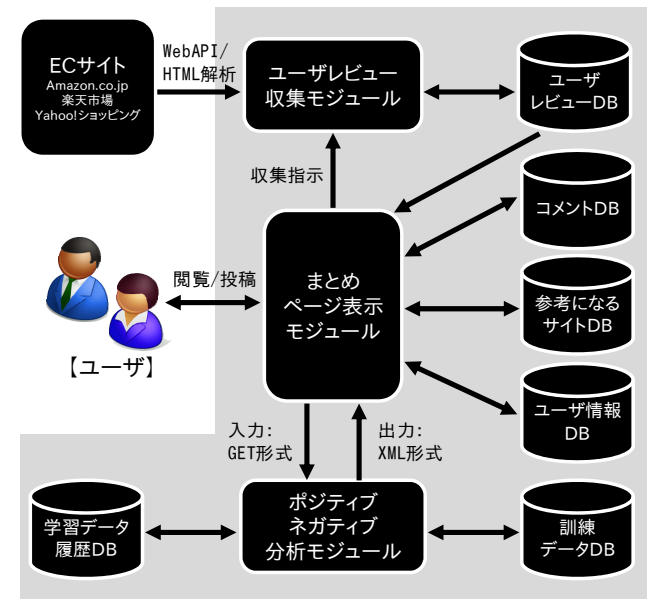


図 2 システム構成 (網掛け部分が提案システム)

Figure 2 System configuration.

提案システムは Web アプリケーションとして開発した。ユーザはまとめページ表示モジュールを通してユーザレビューの閲覧や投稿を行う (図 2)。ユーザがユーザレビュー収集指示を行うと、ユーザレビュー収集モジュールが3つの EC サイトに対して WebAPI と HTML 解析によるデータ取得を行い、ユーザレビューデータベース上に保存する。

ポジティブ・ネガティブ判定モジュールはユーザレビュー収集モジュールが収集したユーザレビューに対してナイーブベイズ分類器によるポジティブ・ネガティブ判定を行い、結果を XML 形式で返す。また、ユーザによって判定結果の修正が指示された場合は、訓練データデータベースおよび、学習データデータベースに修正後の値を保存し、以後の判定に利用する。

6.3 ユーザレビュー収集

ユーザレビュー収集処理は、3つの EC サイト (Amazon.co.jp・Yahoo!ショッピング・楽天市場) からユーザレビューを収集する。

各サイトからの商品情報取得には、各サイトが公式に提供している WebAPI を使用し、ユーザレビューと評価点数の取得には HTML 解析を使用した (図 3)。

商品によっては膨大な数のユーザレビューが投稿されている場合があり、そのすべてを取得すると、処理時間が増加し、取得対象のサイトにも連続アクセスによる負荷がかかってしまう。そのため、取得するユーザレビューの上限を各サイト 100 件までとした。すなわち、3 つのサイト合わせて最大 300 件のユーザレビューを取得する。

また、3 つのサイトからユーザレビューを収集する際には、ASIN コードと JAN コードの 2 種類の商品コードを用いて、商品の識別を行っている。ASIN コードとは、Amazon.co.jp 独自の商品コードであり、商品ごとにユニークなコードが割り当てられている。JAN コードとは、商品のバーコードに使用されている数値であり、ASIN コードと同じくユニークなコードであるが、JAN コードを持たない商品も存在する。これら二つの商品コードは、Amazon.co.jp の商品コード検索 API を用いることにより、相互変換が可能である。

提案システムでは、Amazon.co.jp での商品ページ特定に ASIN コードを使用し、楽天市場と Yahoo!ショッピングでの商品検索には JAN コードを使用している。このため、JAN コードを持たない商品のユーザレビューは収集できない。JAN コードを持たない商品は提案システムへのユーザレビューの投稿と参考になるサイト情報の投稿のみ可能になる。

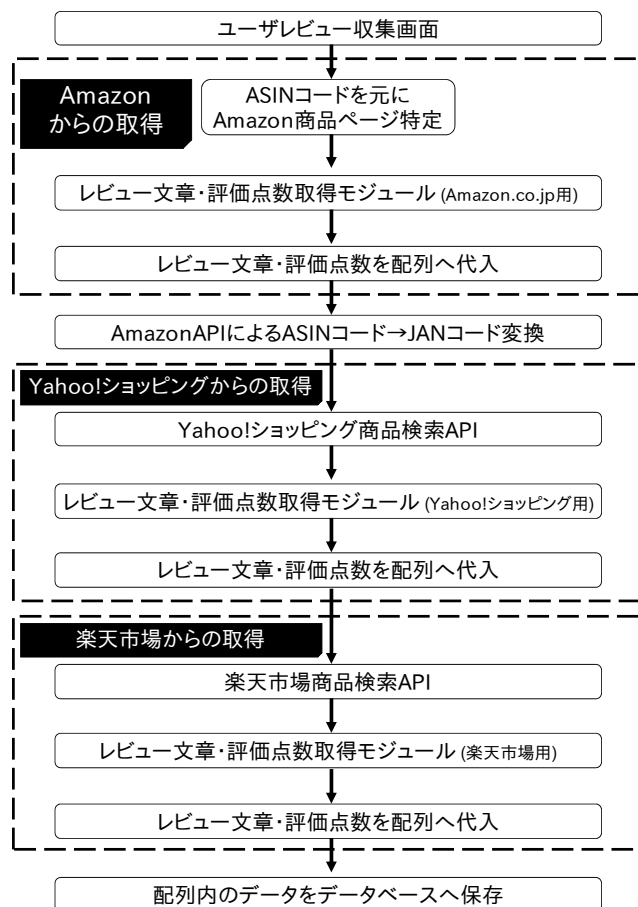


図 3 ユーザレビュー収集処理の流れ

Figure 3 The flow of user review collection processing.

6.4 ポジティブ・ネガティブ分析

ポジティブ・ネガティブ分析モジュールは入力されたレビュー文章を句点や各種記号で一文ごとに分割する。その後一文ごとに MeCab で形態素解析を行い、ナイーブベイズ分類器と訓練データデータベースによるポジティブ・ネガティブ判定と、学習データ履歴データベースとの比較によるポジティブ・ネガティブ判定を行い、結果を XML 形式で出力する (図 4)。

まとめページ表示モジュールはポジティブ・ネガティブ分析モジュールから受け取った結果を整形して円グラフとともに表示する。

まとめページ表示モジュールで結果を表示する際には、判定結果の文にマウスオーバー (マウスを重ねる) をすることで、元のレビュー文章全文を確認できるように工夫した (図 5)。これは、レビュー文章全体から一文を切り取って表示すると、全文を通して読んだときと意味合いが変わってしまうことへの対応である。

また、判定結果に誤りがあった場合には、閲覧者が 1 クリックで訂正できる機能を備えた。判定結果の訂正指示が行われた場合は、即時にポジティブ・ネガティブ判定モジュールが該当文章の再学習および、訓練データデータベース、学習データ履歴データベースへデータを格納し、以後の判定に役立てる。

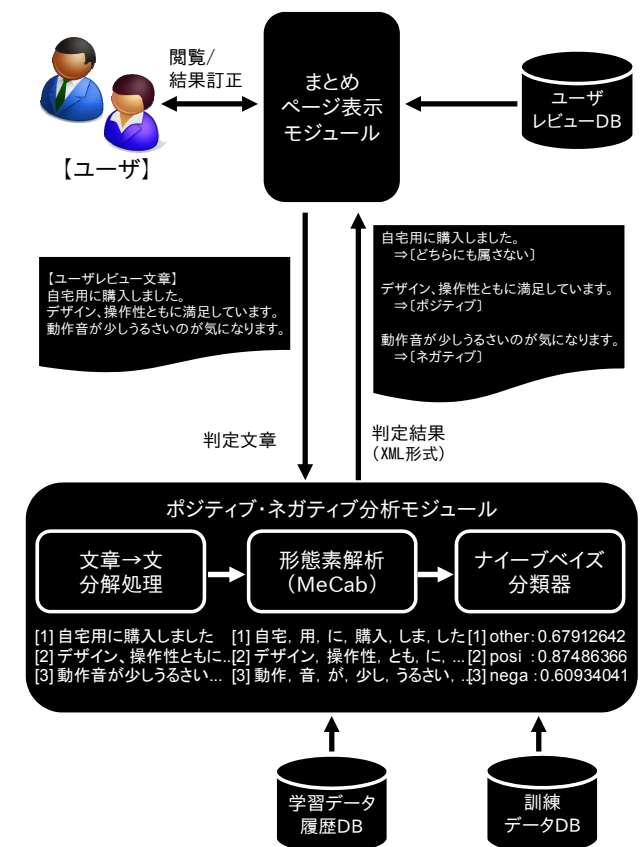


図 4 ポジティブ・ネガティブ分析モジュール構成

Figure 4 Positive and negative analysis module configuration.



図 5 ポジティブ・ネガティブ分析結果の例

Figure 5 Examples of positive and negative results of the analysis.

6.5 ナイーブベイズ分類器

ナイーブベイズ分類器は事前に与えられたデータをもとに、新しく与えられたデータを分類する機械学習の手法である。ナイーブベイズ分類器はその名前の通り、単純な計算を用いて分類を行うが、比較的精度も高く、プログラムに組み込んだ際にも処理時間が短く済むため、電子メールのスパムフィルタや、ウェブ上の文章のカテゴリ分類などに用いられている。

提案システムでは、入力されたユーザレビュー文章を、ナイーブベイズ分類器を用いて、「ポジティブ」「ネガティブ」「どちらにも属さない」の3つの評価カテゴリそれぞれに属する確率を計算する(図6)。

レビュー文章RがカテゴリCiに含まれる確率

$$P(Ci | R) = \frac{P(R | Ci) \times P(Ci)}{P(R)}$$

レビュー文章:R
 カテゴリ:Ci
 Rに含まれる単語がカテゴリCiの中に現れる確率(尤度):P(R|Ci)
 与えられたカテゴリが選ばれる確率(事前確率):P(Ci)



分母のP(R)はどのカテゴリでも同じになるため無視

$$P(Ci | R) = P(R | Ci) \times P(Ci)$$

図 6 ナイーブベイズ分類器によるカテゴリ分類の計算式
 Figure 6 Formula categorization by Naive Bayesian Classifier.

ナイーブベイズ分類器を用いてカテゴリ分類を行うには、まず単語ごとに分解する必要がある。提案システムでは、日本語に対応した形態素解析エンジンの一つであるMeCabを使用し、レビュー文章を単語ごとに分解した後、ナイーブベイズ分類器にてカテゴリ分類を行う(図7)。

形態素解析で使用する品詞は形容詞、形容動詞、感動詞、副詞、連体詞、接続詞、接頭詞、接尾詞、名詞、動詞である。

レビュー文章Rを形態素解析(n個の形態素Tに分解)

$$R = \{T_1, T_2, \dots, T_n\}$$

レビュー文章Rが「ポジティブC₁」に分類される確率

$$P(R|C_1) = P(T_1|C_1) \times P(T_2|C_1) \times \dots \times P(T_n|C_1)$$

図 7 形態素解析とナイーブベイズ分類器による計算式

Figure 7 Formula by morphological analysis and Naive Bayesian Classifier.

前述したようにナイーブベイズ分類器は単語ごとの出現確率を掛け合わせているため、学習データに存在しない単語がレビュー文章に含まれていた場合、出現確率が0となってしまう、最終的に導き出される確率も0になってしまう。これはゼロ頻度問題と呼ばれている。この問題を解決する方法として、あらかじめ一定の確率を与えておく加算スムージングと呼ばれる処理がある。提案システムでは、出現確率が0の単語に対して、仮の確率である0.5を加算して計算を行う。

7. 評価実験

ポジティブ・ネガティブ分析機能に関して、実際にECサイトに投稿されたユーザレビューを用いて、レビューの学習と判別実験を行った。

7.1 使用するデータ

提案システムは、ナイーブベイズ分類器を用いてポジティブ・ネガティブ分析を行う。前述したように、ナイーブベイズ分類器はあらかじめ学習データを与える必要がある。

評価実験では、あらかじめ与える学習データおよび、分析精度の評価を行うためのテストデータに「楽天データ」を利用した⁸⁾。

「楽天データ」は楽天技術研究所が大学や公的研究機関向けに無償提供しているデータセットであり、楽天株式会社保有する膨大な数のユーザレビューデータが含まれている。

7.2 評価実験 1 : 学習データ数およびカテゴリ領域の広さによる分析精度の比較

7.2.1 実験概要

ナイーブベイズ分類器は、一般的に学習データの数が多ければ分類精度が上がる。さらに、ユーザーレビューは商品ジャンルカテゴリによって、特徴的な語句や、頻出語句が変わる。例えば、食品や消耗品のジャンルカテゴリでは、同じ商品を繰り返し購入することを「リピート」「リピ」と呼ぶことがある。また、パソコン周辺機器や、家電製品のジャンルカテゴリでは機器が正しく動作しないことを表す「エラー」や「不良」という単語がよく使われる。このことから、細分化した商品ジャンルカテゴリごとにユーザーレビューの学習データを用意した方が、分析精度が向上すると考えられる。

本実験では、商品ジャンルカテゴリごとに学習データを作成し、さらに学習データ数を増やして複数回実験を行うことで、分類精度にどの程度の差が出るのか評価を行った。

7.2.2 実験手順

学習データの登録は、ユーザーレビューを手作業で「ポジティブ」「ネガティブ」「どちらにも属さない」の3つの評価カテゴリに分け、登録を行った。

各評価カテゴリの学習データが100件程度貯まるごとに、未学習のユーザーレビュー50件をナイーブベイズ分類器を用いてポジティブ・ネガティブ分析を行い、分類成功率、適合率、再現率を算出した。実験は計5回行った。

なお、評価実験に利用したユーザーレビューデータの提供元である楽天市場では、ルート直下のジャンルカテゴリが31個あり、末端のジャンルカテゴリは38,515個にもおよぶ(図8)。本実験では、ルート直下の「パソコン・周辺機器」ジャンルカテゴリおよび配下の「プリンタ・インク」、「複合機」、「キャノン」の計4ジャンルカテゴリに含まれるユーザーレビューを使用した。

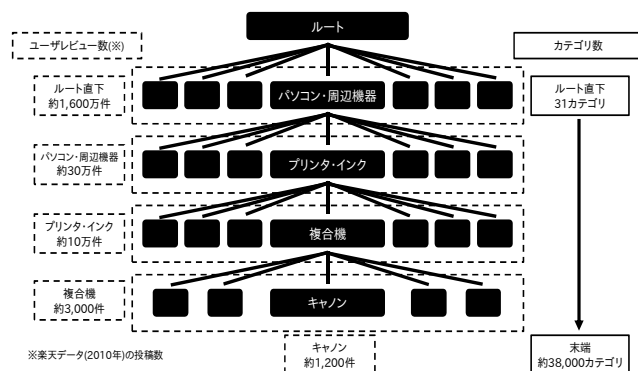


図 8 楽天市場のカテゴリ階層とユーザーレビュー投稿数
 Figure 8 Posts user reviews and hierarchical category Rakuten.

7.2.3 実験結果

「パソコン・周辺機器」および「キャノン」ジャンルカテゴリでの実験結果を表1~表8に示す。

なお、「キャノン」ジャンルカテゴリに関しては、2回目の評価実験を終えた段階で、ネガティブなユーザーレビューがほぼ無くなってしまったため、3回目以降の実験は中止した。

表 1 実験回数ごとの学習データ数(パソコン・周辺機器)
 Table 1 The number of training data and the number of experiments. (Category : Personal Computer and Peripherals)

	1回目	2回目	3回目	4回目	5回目
ポジティブ	101	200	301	401	507
ネガティブ	99	192	295	406	508
どちらにも属さない	99	204	300	400	504
合計	299	596	896	1,207	1,519

表 2 実験回数ごとの成功率(パソコン・周辺機器)
 Table 2 Success rate and the number of experiments. (Category : Personal Computer and Peripherals)

	1回目	2回目	3回目	4回目	5回目
成功	49%	53%	42%	51%	54%
失敗	51%	46%	54%	47%	41%
行分割ミス	0%	1%	4%	2%	5%
その他	0%	0%	0%	0%	0%

表 3 実験回数ごとの適合率(パソコン・周辺機器)
 Table 3 Compliance rate with the number of experiments. (Category : Personal Computer and Peripherals)

	1回目	2回目	3回目	4回目	5回目
ポジティブ	0.47	0.57	0.27	0.46	0.35
ネガティブ	0.2	0.13	0.65	0.36	0.5
どちらにも属さない	0.58	0.56	0.59	0.67	0.78

表 4 実験回数ごとの再現率(パソコン・周辺機器)
 Table 4 Recall a number of experiments. (Category : Personal Computer and Peripherals)

	1回目	2回目	3回目	4回目	5回目
ポジティブ	0.78	0.79	0.72	0.82	0.82
ネガティブ	0.07	0.15	0.3	0.13	0.21
どちらにも属さない	0.32	0.27	0.4	0.42	0.63

表 5 実験回数ごとの学習データ数 (キャノン)

Table 5 The number of training data and the number of experiments. (Category : Canon)

	1回目	2回目	3回目	4回目	5回目
ポジティブ	113	197	-	-	-
ネガティブ	101	182	-	-	-
どちらにも 属さない	87	204	-	-	-
合計	301	583	-	-	-

表 6 実験回数ごとの成功率 (キャノン)

Table 6 Success rate and the number of experiments. (Category : Canon)

	1回目	2回目	3回目	4回目	5回目
成功	56%	57%	-	-	-
失敗	36%	39%	-	-	-
行分割ミス	3%	3%	-	-	-
その他	4%	1%	-	-	-

表 7 実験回数ごとの適合率 (キャノン)

Table 7 Compliance rate with the number of experiments. (Category : Canon)

	1回目	2回目	3回目	4回目	5回目
ポジティブ	0.42	0.58	-	-	-
ネガティブ	0.28	0.13	-	-	-
どちらにも 属さない	0.68	0.69	-	-	-

表 8 実験回数ごとの再現率 (キャノン)

Table 8 Recall a number of experiments. (Category : Canon)

	1回目	2回目	3回目	4回目	5回目
ポジティブ	0.72	0.75	-	-	-
ネガティブ	0.15	1	-	-	-
どちらにも 属さない	0.73	0.4	-	-	-

7.2.4 実験考察

学習データ数を増やしても、分類成功率はほとんど上昇しない結果となった。また、分類成功率は50%前後にとどまった。

分類成功率と適合率に関しては、若干ではあるが、商品ジャンルを細分化して学習データを作成した方が向上した。

再現率に関しては、商品ジャンルを細分化して学習データを作成しても、向上することはなかった。

商品ジャンルを細分化して学習データを作成した結果、分類成功率と適合率が若干向上したが、その差は大きくな

かったことから、分類精度には形態素解析の精度など、学習した語彙の量以外の要素も関係していると考えられる。

7.3 評価実験 2 : ユーザレビュー閲覧までの所要時間の比較

7.3.1 実験概要

3つのECサイトでユーザレビューを閲覧する際に、これまでのようにECサイトを横断してユーザレビューを閲覧する場合と、提案システムを使用してユーザレビューの閲覧する場合とで、ユーザレビュー閲覧までの時間に差が出るのか評価を行った。

また、提案システムを使用することによって、ポジティブ・ネガティブなユーザレビューを探し出すまでの手順が改善できたのか評価を行った。

評価は42の商品でおこなった。なお、商品ジャンルは指定していない。

7.3.2 実験手順

3つのECサイトでユーザレビューを閲覧する手順と、提案システムを使用してユーザレビューを閲覧する手順を図9に示す。なお、提案システムでは、ECサイトからユーザレビューを収集済みであるか否かで、ユーザレビュー閲覧までの時間に差が出るため、収集済み状態、未収集状態をそれぞれケースI、ケースIIとして実験を行った。

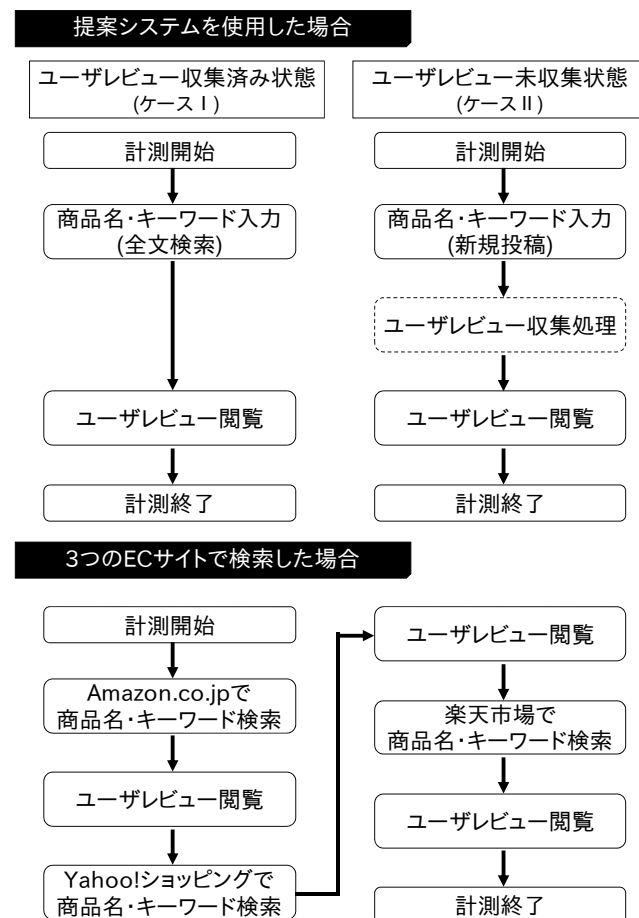


図 9 ユーザレビュー閲覧までの手順
 Figure 9 Steps to browse user reviews.

7.3.3 実験結果

提案システムのユーザレビュー収集済み状態（ケース I）では、42 品中 37 品でユーザレビュー閲覧までの時間の短縮を実現できた。短縮時間は最大で 4 分 8 秒であった。

提案システムのユーザレビュー未収集の状態（ケース II）では、42 品中 34 品でユーザレビュー閲覧までの時間の短縮を実現できた。短縮時間は最大で 3 分 39 秒であった。

また、ユーザレビューの件数が大量にある商品の場合、商品に対してポジティブあるいはネガティブなレビューのみを閲覧したい際には、評価点数などによるソート機能を使用したとしても非常に手間がかかる。

さらに、モール型の EC サイトでは、ユーザレビューの投稿を条件に通常より多いポイントの付与や、送料の優遇をおこなっているショップが存在し、このような特典を以て「ポイント〇倍はお得ですね」「送料無料でお得に購入できました」といった、商品に全く関係のないユーザレビューや、「迅速な対応に感謝します」「メールでの連絡が丁寧」といったショップの対応に対するユーザレビューが投稿されるケースがある。

提案システムのポジティブ・ネガティブ分析機能では、ユーザレビュー文章を一文単位に区切り、ポジティブとネガティブに分けて表示するため、このような商品に関係のないユーザレビューを省くことができ、ユーザレビュー閲覧の際の手間を解消できた（図 10）。

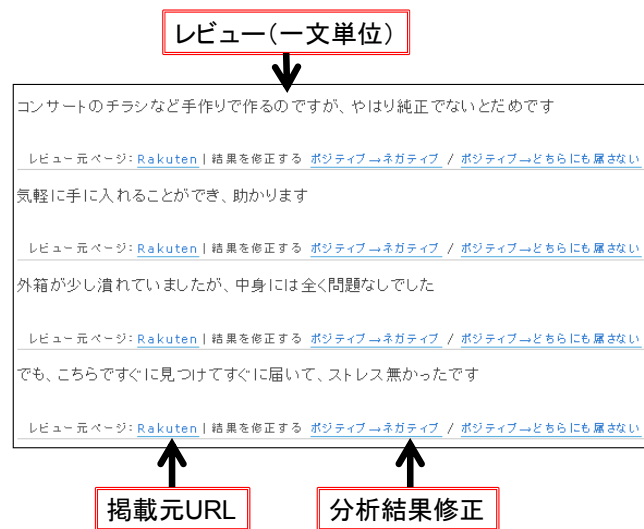


図 10 提案システムでユーザレビューを表示した例
 Figure 10 Example of Viewing reviews in the proposed system.

7.3.4 実験考察

EC サイトからのユーザレビュー収集を自動化することで、ユーザレビュー閲覧までの時間を削減することができた。

また、モール型 EC サイトの欠点である、「商品キーワー

ドに関連する他の商品が表示される」「ショップごとにユーザレビューが掲載される」「特典目当てのユーザレビューやショップの対応に対するユーザレビューが投稿される」といった点に対して、提案システムのまとめページを使用することで、これらの欠点を解消することができた。

8. おわりに

本論文では、EC サイトのユーザレビューに存在する問題点に着目し、それらの問題点を解決する Web アプリケーションを提案・開発した。

まず、ユーザレビューが複数のサイトに分散している問題点に対しては、大手 EC サイトからのユーザレビュー収集機能と手作業でのまとめ機能を提案した。

次に、現在の評価点数によるソートや絞り込み機能だけでは、ポジティブな評価やネガティブな評価だけを閲覧したい際に不便であるという問題点に対しては、ナイーブベイズ分類器を用いたポジティブ・ネガティブ分析機能を提案した。

また、ポジティブ・ネガティブ分析機能に関して、実際に EC サイトに投稿されたユーザレビューを用いて学習データの数と分析精度の関係、学習データを作成する際のジャンルカテゴリの大きさと分析精度の関係を実験した。

その結果、分類成功率は 50%前後にとどまる結果となった。ユーザレビュー全体から 50%程度の分別ができれば実用に耐えうると考えられるが、分類精度をさらに向上させるためには、分析アルゴリズムの改善が必要であろう。

ユーザレビュー閲覧までの手順の比較では、提案システムを使用することで、ユーザレビュー閲覧までの所要時間を削減でき、ユーザレビューの見やすさ・探しやすさに関しても、向上することができた。

今後は、分類成功率の向上を目指して、形態素解析器や、ナイーブベイズ分類器の精度改善、新しい分析アルゴリズムの模索を行いたい。

参考文献

- 1) 富士通総研：インターネットショッピング 2009, <http://jp.fujitsu.com/group/fri/report/cyber/report/shopping2009.html>
- 2) マクロミル：ネットショッピングの利用実態調査 2009, http://www.macromill.com/r_data/20090327netshopping/
- 3) 経済産業省：「消費者購買動向調査」～リーマンショック以降の日本の消費者の実像～, 2010 年 4 月 21 日, <http://www.meti.go.jp/press/20100421002/20100421002-2.pdf>
- 4) 社団法人日本通信販売協会：インターネット通販利用者実態調査 2008, <http://www.jadma.org/pdf/press/press20081117.pdf>
- 5) 清水隆太, 東基衛：評判情報検索のための意見文収集に関する研究, 第 70 回情報処理全国大会 講演論文集, pp.635-636(2008)
- 6) 辻井康一, 津田和彦：テキストマイニングを用いた宿泊レビューからの注目情報抽出方法, 情報処理学会デジタルプラクティス Vol.3 No.4, pp.289-296(2012)
- 7) 高精度日本語解析エンジン「なずき」
<http://www.nttdata-nazuki.jp/>
- 8) 楽天データ公開, <http://rit.rakuten.co.jp/rdr/>