

商品選別支援のためのユーザの興味に基づいた 商品レビューランキング

諭 思綺[†] 牛尼 剛聡[‡]

[†]九州大学芸術工学府 〒815-8540 福岡市南区塩原 4-9-1

[‡]九州大学芸術工学研究院 〒815-8540 福岡市南区塩原 4-9-1

E-mail: [†]yu.siqian.291@s.kyushu-u.ac.jp, [‡]ushiama@design.kyushu-u.ac.jp

あらまし ユーザが E コマースサイトで商品を選ぶ際には、レビューを参照することが一般的である。しかし、商品情報と一緒に掲載されているレビューには、全てにユーザが興味のある情報を含んでいるわけではないという問題がある。レビューの中から興味がある情報を取得するために、全てのレビューを読むことは時間がかかり、ユーザにとっては大きな負担となる。本研究では、ユーザの興味に基づいて、関連文を推薦するシステムを提案する。本手法では、関連文を抽出するため、Word2vec や TF-IDF などを利用する 2 種類の手法を提案する。1 つの手法は単語の埋め込みに基づいた手法であり、もう 1 つの手法は文の埋め込みを利用する手法である。評価実験の結果、2 つの手法の結果を評価し、文埋め込み手法の方が性能が良いことがわかった。

キーワード 自然言語処理, Word2vec, 情報検索

1. はじめに

現在、インターネットの急速な発展と普及に伴い、Amazon などの E コマースサイトを利用するユーザが増加している。カスタマーレビューが欠かせない情報源となっている。また、カスタマーレビューが、ユーザのオンラインショッピングの意思決定プロセスに影響を与える可能性がある。しかし、商品ページに大量のレビューが存在している場合、レビューの中からユーザが興味を持つ情報を得るために、全てのレビューを読むことはユーザにとって大きな負担となる。例えば、ユーザがマスクを購入しようとしている場合、ユーザはそのマスクを購入するかどうかを決めるために、「保湿」という特徴について、他のレビューにおいてどの表に評価されているかを知りたいという状況を考える。このとき、このマスクに関して投稿されたほかのたくさんのレビューを読まなければならない。しかし、レビュー中「保湿」に関する情報を収集するために、全てのレビューを読むことは、時間がかかりユーザにとって負担となる。

このように、大量のレビューからユーザが興味を持つ情報を効果的に取得するために、E コマースサイトではいくつかの工夫が行われている。例えば、代表的な E コマースサイトの一つである Amazon では、図 1 に示すように、ユーザのスコアの分布が示され、ユーザがスコアを選択することにより、そのスコアをもつレビューが提示される。しかし、レビュー全体のスコアが、レビューに記載されている全ての意見と一致ではないという問題があり、ユーザにとって有益な情報を短時間で効率的に取得することが困難となる場合がある。例えば、大島ら[1]は、ユーザがレビュー検索サ

イトのスコアやランキングから詳細な情報を得ることが困難であることを示している。

この問題を解決するために、ユーザがレビューを読んでいるときに使われている特徴情報を検索するシステムを提案する。短い時間で気になる情報を取得することを目標とする。具体的には、ユーザがレビューに興味を持つ特徴的な評価文に着目し、これを基にユーザにとって有益な情報を収集し、推薦することを目的とする。

カスタマーレビュー

★★★★☆ 星5つ中の4.1 ↓

評価の数 627

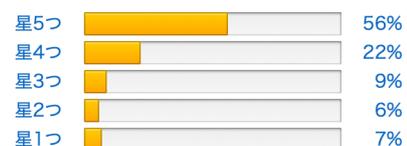


図 1 Amazon における代表点の表示

ユーザによって生成されたビッグデータから重要な情報を抽出するために、Erick ら[2]は、Amazon 上のオンラインレビューを自動的に分析するセンチメント分析ツールと偽レビュー検出ツールを用いて、マーケティング担当者や消費者の意思決定を支援する偽レビュー検出フレームワークを提案している。本研究では、レビューが偽物であるかどうかを分析することではなく、全てのレビューは価値があるレビューを前提とする。

ユーザが興味を持つ商品のレビュー情報を人手で収集するのは難しいという問題を解決するために、本研究では、商品選別支援システムを提案し、オンラインショッピングの効率的な意思決定を支援することを目的とする。提案システムでは、ユーザはレビューを読む際に、興味を持った記述に出会ったときにマークすることができる。システムは、ユーザの興味記述に基づいて、関連度によってレビュー文のランキングをユーザに推薦する。関連文の抽出するために、自然言語処理技術 Word2vec モデルと TF-IDF などを利用する。

2. 関連研究

大量のテキストデータを解析して未知の情報を発見するために自然言語処理 (Natural language processing) が利用される [3]。本研究では、自然言語処理を利用してレビューを分析してユーザの商品選別を支援するシステムを開発することを目的とする。

TF-IDF (term frequency-inverse document frequency) は、ある単語が文書の集合体の中で文書にとってどれだけ重要であるかを評価する統計的な尺度である。TF-IDF 値が高い単語は、その単語が登場する文書と強い関係があることを意味する。この手法は情報検索 [9] やテキストマイニングに広く用いられている。

TF-IDF の計算は次の数式で表す:

$$tfidf_{i,j} = tf_{i,j} \times \log \left(\frac{N}{df_i} \right) \quad (1)$$

ここで、 $tf_{i,j}$ は文書 j に含まれる単語 i の頻度を示し、 N はコーパスに含まれる文書の総数を示し、 df_i は i を含む文書の数を示す。本論文では、TF-IDF 重み付き文の埋め込みを作成する。これは、1 つの文の中のすべての重み付き単語埋め込みの平均を文埋め込みとみなすことを考えたものである。

Mikolov ら [5] は、大量のデータから高品質の単語ベクトルを生成することできる Word2vec 手法を提案している。彼らは、単語の分散表現を学習するために、CBOW モデルと Skip-gram モデルという 2 つの効果的なモデルの構造を提案した。CBOW は文脈に基づいて現在の単語を予測し、Skip-gram は現在の単語の周囲の単語を予測する。Word2vec は、単語ベクトル表現を効果的に表現する。Ombabi ら [6] は、ソーシャルテキストデータからユーザの興味情報を要約するために、事前に学習した Word2vec を用いたアプローチを提案し、ユーザの興味情報分類に対して高い精度を達成してい

る。Word2vec モデルはセンチメント分析分野にも応用できる。しかし、Word2vec 特徴量の次元が高いため、テキスト分類の複雑さが増加する。Alshari ら [12] は、Word2vec を用いて感情分析辞書中の単語に基づいて単語ベクトルのクラスタリングを行っている。単語ベクトルのかわりに類似度を利用し、センチメント分析のための特徴量の次元を小さくする手法を提案している。

ユーザが興味がある情報を表すキーワードをクエリとして入力して検索結果を取得する情報検索システムでは、TF-IDF や BM25 [14] のような語彙ベースのモデルを利用するのが一般的である。しかし、語彙モデルだけに依存した検索は、クエリを持たない関連文書を検索することが困難になる。それに対して、Word2Vec や BERT 等の単語のセマンティックを利用したセマンティック検索では潜在的な文書が抽出できると考えられる。Kuzi ら [13] は、BERT を用いたセマンティックモデルと語彙モデルを結合したアプローチを提案している。

近年、文をベクトルとして埋め込む方法としては、文の集合を異なるトピックに分解する LDA トピックモデル [7] や Doc2Vec モデルなどが提案されている。Doc2Vec モデルは、文の論理構造利用せず、Word2vec での単語ベクトル化手法をベースに、入力に別のベクトル (段落 ID) を追加する事により、文書の埋め込みベクトルを求める。文書の埋め込みベクトルの生成手法の中で最も簡単で一般的な方法は、単語レベルのベクトルの平均を求める方法である。単語レベルの埋め込みベクトルを利用して、文の埋め込みベクトルを生成することの利点は、あらかじめ学習された高品質の単語埋め込みベクトルを活用できることである。

本研究では、類似文を探索するために 2 つの手法を提案する。一つは、学習された Word2vec モデルを用いて単語ベクトルを取得し、類似する単語を抽出し、その単語を含む文を見つける手法である。もう一つの手法は、類似文を見つけるために、同じ学習済みの Word2vec モデルで得られた単語のベクトルを平均化し、文書のベクトルとする手法である。

インターネット上に生み出される大量のレビューを対象にレビューに記載された詳細な特徴を抽出し、トピックモデリング技術を用いてより意味のある高レベルの特徴にグループ化することで、ショップがユーザの意見を分析する手法が提案されている [4]。

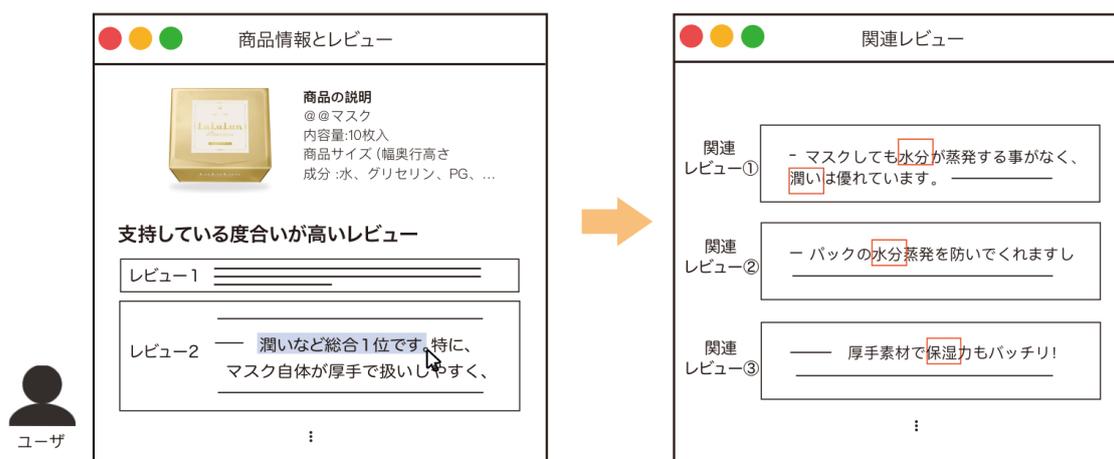


図 2 提案する商品選択支援システム

3. 提案手法

3.1 提案システム

提案システムの概要を図 2 に示す。提案システムでは、図 2 の左の図に示すように、ユーザは、このページで、製品紹介や支持率の高いレビューを閲覧することができる。また、ユーザがレビューの中で興味深い記述を見つけた場合、マウスを利用してその文を選択する。図 2 右の図は、ユーザ選択文によって、その商品レビューの中から抽出した関連文が表示されたページである。赤枠は、ユーザが選択した文に関連する特定の特徴を表示している。

3.2 ユーザ指定分に基づいたレビュー検索手法

本論文では、レビューでユーザが選択した文と類似した文を見つけるために、2つの方法を提案する。

手法 1 は、単語ベクトルを利用する。ユーザーが選択した文章からキーワードを決定し、そのキーワードの単語ベクトルを用いて、レビューの中から類似する単語探し出した。最後に、キーワードと類似する単語を含む文をユーザに推薦する。

しかし、ユーザが選択した文に複数のキーワードが含まれている場合には、手法 1 では効率的な処理ができない可能性がある。そこで、ユーザが選択した文から 1つのキーワードを抽出するのではなく、ユーザが選択した文の文書ベクトルを用いて類似文を抽出する手法 2 を提案する。

2つの手法とも同一学習済み word2vec を利用する。

3.2.1 手法 1：単語ベクトルに基づく手法

手法 1 では、ユーザが興味を持っている文中のキーワードから、類似した文を推薦するという目的を目指

す。まず、日本語の形態素解析機である Mecab¹ を用いて、ユーザが選択した文を mecab-ipadic-NEologd 辞書で前処理し、その文に含まれる全ての名詞を取得した。そして、各単語について TF-IDF 値を計算し、TF-IDF 値が一番高い名詞をキーワードとする。次のステップでは、Word2vec モデルを用いて単語のベクトルを取得し、コサイン類似度を計算して、レビュー中からキーワードと最も類似している単語を抽出する。最後に、正規表現で処理したレビューから、類似語を含む関連文を抽出する。

3.2.2 手法 2：文書ベクトルに基づく手法

ユーザが選択した文の中に、多くのキーワードが存在する場合には、1つのキーワードだけを使うよりも、文全体をベクトルに埋め込むことで、類似した文を探す方が良い可能性がある。そこで、重み付き単語ベクトルの平均化を用いて文のベクトルを計算する手法を手法 2 とする。手法 2 では、まず、1つの文から名詞を抽出し、前節で述べた Word2vec モデルを用いて、全ての名詞のベクトルを取得できる。抽出した全ての名詞の TF-IDF 値を個別に計算し、その TF-IDF 値を対応する単語埋め込みの重みとして使用する。そして、重み付けされた単語ベクトルを平均化し、文のベクトルを得た。コサイン類似度を計算して、類似文を見つけることができる。

ここで、 $\text{vec}(s)$ は文書 s の埋め込み、 $\text{vec}(w_i)$ は単語の埋め込みを表す。各単語の埋め込みは tfidf_i で重み付けされている。

$$\text{vec}(s) = \frac{1}{n} \sum_{i=0}^n \text{vec}(w_i) \times \text{tfidf}_i \quad (2)$$

sentence_1	明るい肌になりましたので効果は期待していいと思います
sentence_2	保湿重視なのであれば
sentence_3	上の方が水分量が少なく下の方が多いなんていうような含有量の違いなどなく
sentence_4	夏場のお風呂上りに毎日使用するのに良いです
sentence_5	マスクのがシワになりにくいので張りやすい

図 3 5 ユーザ選択例文

表 1 5 つ例文の適合率

	sentenc1	sentenc2	sentence3	sentence4	sentence5	NDCG(AVG)
Approch_1	0.86	0.94	0.57	0.83	0.25	0.69
Approch_2	0.91	0.91	0.70	0.89	0.77	0.83

	sentenc1	sentenc2	sentence3	sentence4	sentence5	P(AVG)
Approch_1	0.46	0.48	0.50	0.27	0.16	0.37
Approch_2	0.60	0.47	0.37	0.41	0.30	0.43

表 1 5 つ例文の NDCG

4. 評価実験

提案手法の有効性を評価するために実験を行った。TF-IDF 値を算出するために、Amazon のカスタマーレビューデータセットから 30 カテゴリのレビューと用意したマスクカテゴリ、合計 31 カテゴリを文書セット全体 (TF-IDF では文書数は 31) とした。また、白ヤギコーポレーションが公開した学習済み Word2vec モデル² を用いて、新しい単語を追加し、独自の Word2vec モデルを訓練した。公開した学習済み Word2vec モデルは、mecab-ipadic-NEologd 辞書を用い、Mecab 形態素解析機を使用した。学習済み Word2vec モデルは、日本語 Wikipedia をコーパスとして使用し、エンベディングの次元数は 50 とする。新しい追加単語には、mecab-ipadic-NEologd 辞書を用いて Mecab で前処理した Amazon.co.jp のマスク商品レビュー約 500 件を利用した。

評価を行うための実験データとして、Amazon.co.jp 上のマスク商品の日本語レビュー 15 件を人手で抽出し、正規表現を用いて約 130 文に分割した。これらの文書中から、内容や長さの異なる 5 つの文章をユーザが選択した例文としてランダムに選択した (図 3)。

各例文について類似した文書を見つけるために、セクション 3 で提案されている手法を使用した。消費者の購買行動レポート [10] では、かなりの数のユーザが購入前に 1~10 件のレビューを読んでいることが示されている。そこで、本研究では、各手法の類似文の上

位 40 文のみを抽出し、ランダムで 20 文を選択して 2 つの手法の有効性を評価した。

情報検索、システム評価、精度評価など分野に一般的に使われている適合率と NDCG を用いて、2 つの手法の抽出結果を評価した。

ポジティブラベルの予測におけるモデルの適合率を評価する式は以下の通りである。情報検索における検索結果の場合は、適合性は「適合文書/検索結果の文書」で算出される。

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (3)$$

NDCG: Normalized discounted cumulative gain [11] はランク付けタスクの評価に使用される代表的な指標である。NDCG の目的は、関連性の高い文書を推薦リストの上位に配置することである。関連性の高い項目は、中程度の関連性のある項目の前に配置されなければならない。関連性の低い項目は後に配置されるべきである。NDCG は、ウェブ検索の評価のための代表的な手法である。

アンケートを用いて行った評価は、例文と抽出文の類似性に応じて、ランダムに抽出された 20 文を人手で 0 から 5 まで評価する。アンケートには合計 4 名が参加した。最後に、適合率と NDCG の結果を算出した。

¹ <https://taku910.github.io/mecab/>

² <https://aial.shiroyagi.co.jp/2017/02/japanese-word2vec-model-builder/>

5. 実験結果

表 1 は適合率の結果を示している. 表 2 は NDCG の値を示している. Method_1 は単語ベクトルに基づく手法 1 であり, Method_2 は文書ベクトルに基づく手法 2 である.

5 つの文の適合率の結果から見ると, 2 つの方法は大きな差はなく, どちらの方法とも関連する文を抽出することができることを示している. また, 平均適合率からは, 手法 2 の方が良い結果が得られていることがわかる. この評価方法の欠点は, 2 つの方法の適合率が高いか低いかを証明することができないことである. なぜなら, 上位 40 の類似文の中からランダムに抽出された 20 文を評価しているだけだからである. NDCG の結果から見ると, 手法 1 よりも手法 2 の方が良い結果が得られると考えられる. これは, ユーザが選択した文の中に複数のキーワードが存在する場合には, 1 つのキーワードだけを抽出して関連する文を見つける方法は有効ではないことを示している. このような状況では, 文書埋め込み手法を用いた方が良い結果が得られた. 今回の実験では, 評価対象と被験者数が少ないため, 今後, さらに評価実験を重ねる必要がある.

6. おわりに

本論文では, ユーザがオンラインショッピングをする際に, 短時間で興味がある情報を見つけることができるように, ユーザーの選択に応じて情報を推薦するレビューレコメンドシステムを提案した. 似たような文書を見つけるために, 2 つの方法を提案し, アンケートを用いて 2 つの方法について検討した. 実験の結果から見ると, 文書ベクトル手法の方が良い結果が得られた.

今後の研究では, Word2vec モデルを学習するためのデータセットを増やし, 類似語の検索精度を向上させたいと考えている. 提案した手法で抽出された結果については, より客観的な評価手法についても考えていく必要があると考えている. まだ, ユーザにレビュー文を推薦するだけでなく, センチメント分析を行って, 抽出された各文の感情 (ポジティブ, ネガティブ, ニュートラル) を推定し, それぞれの感情の割合をまとめて, ユーザに提示する. ユーザが短時間で興味のある情報を得ることができるようにする.

謝辞

本研究は JSPS 科研費 19H04219 の助成を受けたものです.

参 考 文 献

- [1] 大島裕明, 山本祐輔, 山本岳洋, 加藤誠, 神門典子, “レビュー情報検索サイトにおける不便の効用に関する検討”, DEIM Forum 2018, D5-4, 2018.
- [2] Erick Kauffmann, Jesus Peral, David Gil, Antonio Ferrandez, Ricardo Sellers, Higinio Mora, “A Framework for Big Data Analytics in Commercial Social Networks: A Case Study on Sentiment Analysis and Fake Review Detection for Marketing Decision-making”, arXiv: 1903.12452v1, 2019.
- [3] Antonio Moreno, Teofilo Redondo, “Text Analytics: The Convergence of Big Data and Artificial Intelligence”, International Journal of Interactive Multimedia and Artificial Intelligence, Vol. 3, No. 6, pp. 57–64, 2016.
- [4] Emitza Guzman, Walid Maalej, “How Do Users Like This Feature? A Fine Grained Sentiment Analysis of App Reviews”, IEEE 22nd International Requirements Engineering Conference, 2014.
- [5] Tomas Mikolov, Kai Chen, Greg Corrado, Teffrey Dean, “Efficient Estimation of Word Representations in Vector Space”, arXiv preprint arXiv:1301.3781, 2013.
- [6] Abubakr H. Ombabi, Onsa Lazzez, Wael Ouarda, Adel M. Alimi, “Deep Learning Framework based on Word2Vec and CNN for Users Interests Classification”, Sudan Conference on Computer Science and Information Technology, 2017.
- [7] David M. Blei, Andrew Y. Ng, Michael I. Tordan, “Latent Dirichlet Allocation”, Machine Learning Research 3, pp. 993-1022, 2003.
- [8] Quoc Le, Tomas Mikolov, “Distributed Representations of Sentences and Documents”, In International conference on Machine Learning, pp. 1188-1196, 2014.
- [9] Juan Ramo, “Using TF-IDF to Determine Word Relevance in Document Queries”, Proceedings of the first instructional conference on machine learning, 2003.
- [10] @PowerReviews. All Rights Reserved, “The Growing Power of Reviews”, 2018.
- [11] Kalervo Ja rvelin, Jaana Keka la inen, “IR Evaluation Methods for Retrieving Highly Relevant Documents”, SIGIR 2000: Proceedings of the 23th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 41–48, 2000.
- [12] Eissa M. Alshari, Azreen Azman, Shyamala Doraisamy, Norwati Mustapha, Mustafa Alkeshr, “Improvement of Sentiment Analysis based on Clustering of Word2Vec Features”, 28th International Workshop on Database and Expert Systems Applications, 2017.
- [13] Saar Kuzi, Mingyang Zhang, Cheng Li, Michael Bendersky, Marc Najork, “Leveraging Semantic and Lexical Matching to Improve the Recall of Document Retrieval Systems: A Hybrid Approach”, arXiv:2010.01195v1 [cs.IR], 2 Oct 2020.
- [14] K. Sparck Jones, S. Walker, S.E. Robertson, “A Probabilistic Model of Information Retrieval: Development and Comparative Experiments”, Information Processing and Management, 36(6), 779-808, 2000.