

E-Commerceにおける不満情報とレビューに基づいた 不満解決商品推薦手法の提案

林利憲^{1,a)} 王元元² 河合由起子³ 角谷和俊¹

概要: E-Commerceにおける商品推薦技術が広く普及しており、他ユーザの購入履歴情報だけでなく、購入後に記載される評価値やコメントといったレビューも考慮する、評価の高いポジティブレビューに基づいた関連商品の推薦手法が代表的な推薦手法となっている。また、購入履歴からその後に必要な付属品といった関連商品も推薦されるが、商品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（例えば画質やサイズ）といった評価尺度を選択する必要があり、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、Amazonのレビューおよび不満買取センターのネガティブな不満情報の2種類の評価データを用いて、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。提案手法では、まず、ユーザが選択した製品の不満情報のネガティブ評価から特徴（不満）ベクトルを生成し、次にレビューの高評価と低評価の差異から商品の特徴（ポジティブ）ベクトルを生成し、不満ベクトルとのコサイン類似度を算出し、不満を解決するレビューを発見する。これにより、閲覧している製品に対して、不満情報とその不満を解決する代替品およびレビュー情報をユーザへ提供できる。本稿では、不満情報とレビューに基づいた商品の不満および解決可能な代替品推薦手法について述べ、定性的評価による検証を行う。

Proposal of Recommendation System based on Dissatisfaction Data and Review Data for E-Commerce

TOSHINORI HAYASHI^{1,a)} YUANYUAN WANG² YUKIKO KAWAI³ KAZUTOSHI SUMIYA¹

1. はじめに

近年、E-Commerceにおける人々の商品購買行動で、他ユーザのレビューを参考に商品選択することが主流となった。Amazonや楽天など多くのE-Commerceサイトでは、ユーザの商品選択の要素としてレビュー情報は重要であり、E-Commerceのレビュー情報のランキングに関する研究が広く行われている。

本研究では、レビュー情報だけでなく、不満買取センターのネガティブな不満情報を用いて、商品に対する固有

の問題点を抽出し、さらにレビュー情報と不満情報から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、不満買取センター^{*1}より、商品ごとに分類された不満情報を抽出する。不満買取センターでは、ユーザが世の中に存在する不満を自由に投稿できるシステムとなっている。投稿する際、不満とそのカテゴリの入力は必須となっているが、不満に該当する商品、もしくはサービスは任意で入力できる。本研究で用いる不満は、その該当商品が入力された不満に限り、商品ごとに分類された不満を入手する。

次に、E-Commerceから商品ごとのレビュー情報^{*2}を抽出し、各商品のポジティブ評価となる特徴語を抽出する。各商品のポジティブな特徴語抽出は、各レビューに星の数

¹ 関西学院大学
Sanda-shi, Hyogo 669-1337, Japan

² 山口大学
Ube-shi, Yamaguchi 755-8611, Japan

³ 京都産業大学
Kyoto-shi, Kyoto 603-8555, Japan

a) den82687@kwansei.ac.jp

^{*1} <http://www.nii.ac.jp/dsc/idr/fuman/fuman-user.html>

^{*2} <https://www.amazon.co.jp/>

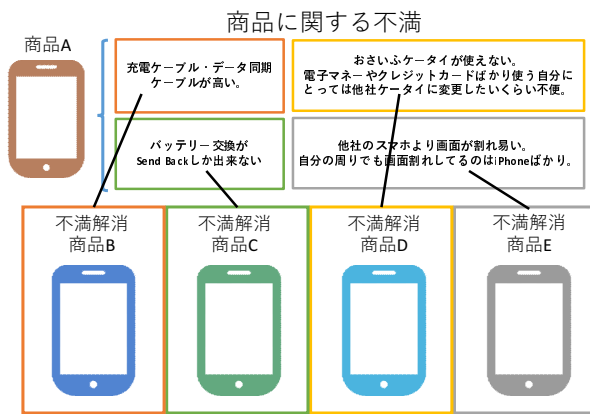


図 1 システムイメージ

などで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、高評価値であっても実際のレビュー文章では、「××は良くなかったが、○○は良かった」というように、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、高評価レビューの特徴語の値から低評価レビューに含まれる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算出法に基づき算出する。

最後に、商品の不満を示す特徴語とレビューのポジティブな評価値のみをもつ特徴語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、不満を解消する代替商品の推薦が可能となる。また、本システムでは、不満から商品選択が可能のため、「画質」や「値段」といった規定の条件以外のユーザの潜在的不満となりえる新たな不満の発見も期待できる。

本稿では、まず2章でこれまでに行われた先行研究を述べ、既存の研究上と本研究の関連性を示す。3章では本研究で使用する不満調査データセットに関するデータの特徴を分析する。その後、4章で提案システムの概要と詳細を述べ、5章で実データを用いた算出を行った結果・考察を述べる。そして最後に今後の課題を含めたまとめとする。

2. 関連研究

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。三澤らは、不満情報を用いた研究としてコーパスを用いてユーザ分析を研究した [1], [2]。また、不満情報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 [3]。北山らは、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視覚化を行った [4]。末廣らは、bag-of-wordsを用いて不満情報の素性化を行い、意味情報を分析した [5]。しかし、本研究は不満情報を用いた情報推薦システムを提案するので、その

点が異なる。

さらに、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 [6]。平山らは、レビューの極性分析を行い、レビュー情報の可視化するシステムを提案した [7]。相澤は [8]、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法の改良手法を検証した。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、多くあるテキスト分析手法で、TF-IDFを用いた確率と情報量を考慮したテキスト分析を使ったシステムを提案する。

そして、E-Commerceにおける購買行動の促進を目的とした研究もいくつか挙げられる。中野らは、レビューの意見ペアを抽出し、商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 [9]。山本らは、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 [10]。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、別データである不満のみの情報を用いて、より活用的な不満解消商品推薦を提案する。また、本研究では、高評価レビューには純粋な満足レビューだけでなく否定的内容もあるとし、算出過程で低評価レビューを用いることにより、ポジティブな特徴語のみで構成する特徴ベクトル生成を行う。

これまで不満情報の分析・可視化、テキストマイニングを用いた商品レビュー分析、そして商品レビューを用いたユーザの購買行動促進システムの提案などが研究された。本研究では、それらの研究に基づき、不満情報を用いたネガティブな意見から相反するポジティブな好評意見をもつ商品を推薦する。これにより、これまでになく不満から商品を導き出す商品推薦システムを実現する。

3. 不満調査データセット

3.1 データセット概要

提案システムで利用する不満調査データセットは、株式会社 Insight tech (旧株式会社不満買取センター) が所持する Web サービス「不満買取センター」で投稿された情報を国立情報学研究所 (NII) によって提供されたものである。2015年3月18日から2015年9月23日に投稿された254,683件の不満投稿情報を用いる。

この Web サービスは、利用ユーザがアカウントを作成し、投稿するシステムであり、ユーザが不満を投稿することで、ポイントを集めることができる。そのポイントを交換することで、現実世界に還元される仕組みである。ありとあらゆる不満を投稿することができ、本研究で用いる商品に関する不満や、電車やホテルといったサービスの不満に加え、

表 1 不満カテゴリ

メインカテゴリ	サブカテゴリ数	メインカテゴリ	サブカテゴリ数
暮らし・住まい	14	自動車	14
ファッション	10	趣味・エンタメ	13
食品・飲料	9	外食・店舗	9
医療・福祉	5	アウトドア・スポーツ	6
デジタル・家電	6	業界・業種	14
美容・健康	5	宿泊・観光・レジャー	3
公共・環境	6	教育	10
国際・文化	5	政治・行政	4
人間関係	8	仕事	5
ペット	6	その他	1

政治への不満も投稿できる。

3.2 不満情報分析

不満情報の投稿項目には必須項目と任意項目があり、次に一部を示す。

- (1) user_number : 投稿ユーザ ID
- (2) time : 投稿時刻
- (3) id : 投稿 ID
- (4) fuman : 不満
- (5) category : メインカテゴリ (任意)
- (6) sub_category : サブカテゴリ (任意)
- (7) product_name : 不満を持っている対象 (任意)
- (8) proposal : 不満に対する改善提案 (任意)

本研究で利用するのは、「fuman」、「product_name」、そして「category」「sub_category」とする。「product_name」は任意項目であるため、入力されている不満データに限り用いる。カテゴリ情報は表 1 に示すように、20 のメインカテゴリ、153 のサブカテゴリから構成され、ユーザが判断し選択する。

4. 提案システム

4.1 システム概要

本システムは、レビュー情報を活用し、ユーザに適した商品を示すことを目的とする。システム概要を図 2 に示す。ユーザが閲覧している商品 X に対して、その商品の不満を解消する商品 Y を発見する。まず、不満情報データベースより各商品に分類された不満情報を抽出する。次に、E-Commerce から各商品のレビュー情報を抽出し、商品ごとのポジティブ評価された特徴語を抽出する。そのために、最初にレビューごとに星による数値で示される評価値で判断し、高評価レビューを抽出する。しかし、高評価レビューでも、「××は良くはなかったが、○○は良かった」というように、高評価に相応しくない低評価内容も含まれていることが多い。そこで本システムは、次に、高評価レビューの特徴語値から低評価レビューの特徴語値を引き、ポジティブな特徴語のみをもつ特徴ベクトルを算出

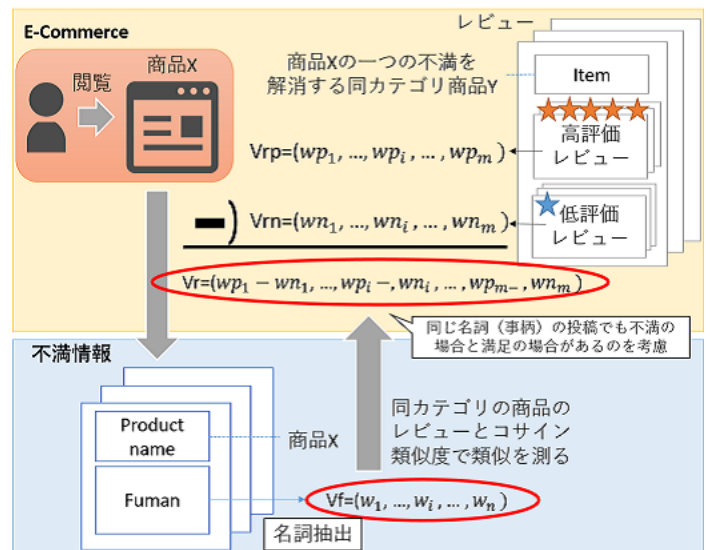


図 2 システム概要

する。

本システムでは、TF-IDF 算出法に基づき各レビューの特徴語値を算出する。その後、任意の商品の不満を示す特徴語と同カテゴリ商品のポジティブ評価を持つ特徴語の類似度を算出する。そして、その商品の不満となっている特徴語が評価されている他商品の推薦を行う。任意の商品に対する不満対象がポジティブ評価されているレビューの対象商品を推薦することで、その不満を解消する商品推薦が可能となる。また、不満から商品選択が可能となるため、ユーザの潜在的不満も解消することができる。

4.2 不満情報抽出

提案システムでは、第章で述べた不満調査データセットを利用し商品ごとの不満を抽出する。全体の不満の内、およそ 50% の不満情報には、不満の対象となる商品名・サービス名が付与されている。本研究では、E-Commerce サイト上の商品推薦システムとしているため、商品名のみを対象とする。

次に、商品ごとに分けられた不満情報を統合し、一つの文書とし、形態素解析より名詞のみを抽出する。その後、以下の TF-IDF 算出法の式に基づき商品に対する不満を表す特徴語 w 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 $v_F = (w_1, \dots, w_i, \dots, w_n)$ を生成する。

$$TF = \frac{\text{任意の商品不満情報の単語 } i \text{ の出現数}}{\text{任意の商品不満情報の単語総数}}$$

$$IDF = \log \frac{\text{同カテゴリの総商品数}}{\text{単語 } i \text{ が不満情報に出現した商品数}}$$

TF は、任意の商品不満情報の単語頻度を表す。IDF は、任意の商品の不満情報を 1 つの文書とした時の文書頻度の逆数を表す。ここでの同カテゴリとは、不満のサブカテゴリに属する商品数とする。

4.3 レビュー情報に基づくポジティブ評価の商品抽出

次に、E-Commerce 上のレビュー情報を抽出する。ここで、レビュー情報の高評価値だけでなく低評価値を用いて商品に対するポジティブな評価となる特徴語を抽出する。これにより、任意の商品に対する不満を解消する商品（代替品）を結びつける際に、ポジティブな特徴語を用いることで不満解消商品を推薦できる。

抽出するレビュー情報は、商品ごととする。不満の対象商品に対し、その不満を解消する商品を求めるために、レビュー情報も商品ごとにする。そこで、まず高評価レビューを抽出する。レビュー評価の判断材料として、本研究ではレビューに付与された評価値（星の数）を用いる。評価値は一般的に、5段階のリッカート尺度が用いられており、評価値が5や4であれば、その商品に対するレビュアーの評価（満足度）が高いことを示しており、評価値が1の場合は、低い評価（問題点）を示す。

しかし、高評価レビューの投稿情報の中には、低評価の内容が含まれていることがある。例えば、「××は良くなかったが、〇〇はよかった」というものである。このようなノイズを除くため、低評価レビューを用いる。低評価レビューには、「××」に関する情報が多く存在する。そこで、高評価レビューの特徴語値から低評価レビューの特徴語値を引くことで、「××」といった特徴語を除く。

特徴語算出には4.1節と同様にTF-IDF算出法を用いる。まず、任意の商品の高評価レビューの文章を全て抽出し、統合することで1つのドキュメントを生成する。同様に、低評価のレビューの文章を全て抽出し、統合し1つのドキュメントを生成する。次に、形態素解析し、名詞のみを抽出し、TF-IDFに基づく以下の式より、単語の要素値を算出する。

$$TF = \frac{\text{任意のレビュー情報の単語 } i \text{ の出現数}}{\text{任意のレビュー情報の単語総数}}$$

$$IDF = \log \frac{\text{同カテゴリ商品の総文書数}}{\text{単語 } i \text{ がレビュー情報に出現した文書数}}$$

レビュー情報のTF-IDF算出法は、不満情報の算出法とほぼ同じである。ただし、各商品に対し高評価レビューと低評価レビューの2種類の文書を抽出しているため、「同カテゴリの総商品数」には、同カテゴリに属する商品数を2倍した数値が入る。そして、求めた要素値を用いて、高評価ベクトル v_{rp} と低評価ベクトル v_{rn} を生成する。

$$v_{rp} = (wp_1, wp_2, \dots, wp_i, \dots, wp_m) \quad (1)$$

$$v_{rn} = (wn_1, wn_2, \dots, wn_i, \dots, wn_m) \quad (2)$$

最後に、商品ごとによって導き出した高評価レビューの式(1)と低評価レビューの式(2)の差を算出し、ポジティブな評価ベクトルを生成する。

$$v_R = (v_{rp} - v_{rn}) =$$

$$(wp_1 - wn_1, \dots, wp_i - wn_i, \dots, wp_m - wn_m) \quad (3)$$

ここで、負の値をもつ名詞はネガティブ評価と考えられる。そこでポジティブな評価となる名詞のみを抽出するために、負の場合は0と換算し、各商品の特徴ベクトルは、ポジティブ評価となる名詞のみとなり、1に近いほど高評価とされる特徴語、0に近いほど高評価とされない特徴語となる。

4.4 不満と商品の類似度算出

不満に対して、解消する商品推薦をするために、不満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、前節で求めた二種類の特徴ベクトルの類似度を算出方法を述べる。

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。それらの特徴語は不満情報では不満の対象とされているが、レビュー情報ではポジティブ評価の対象となっているので、不満情報と類似度の高い特徴語をもつポジティブな評価レビューをもつ商品が、その不満を解決する商品となる。

類似度算出には、4.1節、4.2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。

$$Sim(v_F, v_R) = \frac{\sum w_n \cdot (wp_i - wn_i)}{\sqrt{\sum_{i=1}^{|V|} (w_n)^2} \cdot \sqrt{\sum (wp_i - wn_i)^2}}$$

ここで v_f は、不満情報から算出した商品ごとの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特徴ベクトルを示している。本システムでは、同カテゴリ内において、不満を解消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、不満情報、レビュー情報という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もつ商品レビュー情報を求める。

5. 評価実験

第4章では、提案システムの流れを述べた。E-Commerceにおいてユーザの不満を解消する商品推薦を行うために、不満調査データセットを用いる。本研究では、E-Commerceにおけるレビュー情報としてAmazonのレビュー情報を取得し、検証した。まず、5.1節で、iPhone5の不満を対象に、E-Commerce上で商品レビューを多く持つ同カテゴリ5商品、また別カテゴリ2商品を対象に類似度算出をした。また、5.2節では、iPhone5の不満だけでなく、同カテゴリ

表 2 「iphone5」の不満情報より抽出した特徴語上位 40 語

特徴語 (TF-IDF 値)
交換 (0.23), G (0.17), 機 (0.15), 量 (0.15), 写真 (0.15), 最近 (0.15), 入力 (0.13), 文字 (0.13), 容量 (0.13), 電池 (0.12), - (0.12), 電波 (0.17), 修理 (0.15), 変 (0.15), Softbank (0.10), 携帯 (0.15), 設定 (0.13), ショップ (0.13), データ (0.13), カード (0.08), アプリ (0.08), アップデート (0.07), 子供 (0.06), 不便 (0.06), 笑 (0.06), いっぱい (0.06), アップル (0.06), 無料 (0.06), インターネット (0.06), 半 (0.06), 機能 (0.06), パソコン (0.06), iPhone (0.06), カー (0.06), 充電 (0.05), 悪 (0.05), 削除 (0.05), GB (0.04), コード (0.04), 無駄 (0.04),

表 3 不満情報の「iPhone5」と7商品(レビュー)との類似度

Items	iPhone5
iphone5	0.28
iphone6	0.63
AQUOS	0.46
Xperia	0.46
Huawei	0.60
PC_ASUS	0.53
CAMERA	0.23

3つの不満を分析し、同じように類似度算出をした。

5.1 不満解消商品の算出

まず、iPhone5を対象とした不満を分析する。iPhoneに関する不満情報は、68件投稿されており、抽出した名詞は232件であった。表2はiPhone5を対象とした不満から抽出した上位特徴語40件とそのTF-IDF値の結果である。

また、Amazonから抽出した商品レビューは以下の7つとした。商品選択の基準は、不満情報の投稿数と同等のレビュー数が約50件程度の商品とした。また、同カテゴリの携帯端末製品5点と異なるサブカテゴリのパソコンとカメラの2点を選択した。

iphone5: iPhone5

iphone6: iPhone6

AQUOS: AQUOS PAD SH-08E

Xperia: Xperia Z5

Huawei: Huawei 5.9型 Mate9

PC_ASUS: ASUS 2in1 パソコン T101HA/10.1型

CAMERA: MUSON アクションカメラ 4K

iPhone5と他7商品との類似度の算出結果を表3に示す。また、表4は、その他7商品から抽出した特徴語上位10件である。不満情報の特徴ベクトルの要素数は307単語あり、商品レビューの特徴ベクトルは2,329単語となった。

類似度の結果は、最小値が「iphone5」の0.23となり、最大値が「iphone6」の0.63となった。iphone6との上位で共通する特徴語は、「容量」「データ」「電池」であった。次に類似度の高い商品は「Huawei」の0.60となり、iPhone5

表 4 レビュー特徴語上位 10 件 (TF-IDF 値)

商品	特徴語上位 10 件
iPhone5	mvno (1), 包装 (0.49), めちゃくちゃ (0.49), 運用 (0.47), v (0.47), アマゾン (0.39), 満足 (0.35), クリーニング (0.33), 省略 (0.33), 一瞬 (0.33)
iPhone6	アクティベーションロック (0.33), OCN (0.32), SIM (0.29), 格安 (0.26), ロック (0.26), 無事 (0.26), IJ (0.24), モバックス (0.22), プロフィール (0.22), IJmio (0.22)
AQUOS	フルセグ (0.40), フル (0.35), テレビ (0.27), アンテナ (0.18), デジ (0.17), インチ (0.17), ネクサス (0.17), docomo (0.16), 端 (0.15), 7 (0.14)
Xperia Z5	Z (0.42), Compact (0.33), ac (0.26), 認証 (0.24), Xperia (0.23), docomo (0.23), ケータイ (0.22), p (0.21), 指紋 (0.20), サイフケータイ (0.17)
HUAWEI	mate (0.47), lite (0.43), 指紋 (0.41), 認証 (0.38), ガラス (0.22), 撮影 (0.20), 指 (0.19), 写真 (0.18), 高級 (0.17), Mate (0.15)
ASUS	Windows (0.35), キーボード (0.30), ボード (0.25), PC (0.23), キー (0.23), in (0.17), B (0.17), Wi (0.17), コス (0.17), パソコン (0.17)
MUSON	アクション (0.93), アタッチメント (0.75), 豊富 (0.43), 画質 (0.41), カメラ (0.25), 付属 (0.24), 旅行 (0.23), 活躍 (0.21), シュノーケル (0.21), 記録 (0.21)

の不満情報との上位共通特徴語として、「写真」「交換」「認証」「指紋」であった。これらのiPhone6とHuaweiの共通特徴語は、iPhone5では不満として挙げられているが、それぞれの商品レビューでは高評価のポイントとして挙げられていた。つまり、「容量」「データ」「電池」というiPhone5の不満をiPhone6が解消し、「写真」「交換」「認証」「指紋」の不満を解消するのはHuaweiとなる。以上の結果から、iPhone5の複数の不満を発見し、それを解決可能な商品であるiphone6やHuaweiといった商品が発見できた。

今後、特徴語不満に応じた商品の推薦による定性的評価を行う。具体的には、例えば、今回の実験結果ではiPhone5には、2つの高い類似度を持つ商品、iPhone6とHuaweiを推薦し、異なるそれぞれの不満となる特徴語を提示し、解消につながるかを検証する。また、特徴語から適切な不満情報とレビュー情報の発見と提示に関する検討も必要となる。

5.2 同カテゴリ商品の類似度算出

5.2節では、5.1節で取り上げたiPhone5以外に、同カテゴリの他3商品の不満データを加えて算出を行った。不満

表 5 不満特徴語上位 10 件

商品	特徴語上位 10 件
iPhone5	交換, G, 機, 量, 写真, 最近, 入力, 文字, 容量, 電池
AQUOS	傷, カバー, 一, シート, 処理, ブック, 内蔵, 型, グリップセンサー, プレイヤー
Xperia	ステレオ, 音楽, XPERIA, メモリ, 電池 カー, 音, フィルム, USB, 皮脂
Huawei	発信, つもり, 番号, 電話, 指先, 目的, 相手, 限り, 心配, 履歴

表 6 不満情報の 4 商品とレビュー 6 商品との類似度

Items	iPhone5	AQUOS	Xperia	Huawei
iphone5	0.28	0.20	0.18	0.10
iphone6	0.63	0.25	0.33	0.12
AQUOS	0.46	0.17	0.40	0.20
Xperia	0.46	0.16	0.60	0.05
Huawei	0.60	0.27	0.43	0.06
PC_ASUS	0.53	0.17	0.21	0.04
CAMERA	0.23	0.06	0.03	0.02

調査データセットから抽出したスマホカテゴリの商品は、「iPhone5」「AQUOS phone」「Xperia Z3」「Huawei」の 4 つとした。

これら 4 種類のデータを用いて、TF-IDF 算出法に基づき、各商品の不満を表す特徴語を抽出する特徴ベクトルを生成した。表 5 は、それぞれの商品の不満から抽出した結果、上位 10 件を示す。また、表 6 は、不満情報を持つ商品 4 件と、商品レビューを持つ商品 7 件の類似度を算出した結果を示す。

結果より、Xperia Z3 と Xperia Z5 で高い類似度算出値が見られ、「音楽」「指紋」「ゲーム」というような特徴語が共通して上位に見られた。iPhone5 だけでなく他商品でも、提案システムによる不満解消商品の算出が可能であることを示すことができた。

また、iPhone5 と PC_ASUS の間にも高い類似度が得られた。共通で持つ特徴語には、「容量」「アプリ」「ストレージ」などが見られた。別カテゴリの商品でも高い類似度が算出されたので、別カテゴリの商品推薦も不満解消につながるという結果であった。

今回の実験により、携帯端末のスマホカテゴリに対する 5 点の不満情報と他のカテゴリ 2 点を加えた 7 商品のレビューから抽出したポジティブな評価情報を用いて、商品ごとの不満解消商品を算出した。結果により、双方のベクトルがもつ共通特徴語に関する不満を解消する商品を算出できただけでなく、別カテゴリでも不満解消商品を算出できることを実証できた。今後は、商品の不満の特徴ごとに異なる商品の推薦や、別カテゴリの不満解消商品の推薦手法について検討予定である。

6. おわりに

本稿では、不満を解消する商品を推薦するシステム提案を目的として、TF-IDF 法、コサイン類似度法を用いた不満解消商品の算出方法を述べた。また、不満買取センターの 4 商品約 200 件の不満情報、Amazon の 7 商品約 350 件のレビュー情報の実データを用いた実験を行い、提案手法の有用性を実証した。

今後は、特徴語を用いることにより、同一商品でも不満に応じた異なる商品の推薦および提示手法の検証を予定している。また、同一カテゴリだけでなく、別カテゴリを含めた解決につながる商品の発見および推薦手法も検討予定である。

謝辞 本研究では、株式会社 Insight Tech 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「不満調査データセット」を利用した。ここに謝意を表す。

参考文献

- [1] 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.
- [2] 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.
- [3] 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.
- [4] 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 2017), P7–1, 2017.
- [5] 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, MAR 2017.
- [6] 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, MAR 2003.
- [7] 平山拓央, 湯本高行, 新居学, 佐藤邦弘. 語の共起と極性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, NOV 2012.
- [8] 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, DEC 2000.
- [9] 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, NOV 2014.
- [10] 山本竜太郎, 藤本悠, 大原剛三. ユーザの重視する不満意見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 2012), A9–2 2012.

正誤表

下記の箇所に誤りがございました．お詫びして訂正いたします．

訂正箇所	誤	正
1 ページ 右下 脚注 1	http://www.nii.ac.jp/dsc/idr/fuman/fuman-user.html	http://fumankaitori.com/
2 ページ 右下 35 行目	Insight tech	Insight Tech
2 ページ 右列 37 行目	国立情報学研究センター (NII)	国立情報学研究所 (NII)