

E コマースにおける商品レビュー分析に基づく 不満解決商品推薦手法の提案

吉川 耀敬[†] 王 元元[†] 河合由起子^{††,†††}

[†] 山口大学工学部知能情報工学科 〒755-8611 山口県宇部市常盤台 2-16-1

^{††} 京都産業大学情報理工学部 〒603-8555 京都府京都市北区上賀茂本山

^{†††} 大阪大学サイバーメディアセンター 〒567-0047 大阪府茨木市美穂ヶ丘 5 番 1 号

E-mail: [†]{v079ff, y.wang}@yamaguchi-u.ac.jp ^{††}kawai@cc.kyoto-su.ac.jp

あらまし 現在、インターネットの普及とともに E コマースが注目され、ネットショッピングにおける商品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、購入履歴からその後に必要となる付属品といった関連商品も推薦されるが、商品に不満があった場合はレビューの評価値のみでは判断できず、そのものの問題点を解決する代替品は推薦されない。そのため、既存サービスでは、ユーザは規定項目（たとえば、値段や対象者）といった評価尺度を選択する必要がある、新たなニーズ（問題）への対応が困難である。そこで本研究では、楽天市場の商品レビューから抽出されるネガティブな情報とポジティブな情報を用いて、商品に対する固有の問題点を抽出し、それらを解決する代替品となる商品を発見し推薦する手法を提案する。本論文では、ゲーム商品を対象にし、レビュー文を解析してその中から低評価のネガティブな特徴ベクトル（不満情報）と高評価のポジティブな特徴ベクトル（満足情報）を生成し、不満情報と満足情報の類似性を算出することで、不満を解決するレビューを発見する。これにより、閲覧しているゲームに対して、不満情報とその不満を解決するレビューを持つ他のゲームをユーザへ推薦できる。本稿では、商品レビュー分析に基づく不満情報抽出および不満解決可能な代替品推薦手法について述べ、不満情報および不満を解決できる情報抽出の有効性を検証する。

キーワード E コマース, レビュー分析, 不満情報, 商品推薦

1. はじめに

近年、E コマースの発達により、商品レビューを参考に多種多様な商品を比較し、自分の趣味・嗜好に合わせた商品を購入することが主流となった。楽天市場や Amazon など多くの E コマースサイトでは、ユーザの商品選択の手段としてレビュー情報は重要であり、E コマースのレビュー分析に関する研究が広く行われている。しかし、基本的に評価の高い商品やユーザの興味にあった商品といった、高評価のポジティブなレビューに基づいた関連商品を比較することが多く、低評価のネガティブなレビューは無視される傾向にある。

そこで、本研究では商品レビューから抽出されるネガティブな情報とポジティブな情報を用いて、商品に対する固有の問題点を抽出し、それらを解決可能な代替品となる商品を発見し推薦する手法を提案する（図 1）。提案手法では、まず、「楽天公開データ」^(注1)より、楽天市場における各商品のレビュー情報を取得し、低評価レビューと高評価レビューを分類する。

次に、低評価のネガティブな情報と高評価のポジティブな情報を抽出する。各商品の低評価のネガティブな特徴語（不満情報）と高評価のポジティブな特徴語（満足情報）抽出は、各レビューに示される評価値で判断する。評価値は一般的に 5 段階評価となっている。しかしながら、低（高）評価値であっても

実際のレビュー文章では、「××は良くはなかったが、○○は良かった」というように、高（低）評価の単語も含まれていることが多い。そこで、本研究では、低評価のネガティブな特徴語の値から高評価のポジティブな同一の特徴語の値との差分を算出し、ネガティブな評価と成り得る特徴語を抽出する。同様に、高評価のポジティブな特徴語の値から低評価のネガティブな同一の特徴語の値との差分を算出し、ポジティブな評価と成り得る特徴語を抽出する。なお、特徴語の値は *TF-IDF* 手法に基づき算出する。

最後に、商品の低評価のネガティブな特徴語（不満情報）と高評価のポジティブな特徴語（不満解決情報）との類似度を算出し、商品の不満となっている特徴語が、逆にポジティブ高評価されている他商品の推薦を行う。不満対象が高評価としてレビューに記された商品を推薦することで、不満を解決する代替商品の推薦が可能となる。

本論文の構成は以下のとおりである。次章では商品推薦システムに関する研究や不満情報を用いた研究を紹介し、3 章では、商品レビューを分析し、不満を解決する商品の推薦について説明する。4 章では、実データを用いた提案手法の評価実験について述べる。最後に、5 章でまとめと今後の課題について述べる。

2. 関連研究

E コマースにおける購買行動の促進を目的とした研究もいくつかあげられる。中野ら [1] はレビューの意見ペアを抽出し、商

(注1) : <https://www.nii.ac.jp/dsc/idr/rakuten/rakuten.html>

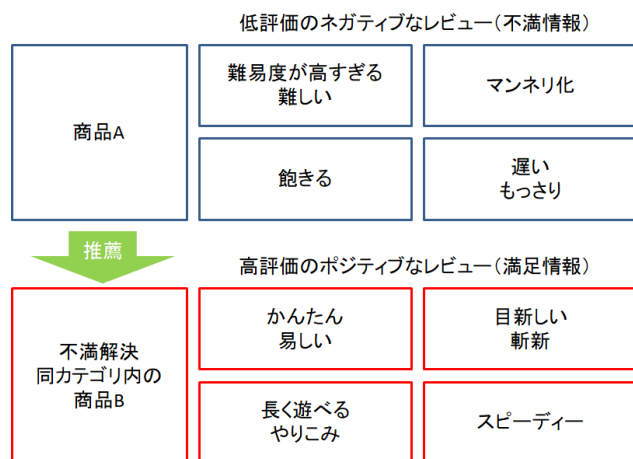


図 1 不満解決商品推薦システム

品レビューの要約を可能にしたユーザの購買行動促進システムを提案した。山本ら [2] は商品レビューに存在する不満意見と好評意見を考慮し、ユーザのニーズに適した商品推薦システムを提案した。本研究では、商品の低評価レビューの内容は、商品の評価が目的であるとし、純粋な不満意見でないと考える。そこで、レビューのポジティブとネガティブの判別によって低評価のネガティブな情報を用いて、より活用的な不満解決商品推薦を提案する。

また、本研究で扱うテキスト分析に関しても多く研究もされている。小林ら [3] はテキストによる評価表現に関して、領域・分野に依存する文章表現の効率的に収集した。平山ら [4] はレビューの極性分析を行い、レビュー情報を可視化するシステムを提案した。相澤 [5] は語と文書の共起性に基づきテキスト分析に多く用いられる TF - IDF 手法の改良法を検証した。小林ら [6] はレビュー文内に出現する単語の属性を分析、収集しレビューの評価値を算出した。東山ら [7] は名詞の評価極性について検証した。このように、テキスト分析に関する研究は多く、レビュー情報を用いた検証も多い。本研究は、 TF - IDF を用いた確率と情報量を考慮したレビュー分析をに基づく不満解決商品推薦手法を提案する。

これまでに不満調査データセットを用いた研究はいくつか行われてきた。三澤ら [8], [9] は不満調査データに特化したコーパスの構築と分析を行った。また、不満調査データセットを用いてユーザの意見をグループ分けするクラスタリングを検証した [10]。北山ら [11] は不満調査データセットを用いて職種・年齢別のグループごとに分けた不満対象の可視化を行った。末廣ら [12] は Bag-of-Words を用いて不満調査データの素性化を行い、意味情報を分析した。Hayashi ら [13] は不満調査データとレビューを用いた不満解決商品推薦手法を提案した。本研究は商品レビューから抽出される不満情報をを用いた情報推薦手法を提案するので、その点が異なる。

3. レビュー分析に基づく不満解決商品推薦

3.1 不満情報抽出

本研究では、商品レビューを分析し商品ごとの不満を抽出す

る。対象商品の評価値が 3 以下のものを低評価とし、それに付与されているレビューを低評価レビューとする。しかし、低評価レビューにも高評価の内容が含まれていることがある。たとえば、「購入前は楽しみだったが、実際にやってみたらがっかりだった。」といった文章では、「楽しみ」という単語がポジティブな意味合いを持つ場合には低評価レビューからは除外した。そして、低評価レビューのうちにネガティブな情報を不満情報とする。不満情報に特に多く出現し、さらに高評価レビューに出現していない単語を不満情報の特徴語として抽出する。なお、単語ごとのポジティブ、ネガティブ判別については菅・鈴木研究室にて公開されている日本語評価極性辞書^(注2)を用いた。以下の TF - IDF 算出式に基づき商品に対する不満を表す特徴語 wn の要素値として算出する。算出した要素値から不満を示す特徴語ベクトル $v_F = (wn_1, \dots, wn_i, \dots, wn_m)$ を生成する。

$$TF = \frac{\text{商品の不満情報の単語 } wn \text{ の出現数}}{\text{商品の単語総数}}$$

$$IDF = \log \frac{\text{商品のレビュー総数}}{\text{単語 } wn \text{ が出現したレビュー数}}$$

TF は、任意の商品の不満情報の単語頻度を表す。 IDF は、任意の商品のレビューを 1 つの文書とした時の文書頻度の逆数を表す。

3.2 ポジティブ評価の商品抽出

対象商品に付与されている評価値が 4 および 5 のものを高評価として、それに付与されているレビューを高評価レビューとする。しかし、高評価レビューには、低評価の内容が含まれていることがある。たとえば、「××は良くなかったが、○○はよかった」といった文章である。このようなノイズを除くため、不満情報抽出と同様に、ネガティブな意味合いをもつ単語については、高評価レビューに付与されている単語であっても、高評価レビューからは除外した。そして、高評価レビューのうちにポジティブな情報を満足情報とする。その中の特徴語から不満情報に対応できる不満解決情報を抽出する。なお、単語ごとのポジティブ、ネガティブ判別については 3.1 節と同様に菅・鈴木研究室にて公開されている日本語評価極性辞書を用いた。特徴語算出には 3.1 節と同様に TF - IDF 算出式を用いて商品に対するポジティブ評価を表す特徴語 wp の要素値として算出する。算出した要素値から不満を示す特徴語ベクトル $v_R = (wp_1, \dots, wp_i, \dots, wp_n)$ を生成する。

3.3 不満解決商品の推薦

不満に対して、解決する商品推薦をするために、不満情報を持つ商品とポジティブ評価がされた商品の比較が必要となる。この節では、前節で求めた 2 種類の特徴ベクトルの類似度算出について述べる。

ここで、対象とする任意の商品の不満情報と同カテゴリのポジティブな高評価レビューをもつ商品の類似度が高いほど、共通の特徴語を持っていることになる。不満情報の特徴語と類似度の高い特徴語を持つポジティブな高評価レビューを持つ商品

(注2) : <http://www.cl.ecei.tohoku.ac.jp/index.php?Open%20Resources%2FJapanese%20Sentiment%20Polarity%20Dictionary>

表 1 対象とした商品

記号	商品名
A	New スーパーマリオブラザーズ wii (任天堂)
B	Wii Sports Resort (任天堂)
C	太鼓の達人 みんなでパーティ☆3代目! (バンダイナムコゲームス)
D	桃太郎電鉄 2010 戦国・維新ヒーロー大集合! の巻 (ハドソン)
E	星のカービィ wii (任天堂・HAL 研究所)

が、その不満を解決する商品となる。

類似度算出には、3.1 節と 3.2 節で求めた 2 種類の特徴ベクトルがある。これらのベクトルの類似度を算出するために、以下の式に表されるコサイン類似度を用いる。

$$Sim(v_F, v_R) = \frac{\sum wn_m \cdot wp_n}{\sqrt{\sum (wn_m)^2} \cdot \sqrt{\sum (wp_n)^2}}$$

ここで v_F は、レビュー情報から算出した商品ごとの低評価のネガティブな特徴ベクトルを示し、 v_R は、レビュー情報から算出した商品ごとの高評価のポジティブな特徴ベクトルを示している。本研究では、同カテゴリ内において、不満を解決する商品を推薦するため、ここでの v_F と v_R も同商品カテゴリの必要がある。まず、コサイン類似度を算出するために同カテゴリ内において、2 種類の特徴ベクトルの次元数を揃える、また、 $TF-IDF$ 値の算出方法も異なるため、それぞれ正規化をする。その後、コサイン類似度を算出し、不満に対して類似度の高い特徴語を持つポジティブな高評価の商品を求める。

4. 評価実験

4.1 データセット概要

提案手法で利用する「楽天公開データ」は、楽天株式会社が所持する楽天市場の商品データとレビューデータを国立情報学研究所 (NII) によって提供されたものである。これは 2010 年 8 月 4 日から 2014 年 4 月 1 日において、全商品データ約 1 億 5600 万件、レビューデータ約 6400 万が含まれる。各商品に付与されたレビューおよび評価値 (5 段階評価) を用いる。

本研究では楽天市場のデータを用いて低評価のネガティブな特徴語 (不満情報) および高評価のポジティブな特徴語 (満足情報) を抽出した。ただし、「面白い」や「おもしろい」といった、まったく同じ単語でも表記の違いから別の単語として抽出されたものについては、 $TF-IDF$ 値が高い一方のみを表記した。対象は、商品レビューが特に多く付与されていた 5 件の商品とした (表 1)。

4.2 実験 I : 不満情報抽出の検証

実験 I では、レビュー情報から商品の不満情報を抽出し、検証した。表 2 に対象とした商品から抽出した不満情報である低評価のネガティブな特徴語上位 20 件とその $TF-IDF$ 値の結果を示す。

4.3 実験 II : 不満解決消費抽出の検証

実験 II では、コサイン類似度に基づき対象とした商品の不満を解決する商品を抽出し、検証した。対象とした商品の不満を解決する商品上位 5 件から抽出した満足情報である高評価のポジティブな特徴語上位 20 件とその $TF-IDF$ 値の結果をそれぞ

表 2 対象商品の不満情報

商品	抽出した不満情報の特徴語上位 20 件 ($TF-IDF$ 値)
A	コマーシャル (0.335), 見る (0.335), とおり (0.335), 商品 (0.264), 忙しく (0.238), くらい (0.238), 眠る (0.238), 親子 (0.220), 過ぎ (0.215), 人 (0.211), 難易 (0.208), 用 (0.208), ほしい (0.208), 値段 (0.196), 小学校 (0.196), 年生 (0.196), 難しい (0.196), ぶり (0.185), 家族 (0.172), 最初 (0.167)
B	最悪 (0.376), たく (0.376), 利用 (0.307), ください (0.289), 覚える (0.235), ショップ (0.215), 倍 (0.212), ハマる (0.209), ピンク (0.209), 付属 (0.203), プレー (0.198), 画像 (0.198), 星 (0.198), 操作 (0.197), 対応 (0.197), 難しい (0.192), 飽きる (0.192), 大変 (0.192), Christmas (0.188), 開封 (0.188)
C	安かった (0.278), 連絡 (0.229), プレイ (0.222), 同時 (0.222), お願い (0.196), 残念 (0.189), たち (0.186), 夢中 (0.186), キャンセル (0.180), 難しい (0.175), 取る (0.175), 人気 (0.159), 新しい (0.159), 出る (0.159), ほしく (0.159), 主人 (0.159), 過ぎる (0.159), 多く (0.159), 安い (0.155), 性 (0.142)
D	PLAY (0.278), 双六 (0.238), ハズレ (0.238), 移動 (0.208), 兼ね (0.208), シリーズ (0.196), たくさん (0.196), マリオ (0.196), 高かった (0.175), 当たった (0.175), 忘年会 (0.175), ハード (0.175), 戦国 (0.167), 崩し (0.167), 思う (0.167), 値段 (0.159), 家族 (0.155), 子 (0.155), DS (0.145), 版 (0.145), 勉強 (0.140),
E	対応 (0.224), 今後 (0.224), 買う (0.196), 早く (0.196), 点 (0.184), はまった (0.184), 駄目 (0.184), 届く (0.174), 利用 (0.174), 買った (0.149), こども (0.143), くれる (0.143), 無視 (0.132), もらえ (0.123), 日 (0.116), 値段 (0.116), みたい (0.113), 欲しい (0.111), 難易 (0.108), 息子 (0.106)

表 3 対象商品 A の不満解決商品

順位	不満解決商品 (コサイン類似度)	抽出した満足情報の特徴語上位 20 件 ($TF-IDF$ 値)
1	SD ガンダム ジェネレーションウォーズ (0.693)	シリーズ (0.252), 楽しみ (0.173), 大 (0.173), 画面 (0.173), wii (0.173), はまり (0.173), 中 (0.173), 久しぶり (0.163), 出 (0.163), 言う (0.163), 年々 (0.163), 進化 (0.163), 夫 (0.154), 面白い (0.152), 今回 (0.152), フォルム (0.139), たまらない (0.139), ジェネ (0.139), 通常 (0.132), 若干 (0.132)
2	ドラゴンボール 天下一大冒険 (0.681)	長く (0.250), 早かった (0.233), ハマリ (0.206), 綺麗 (0.206), ナイショ (0.206), 遊べ (0.200), 買い (0.187), 所 (0.175), 良い (0.175), 物 (0.175), 機会 (0.175), 予定 (0.175), クリア (0.171), 以上 (0.165), 発送 (0.160), 面白く (0.159), 映画 (0.159), 観 (0.159), 場面 (0.159), 子ども (0.146)
3	大乱闘スマッシュブラザーズ X (0.654)	ずっと (0.442), 面白く (0.442), 危険 (0.442), 気に入る (0.374), 盛り上がり (0.347), 必須 (0.347), スンチャク (0.347), ものすごく (0.347), はまり (0.342), 見た目 (0.324), 難しい (0.324), 喜ば (0.324), 数 (0.324), 最近 (0.324), たくさん (0.321), 迅速 (0.313), 届く (0.298), 良く (0.298), 登場 (0.289), いただき (0.289)
4	スーパーマリオギャラクシー 2 (0.653)	名作 (1.051), あたり (0.877), 良い (0.790), 訳 (0.683), 飛び立つ (0.621), 興味 (0.621), たち (0.606), ナイショ (0.569), ほしい (0.569), 念願 (0.569), 嫁さん (0.569), 年齢 (0.569), 間々 (0.569), 本日 (0.559), 宇宙 (0.558), 冒険 (0.558), めちゃくちゃ (0.558), 定番 (0.558), 試し (0.558), おもしろく (0.544)
5	スーパーマリオコレクション スペシャルパック (0.652)	着き (0.585), だんな (0.585), キャラクター (0.573), がる (0.531), 派 (0.531), ラッキー (0.515), はやく (0.487), 程 (0.487), 楽しかった (0.471), 物 (0.468), 怒り (0.450), 笑え (0.450), 練習 (0.432), 最近 (0.432), ねだら (0.418), ハマリ (0.418), 堪能 (0.418), 得意 (0.418), 丁寧 (0.404), 初代 (0.396)

れ表 3～表 7 に示す。

4.4 考察

実験 I と II の結果、ゲームによってユーザが求めていることが浮き彫りになった。全体的な傾向としては親子を主に、家族で一緒に遊ぶといった内容が、多くのレビュー内に散見され、不満解決商品もそれに付随するものが多く抽出された。これは、1 つの対象商品内だけでなく、ほぼすべての対象商品に該当し

表 4 対象商品 B の不満解決商品

順位	不満解決商品（コサイン類似度）	抽出した満足情報の特徴語上位 20 件（ <i>TF-IDF</i> 値）
1	ファミスタ DS 2010 (0.716)	使い (0.393), 常態 (0.393), よかった (0.328), <u>パパ</u> (0.328), 言った (0.328), おもしろい (0.324), やすい (0.283), ファミガシャ (0.281), 送料 (0.270), 込み (0.270), ポイント (0.270), 最高 (0.270), ファミコン (0.270), 即納 (0.262), とも (0.262), 通り (0.262), 世代 (0.255), なし (0.246), 酷 (0.231), すぎ (0.231)
2	イナズマイレブン 3 世界への挑戦!!スパーク (0.706)	イナズマイレブンスパーク (0.960), 最高 (0.844), おもしろい (0.594), 言った (0.576), いこう (0.524), 休み (0.501), 親戚 (0.480), 素早い (0.466), 最高 (0.461), 中学生 (0.461), 昨日 (0.443), 昼前 (0.443), 熱意 (0.443), 負け (0.443), ラクラク (0.443), 使った (0.422), 電機 (0.411), 屋 (0.411), 安価 (0.411), イナズマイレブン (0.411)
3	マリオ&ソニック AT バンクーバーオリンピック (0.698)	提供 (0.436), たい (0.434), 終了 (0.396), 間近 (0.396), 楽しい (0.396), もらい (0.396), 嬉しかった (0.363), 親子 (0.363), 良く (0.311), 星 (0.311), 四つ (0.311), おもしろい (0.311), 遊べる (0.311), 毎晩 (0.311), バトル (0.311), 激 (0.311), 喜ん (0.305), はまった (0.305), 感動 (0.290), 付いた (0.290)
4	Wii Fit Plus (0.694)	相方 (0.663), 大 (0.606), 敏速 (0.603), おとそ (0.553), <u>ボディ</u> (0.553), <u>づくり</u> (0.553), 付き合え (0.553), 1 つ (0.553), <u>頑張</u> り (0.529), 撃退 (0.510), 良し (0.510), 感覚 (0.507), 最高 (0.495), 作れる (0.495), スリム (0.495), 兼ね (0.457), 不満 (0.457), 効く (0.457), つかえ (0.457), <u>リハビリ</u> (0.443)
5	ファミリーフィッシング (0.675)	ほしい (0.611), 燃え (0.558), 毎晩 (0.508), 有り難う (0.508), 御礼 (0.508), 申し上げ (0.508), お正月 (0.489), 皆 (0.489), 豆 (0.465), 期日 (0.465), 過ごせ (0.465), 盛り上がり (0.465), 弟 (0.445), たかい (0.429), ものすごく (0.408), 間に合った (0.408), 遊び (0.382), おすすめ (0.382), 無事 (0.374), 出来る (0.350)

表 5 対象商品 C の不満解決商品

順位	不満解決商品（コサイン類似度）	抽出した満足情報の特徴語上位 20 件（ <i>TF-IDF</i> 値）
1	太鼓の達人 ちびドラゴンと不思議なオーブ (0.701)	欲しい (0.572), 包装 (0.531), お買い得 (0.510), イイ (0.510), 入り (0.464), たのしかった (0.425), ゴールデンボンバー (0.425), おもしろい (0.425), すすめ (0.425), 欲しかった (0.401), 手 (0.401), 大好き (0.367), 内緒 (0.364), スゴク (0.340), 仕入れ (0.340), 友達 (0.333), 楽しく (0.333), 家族 (0.333), 親 (0.319), 小さい (0.319)
2	太鼓の達人 Wii 決定版 (0.699)	続く (0.400), 問題 (0.369), 即決 (0.369), 配達 (0.369), 十分 (0.343), どこ (0.342), 早く (0.335), あげる (0.320), 肝心 (0.320), ゆう (0.320), 内容 (0.316), 無い (0.316), 値段 (0.316), 迅速 (0.316), 頂き (0.316), 素晴らしい (0.309), おもしろい (0.308), 新旧 (0.300), こども (0.300), 軒並み (0.300)
3	JUST DANCE Wii (0.698)	幅広く (0.497), スムーズ (0.475), ノリ (0.432), 懐かし (0.414), ダンレゴ (0.414), 思い出す (0.414), 遊べる (0.382), メチャクチャ (0.382), すぎ (0.355), エクササイズコース (0.355), 無理 (0.355), 楽しい (0.347), 大人 (0.336), 親 (0.331), 賑やか (0.331), 早い (0.329), 難しい (0.323), 中学年 (0.311), パート (0.311), 暖房 (0.311)
4	太鼓の達人 Wii 超ごうか版 (0.671)	親子 (0.317), 達 (0.317), 楽しく (0.296), 遊ぶ (0.296), 最速 (0.278), 早かった (0.278), 曲目 (0.261), 勉強 (0.250), 夢中 (0.250), こども (0.250), 嬉しい (0.239), 顔 (0.236), みんな (0.231), 喜ぶ (0.225), 欲しかった (0.223), 新作 (0.222), 機会 (0.212), 利用 (0.212), 問題 (0.212), 対戦 (0.212)
5	みんなのリズム天国 (0.669)	なし (0.899), 問題 (0.775), 感謝 (0.711), 歳児 (0.638), 取引 (0.582), 満載 (0.533), 待ち遠しい (0.519), 一言 (0.492), たつ (0.492), 踊り (0.492), 深い (0.492), 無事 (0.475), 買える (0.457), お父さん (0.457), 下手くそ (0.457), 定価 (0.442), うれしい (0.442), とどき (0.442), 是非 (0.442), 取り合い (0.442)

表 6 対象商品 D の不満解決商品

順位	不満解決商品（コサイン類似度）	抽出した満足情報の特徴語上位 20 件（ <i>TF-IDF</i> 値）
1	ぶよぶよ 7 スペシャルプライス Wii 版 (0.748)	発売 (0.296), 当初 (0.296), 欲しかった (0.296), ベスト (0.296), 大変 (0.296), <u>定価</u> (0.257), <u>安かった</u> (0.257), <u>無料</u> (0.257), くらい (0.226), 絶賛 (0.226), 満足 (0.226), 盛り上がり (0.226), ぶよぶよをやりましたが (0.226), はまった (0.226), 前作 (0.214), 向上 (0.214), 期待 (0.214), 事 (0.214), <u>プライス</u> (0.212), メール (0.210)
2	THE テーブルゲーム ～麻雀・囲碁・将棋・カード・ リバーシ・五目ならべ～ (0.723)	時間 (0.268), 値段 (0.268), テーブル (0.268), 必要 (0.268), 幼児 (0.268), がんばれ (0.268), <u>WII</u> (0.268), 考え (0.268), 与え (0.268), 良 (0.230), 母親 (0.215), 頼ま (0.215), 早く (0.215), <u>暇</u> (0.215), 届い (0.215), <u>持て余</u> し (0.215), ある (0.210), 時 (0.210), 最高 (0.201), こちら (0.189)
3	桃太郎電鉄 16 北海道大移動の巻！ みんなのおすすめセクション (0.712)	歴史 (0.241), くる (0.241), 面白かった (0.241), やつ (0.241), <u>パーティー</u> (0.206), 最速 (0.206), ヒーロー (0.183), ファン (0.170), <u>安い</u> (0.170), 始め (0.161), 持った (0.157), <u>高かった</u> (0.131), ハズレ (0.131), 無し (0.131) 代々 (0.131), <u>買った</u> (0.129), 一つ (0.128), 大 (0.122), 楽しみ (0.122), うれしい (0.120)
4	カラオケ JOYSOUND Wii SUPER DX (0.694)	おもしろく (0.454), <u>みなさん</u> (0.454), 大人 (0.413), 幼稚園 (0.413), LAN (0.379), めちゃくちゃ (0.379), 速い (0.349), もらった (0.325), <u>手頃</u> (0.325), DS (0.321), 会 (0.313), 損 (0.296), 買う (0.296), おすすめ (0.293), <u>無線</u> (0.287), ワイヤレス (0.284), USB (0.284), 後ろ (0.284), カラ (0.284), 無事 (0.284)
5	いただきストリート Wii (0.693)	おもしろい (0.495), はず (0.495), <u>こども</u> (0.445), 良い (0.374), たのしみ (0.371), カラダ (0.371), 体験 (0.371), 型 (0.371), <u>髭</u> (0.318), 今日 (0.318), <u>安価</u> (0.318), くれる (0.313), おもしろく (0.313), <u>お買い得</u> (0.313), よい (0.304), 売り切れ (0.297), 買った (0.296), おすすめ (0.280), 楽しく (0.279), 夢中 (0.278)

表 7 対象商品 E の不満解決商品

順位	不満解決商品（コサイン類似度）	抽出した満足情報の特徴語上位 20 件（ <i>TF-IDF</i> 値）
1	マリオカート Wii (0.766)	早い (0.226), 家族 (0.198), 面白い (0.198), 大 (0.198), 喜び (0.198), クリスマス (0.149), 楽しん (0.149), 普通 (0.149), たち (0.149), <u>子ども</u> (0.146), CM (0.132), 安く (0.132), 見 (0.132), 満足 (0.132), いつ (0.121), <u>大勢</u> (0.121), やる (0.121), 買わ (0.121), 遊べる (0.121), net (0.116)
2	大乱闘スマッシュブラザーズ X (0.759)	すっごく (0.442), 面白く (0.442), 危険 (0.442), 気に入 (0.374), 盛り上がり (0.347), 必須 (0.347), ムンチャク (0.347), ものすごく (0.347), はまり (0.342), 見た目 (0.324), 数 (0.324), 難しい (0.324), 喜ば (0.324), 最近 (0.324), <u>たくさん</u> (0.321), 迅速 (0.313), 届く (0.298), 良く (0.298), 登場 (0.289), いただき (0.289)
3	毛糸のカービィ (0.757)	着 (0.499), 保育園 (0.499), <u>年長</u> (0.499), プレゼント (0.499), 素早く (0.461), クリスマスプレゼント (0.428), <u>人気</u> (0.384), 死な (0.382), カービィシリーズ (0.378), トイザラス (0.378), 親戚 (0.378), 預け (0.374), 久しぶり (0.374), 姪 (0.365), 孫 (0.354), 最後 (0.354), 入った (0.353), 日用 (0.353), 親子 (0.353), <u>すぎ</u> (0.350)
4	スーパーマリオギャラクシー 2 (0.720)	名作 (1.051), あたり (0.877), 良い (0.790), 訳 (0.683), 飛び立つ (0.621), 興味 (0.621), <u>たち</u> (0.606), ナイショ (0.569), 念願 (0.569), ほしい (0.569), 嫁さん (0.569), <u>年齢</u> (0.569), 問わ (0.569), 本日 (0.559), 宇宙 (0.558), 冒險 (0.558), 定番 (0.558), 試し (0.558), めちゃくちゃ (0.558), おもしろく (0.544)
5	モンスターハンター 3 (0.712)	いどこ (0.454), 気に入った (0.430), 解禁 (0.416), 待ち (0.416), ラッピング (0.391), うれしい (0.384), 速かった (0.384), 品 (0.371), 買える (0.357), 世の中 (0.357), 欠々 (0.333), 価値 (0.333), <u>WII</u> (0.333), きれい (0.328), 変わり (0.312), 即決 (0.307), 機会 (0.301), お願ひ (0.301), 販 (0.299), 現在 (0.294)

た。これは、楽天市場の多くのユーザがゲームを購入する際に、複数人で遊ぶことを前提に購入していることがわかる。

対象商品 A の「New スーパーマリオブラザーズ wii」や対象商品 C の「太鼓の達人みんなでパーティ☆3代目！」では同シリーズの作品が上位 5 位以内に 2 つ以上入っていることから、シリーズの作品が前作の不満を解決できる流れがあることがわかる。満足情報についても目新しさといった、新たに追加

された要素についての評価が高かった。同時に、定番といったシリーズの流れを評価する特徴語も多くされ、新しさと定番のうまくバランスが取れている作品ほど評価が高いと考えられる。逆にいえば、このバランスが崩れると、不自然やマンネリ化といった不満にもつながる可能性があるともいえる。

対象商品 B の「Wii Sports Resort」については名前の通り、ゲーム内で多くのスポーツを体験できるといったゲームであり、同じスポーツカテゴリ内では全体的に高めのコサイン類似度の値となった。ゲームプレーの場面は、親戚、友人内で対戦をす

るといった対戦ツールとして利用されている傾向にあった。さらにここで注目すべきは、ゲーム内でスポーツを楽しむだけでなく、実際に体を動かす事が目的である、Wii Fit Plus が 4 位に入っていることである。元々、wii が体感型という点を前面に押し出したゲームハードではあるが、実際に体を動かすゲームが上位に入るということは、スポーツゲームを購入するユーザは、ゲーム内でゲームをするだけでなく、実際に体を動かすことを好む人が多いと推測できる。

対象商品 D の「桃太郎電鉄 2010 戦国・維新ヒーロー大集合! の巻」はボードゲーム形式のパーティーゲームであるため、価格に関するレビューが散見された。パーティーゲームは長時間プレーする、いわゆる「やり込む」といったことをあまりするゲームではないため、比較的安価かつ、数人で気軽に遊べるといった点が満足情報として多く抽出されたと考えられる。

対象商品 E の「星のカービィ wii」は対象商品 A の「New スーパーマリオブラザーズ wii」と同じアクションゲームカテゴリ内の作品ではあるが、より初心者でも遊びやすい難易度に抑えてあるのが特徴であり、低年齢の子を持つ親からも高い評価を得ている。しかし、本編外のモードでは上級者ですら、手ごたえを感じるほどの高難易度のゲームモードも用意されているため、難易度に関しては幅広い層から支持を得ており、不満も少ない。本作品は、1 人でも遊べるが、最大 4 人で協力できる点を特徴としており、不満解決商品で上位に入った作品は、すべて 1 人でも遊べ、複数人で協力して遊べるものとなっているため、どちらかという関連度が高い商品が抽出されることとなった。ただし、これはあくまでも楽天市場内のレビューの結果によるものであって、他のコアなゲーマー向けの E コマースサイトではあまり重要視されないという可能性も十分にある。

5. おわりに

本研究では、楽天市場の商品レビューから抽出されるネガティブな情報とポジティブな情報を用いて、商品に対する固有の問題点を抽出し、それらを解決する代替品となる商品を発見し推薦する手法を提案した。また、楽天市場の 5 商品約 2000 件のレビュー情報の実データを用いた不満情報および不満を解決できる商品抽出を検証した。

今後の課題として、不満情報抽出の改善を行う予定である。E コマース上の不満情報には、必ず発送といった手順が入る性質上、ショッピングサイトや配送業者に対する不満情報も散見された。しかしこれらの情報は、本研究の商品に対する不満情報とは無関係であり、不満情報からは除外すべきである。こういったノイズを正確に除去することができれば、より不満解決商品の推薦の精度を向上させることが期待できる。また、評価実験において、今回はレビュー件数が多いものを対象商品としたが、レビュー件数が少ないとコサイン類似度の値が極端に高い値や逆に低い値が出やすくなる点や、たとえば、「面白い」や「おもしろい」といった、漢字と平仮名などといったまったく同じ意味ではあるが、表記の違いで別の単語として抽出されている点を改善すれば、コサイン類似度計算もより精度が高められる点が今後の課題となった。

謝 辞

本研究では、楽天株式会社が国立情報学研究所の協力により研究目的で提供している「楽天公開データ」を利用した。ここに謝意を表す。

文 献

- [1] 中野裕介, 湯本高行, 新居学, 佐藤邦弘. 商品レビュー要約のための属性-意見ペア抽出. 研究報告データベースシステム (DBS), Vol.2014, No.15, pp.1-7, 2014 年 11 月.
- [2] 山本竜太郎, 藤本悠, 大原剛三. ユーザの重視する不満足意と好評意見を考慮した商品推薦システム. 第 4 回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2012), A9-2, 2012 年 3 月.
- [3] 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一. テキストマイニングによる評価表現の収集. 情報処理学会研究報告自然言語処理 (NL), Vol.2003, No.23, pp.77-84, 2003 年 3 月.
- [4] 平山拓史, 湯本高行, 新居学, 佐藤邦弘. 語の共起と極性にに基づく商品レビュー閲覧支援システム. 研究報告データベースシステム (DBS), Vol.2012, No.3, pp.1-9, 2012 年 11 月.
- [5] 相澤彰子. 語と文書の共起に基づく特徴度の数量的表現について. 情報処理学会論文誌, Vol. 41, No. 12, pp.3332-3343, December 2000.
- [6] 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一. 意見抽出のための評価表現の収集. 自然言語処理, Vol.12, No.3, pp.203-222, 2005.
- [7] 東山昌彦, 乾健太郎, 松本裕治, 述語の選択選好性に着目した名詞評価極性の獲得, 言語処理学会第 14 回年次大会論文集, pp.584-587, 2008.
- [8] Kensuke Mitsuzawa, Maito Tauchi, Mathieu Domoulin, Masanori Nakashima and Tomoya Mizumoto. FKC Corpus: a Japanese Corpus from New Opinion Survey Service. In proceedings of the Novel Incentives for Collecting Data and Annotation from People: types, implementation, tasking requirements, workflow and results, pp.11-18, Portoro, Slovenia, May 2016.
- [9] 三澤賢祐, 田内真惟人, Mathieu Domoulin, 中島正成, 水本智也. ネガティブ評判情報に特化したコーパスの構築と分析. 言語処理学会第 22 回年次大会発表論文集, pp.501-504, 2016.
- [10] 三澤賢祐, 田内真惟人, Mathieu Domoulin, 中島正成, 水本智也. 意見投稿プラットフォームにおける意見クラスタリングの試み. 言語処理学会第 22 回年次大会発表論文集, pp.1037-1040, 2016.
- [11] 長谷川徹, 北山大輔. 不満調査データセットを用いた不満グループの可視化. 第 9 回データ工学と情報マネジメントに関するフォーラム (DEIM Forum 2017), P7-1, 2017 年 3 月.
- [12] 末廣駿, 斎藤博昭. 不満調査データセットの素性ベクトル化. 言語処理学会第 23 回発表論文集, pp.545-548, 2017 年 3 月.
- [13] Toshinori Hayashi, Yuanyuan Wang, Yukiko Kawai, Kazutoshi Sumiya. An E-Commerce Recommender System using Complaint Data and Review Data. In proceedings of ACM IUI2018 Workshop on Web Intelligence and Interaction (WII 2018), Vol-2068, wii8, pp. 1-5, March 2018.