

# ポジティブ・ネガティブ分析を備えた ユーザレビュー自動分類システムの提案

1185003 小林 亮

(指導教員 速水 治夫 教授)

## 1 はじめに

インターネットショッピングサイト（以下、ECサイトという）の多くにはユーザレビューと呼ばれる商品購入者の感想や意見を投稿・閲覧できる機能が備わっている。ユーザレビューは主に、商品の満足度を5つ星や100点満点での点数で評価し、その評価とともにレビュー文章を投稿する。

ECサイトの普及とともに商品選択時にユーザレビューを参考にするユーザが増加している<sup>[1][2][3]</sup>。しかし、現在は多くのECサイトや価格比較サイト、レビューサイトや個人のブログなど、インターネット上の様々なサイトにユーザレビューが分散しており、より多くのユーザレビューを閲覧したいユーザは、複数のサイト間を横断しなくてはならないという問題が存在している。また、点数評価によるソートや絞り込み機能だけでは、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。特にネガティブな評価は、ユーザレビューや口コミからしか得ることのできない貴重な評価であり、ユーザレビューをポジティブ・ネガティブに分別することは商品の販売者、購入者双方にメリットがあると考えられる。

本研究では、ユーザレビューのこれらの問題点を解決するWebアプリケーションを開発し、評価実験を行った。

## 2 研究のアプローチ

ユーザレビューの分散に対しては、大手ECサイトからのユーザレビューの収集機能と手作業でのまとめ機能を、ポジティブな評価・ネガティブな評価を閲覧するための機能として、収集したユーザレビューのポジティブ・ネガティブ分析機能を備えることで、効率よくユーザレビューを閲覧できるユーザレビュー自動分類システムを提案する。

## 3 関連研究および本研究との関係

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<sup>[4]</sup>。

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、数値化を行っており、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。

本研究の提案システムでは、汎用性を重視し、ナイーブベイズ分類器を用いてECサイトに投稿されたユーザレビューの学習と判別を行う。

## 4 提案システム概要

### 4.1 システム構成

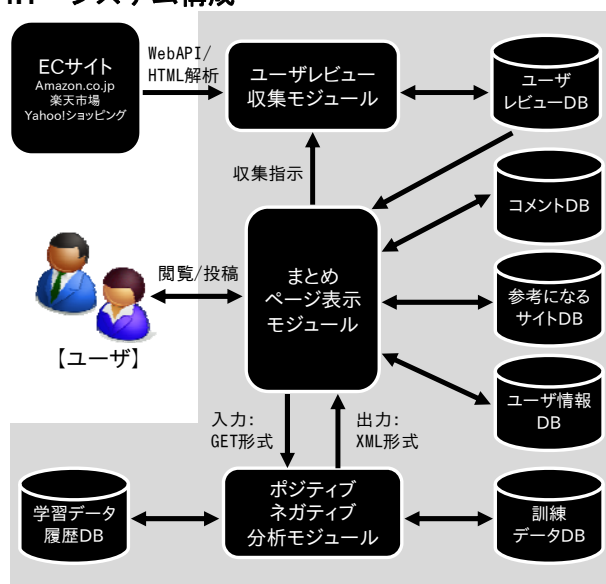


図1 提案システム構成（網掛け部が提案システム）

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。

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・ネガティブ分析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（図1）。

### 4.2 ユーザレビュー収集モジュール

ユーザレビュー収集モジュールは、Amazon.co.jp・Yahoo!ショッピング・楽天市場の3つのECサイト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。

### 4.3 ポジティブ・ネガティブ分析モジュール

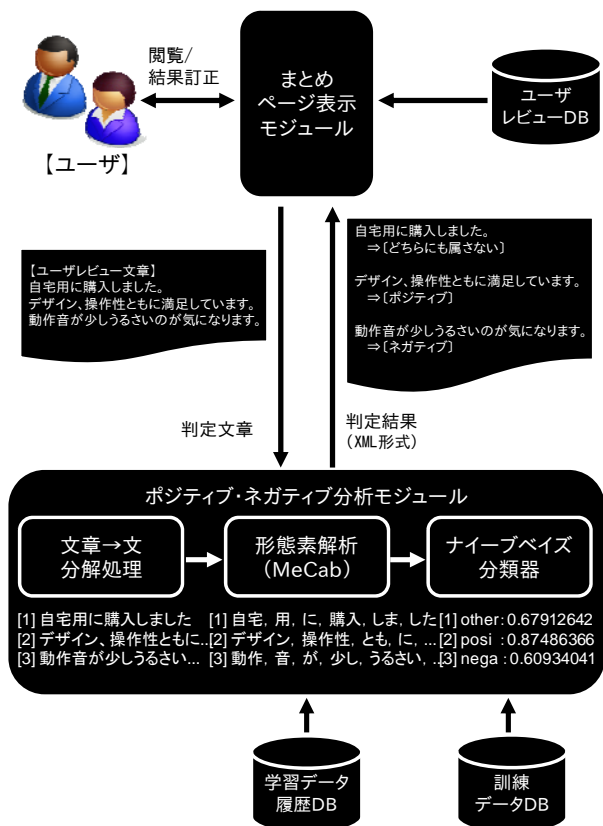


図 2 ポジティブ・ネガティブ分析の流れ

ポジティブ・ネガティブ分析モジュールは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 2)。

## 5 評価実験

### 5.1 ポジティブ・ネガティブ分析精度の評価

楽天株式会社が大学や公的研究機関向けに提供しているデータセットである「楽天データ」のユーザーレビューデータを使用して、ポジティブ・ネガティブ分析機能の精度評価を行った<sup>[6]</sup>。

精度評価では、学習データ数および商品ジャンルのカテゴリ階層ごとに精度にどのような変化が現れるのか評価を行うため、楽天市場の商品ジャンルカテゴリのうち、ルート直下に位置する「パソコン・周辺機器」カテゴリ配下のユーザーレビューを実験の対象とした。

結果、分類成功率は 50%前後にとどまる結果となった。また、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、ほとんど向上することはなかった。

### 5.2 ユーザレビュー閲覧までの手順の比較

3つの EC サイト (Amazon.co.jp・Yahoo!ショッピング・楽天市場) を横断してユーザー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ーレビューの閲覧する場合とで、ユーザーレビュー閲覧までの時間に差が出るのか評価を行った。

42の商品で評価をおこなった結果、提案システムのユーザーレビュー収集済み状態では 42品中 37品で、提案システムのユーザーレビュー未収集の状態でも 42品中 34品でユーザーレビュー閲覧までの時間の短縮を実現できた。

さらに、モール型の EC サイトでは、ポイントの付与や、送料の優遇を目当てに商品に全く関係のないユーザーレビューや、ショップの対応に対するユーザー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、ユーザー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ーレビューを省くことができ、ユーザーレビュー閲覧の際の手間を解消できた。

## 6 おわりに

本研究では、EC サイトのユーザー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開発した。

評価実験の結果、分類成功率は 50%前後にとどまる結果となったが、ユーザーレビュー全体から 50%程度の分別ができれば実用に耐えうると考えられる。ユーザー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、ユーザーレビュー閲覧までの時間を削減でき、ユーザー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/](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] 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)
- [5] 楽天データ公開, <http://rit.rakuten.co.jp/rdr/>