

商品予測タスクを利用した商品選別支援のための 有用なレビュー文の抽出

諭 思セイ† 牛尼 剛聡††

† 九州大学大学院芸術工学府 〒 815-8540 福岡市南区塩原 4-9-1

†† 九州大学大学院芸術工学研究院 〒 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 E コマースサイトで商品を選択する際、ユーザはカスタマーレビューを参考にすることが一般的である。しかし、レビューの中に商品の特徴とは関係ない情報などが含まれることがあり、そのような不要な情報はユーザの購買意思決定に役立たない。本論文では、商品レビューの中からユーザの商品選別に有用なレビュー文を抽出する手法を提案する。本手法では、E コマースサイトのデータセットを利用して、レビュー文が対象とした商品やジャンルを予測する文の分類問題を解くための機械学習モデルを学習させる。そして、対象とする商品の予測確率が上位のレビュー文をその商品の代表的な文とする。さらに、商品予測モデルから抽出された文集合の中に存在するノイズ文を削除するため、商品ジャンル予測モデルと組み合わせる。評価実験によって、提案手法は比較手法より高い評価結果が得られることが分かった。商品レビューの中から商品の特徴を表す文を抽出することで、ユーザが商品選別するときの負担の軽減が期待できる。

キーワード 自然言語処理, Word2Vec, LSTM, E コマース, 商品選別支援

結果と評価について述べ、第5章でまとめを述べる。

1 はじめに

2020年度の電子商取引に関する市場調査[1]によると、新型コロナウイルス下の状況で、人と人との接触を回避すべく、Eコマースが推奨された結果、物販系分野の大幅な市場規模拡大につながった。EC化率は前年比が増加傾向にあり、商取引の電子化が引き続き進展している。

Eコマースは便利な反面、消費者が多くのショップの中から自分の要求に合った商品を見つけ出しにはどうすればよいかわからない場合があるという問題を抱えている。大量なレビューの中から有用な情報を見つけられるように、Eコマースサイトは、キーワードの検索、評価値(おすすめ度)でレビューの検索など機能を提供していることもある。しかし、キーワードの多くは、商品の機能に関するユーザの感想や説明であり、これらのキーワードは、他の類似商品のレビューにも多数存在する可能性もある。また、大量のレビューから商品の特徴を表すキーセンテンスを探し出すことは、ユーザにとって負担となる。

本研究では、ユーザが商品を選別する際に役立つ文をレビューから見つけ出すために、教師ありテキスト分類モデルを訓練する。商品を正しく予測する確率が高い文は、その商品の特徴を表していると仮定し、モデルから抽出したレビュー文の有用性に関する評価実験を行う。本手法によりEコマースサイトにおけるレビューの情報量過多によるユーザの負担を軽減することを目標としている。

第2章では関連研究を挙げ、第3章では提案手法の詳細を述べる。第4章では提案手法を実際のデータを用いて実験を行い、

2 関連研究

2.1 レビューからキーワードの抽出

Eコマースサイトに投稿されたレビューが多い場合には、ユーザが望むような情報を得ることはユーザにとって大きな負担となる。中野[2]らはレビューから属性と意見を抽出して、属性のSVM分類器と意見のSVM分類器を構築し、ユーザが商品の属性と意見を効率的に把握できるインタフェースを提案している。Wang[3]らは、LSTM RNNに基づいた商品のレビューから自動的にキーワードを抽出する手法を提案している。この手法ではアノテーションされたレビュー文を使用し、商品に関係のないレビュー文のフィルタリング分類器を学習させる。フィルタリングされた文のn-gramの頻度に基づいて分類ラベルを決定し、キーワード抽出モデルを学習させる。このキーワード抽出モデルを使って、新しいレビュー文を分類し、各キーワードを商品レビュー全体での出現頻度によって高い順に並べる。キーワードランキングを新規顧客に表示させ、顧客はより良い購買判断ができるようになる。

レビュー要約を利用して、ユーザが商品の主要な情報を短時間で把握でき、商品間の違いを比較する手法も提案されている。立石[4]らは、係り受け関係を利用した抽出ルールを用いて、webの文章から意見を抽出し、レーダーチャートの形式で要約する方式を提案している。小橋[5]らは、自然言語処理ツールStanfordCoreNLPを利用して、レビューから文脈つき評価軸を抽出することにより、より詳細な商品比較ができる手法を検討している。本研究では、分類モデルを利用してレビューの

中の商品選別に役立つ文の抽出に着目する。利用する学習データは、手動で分類ラベルを付けず、その商品 ID とジャンル ID をラベルとして使う。

2.2 商品選別に有用なレビューの抽出

商品選別に有用なレビューを抽出するために、レビューの支持度とレビューに記述されている商品属性の説明の豊富性を考慮する手法が提案されている。

Debanjan [6] らは、高い支持度を持ち、商品のさまざまな属性と感情をよくカバーしているレビューのサブセットを推薦する手法を提案している。この手法では単語分散表現を利用し、レビュー中の類似した商品の属性記述を識別する。レビュー記述を商品の紹介に記載されている属性とマッピングすることで、レビュー集合のポジネガの割合を算出する。動的畳み込みニューラルネットワークを用いて支持度が低い新しいレビューにも対処し、最終的に、ポジネガの割合を利用して、支持度に基づいたレビューランキングからレビューのサブセットを抽出する。

草刈ら [7] は、レビュー評価値の分布を利用し、未購入者のレビューを商品の購入とは無関係な「役に立たないレビュー」の学習正例として使えて役に立たないレビューを除去するという手法の有効性を評価している。大量のレビューが投稿され、かつ評価の分布が高評価に偏っている商品のレビューに注目し、褒めてばかりのレビューが役に立たないレビューの学習例として有望であることを示している。

本研究では、1つのレビューに複数の属性記述と評価が存在する可能性があり、それらの記述と評価が他の商品のレビューにもあるかもしれないことを考慮し、レビューを文に分割し、他の商品のレビュー文と異なる特徴の文を抽出ことを目標としている。

3 提案手法

3.1 提案手法の概要

提案手法では、商品予測モデルが予測したレビュー文が対象とした商品の確率（予測を行うニューラルネットワークモデルの Softmax 値）を用いて文を並べ、上位の文を各商品の代表的な文と仮定する。商品予測モデルから抽出された文集合の中で、商品の特徴を知るのには役に立たない「すぐに届いて役立ちました」「いい買い物できました」「また、お世話になります」等のノイズ文をフィルタリングするために、商品ジャンル分類モデルを訓練する。商品選別の役に立たないノイズ文は、他のジャンルの商品のレビューにも出てくる可能性がある。そこで、ジャンル分類モデルの訓練を行う。商品予測モデルと商品ジャンル予測モデルの結果の確率を組み合わせることで代表的な文を抽出する手法を提案する。

3.2 モデルの全体像

ある商品における代表的な特徴を表しているレビュー文は、その商品に多く現れ、他の商品にはほとんど現れない特徴表現していると考えられる。商品予測モデルの目的は、商品の代表的な特徴文を抽出することである。予測モデルとしては、LSTM(Long Short Term Memory)を利用する。LSTMは1997年にSepp HochreiterとJürgen Schmidhuber [8]によって提案されたニューラルネットワークモデルで、忘却ゲート、選択記憶ゲート、出力ゲートの3つのゲートを有する。LSTMはRNNに存在する長期依存性の問題を解決し、重要な情報をより長い時間記憶できる。

提案手法では、商品レビュー文を商品IDごとにグループ分けし、それぞれのグループの中のレビュー中の文に商品IDのラベルを付ける。図1は商品予測モデルの構造を示す。前処理されたレビュー文の分散表現をLSTMに入力し、最終的には全連結層とSoftmax関数によって、レビュー文に対応する商品ラベルの確率を予測する。交差エントロピー誤差(Cross Entropy Loss)損失関数を用いて、実際の出力と目的の出力の近さを判断する。各商品グループに含まれるレビュー文を出力の確率に応じてランク付けし、確率スコアが高いレビュー文を抽出する。

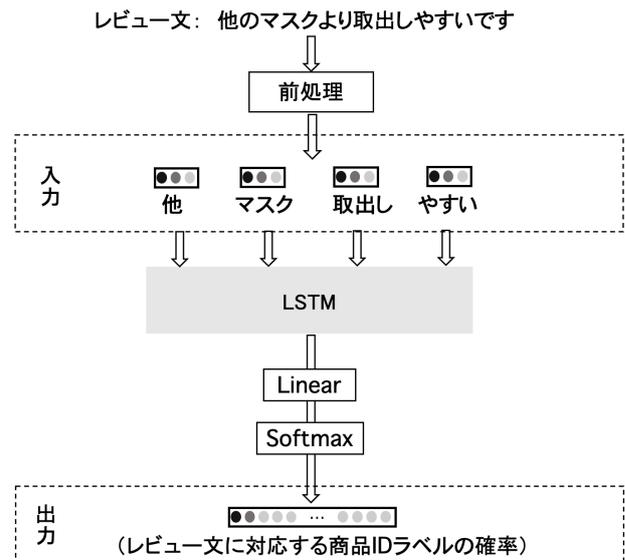


図1 商品予測モデルの構造

3.3 代表的なレビュー文の決定

商品予測モデルが抽出したレビュー文に含まれるノイズ文を削除するために、ジャンル予測モデルを訓練する。すなわち、商品予測モデルとジャンル予測モデルの結果と合わせて代表的な文を決定する。ジャンル予測モデルは、商品レビューをジャンル別（1つのジャンルに複数の商品がある）にグループ分けする。グループ中のレビュー文は、ジャンルに応じてラベル付けられる。ジャンル予測モデルと商品予測モデルは両方ともLSTMを使い、出力は文の対応するジャンルラベルの確率であ

る。図2に示すように、商品AとジャンルBに属するレビュー文は商品予測モデルとジャンル予測モデルに入力し、2つモデルの出力結果を掛け合わせる。この結果に基づいて再度商品グループ中のレビュー文を並び替え、スコアの高いレビュー文をその商品の代表的なレビュー文とする。

レビュー文 s の商品 i に対するスコア $score(s, i)$ を以下の式 (1) で定義する。

$$score(s, i) = p(i|s) \times p(g(i)|s) \quad (1)$$

ここで、 $p(i|s)$ はレビュー文 s が与えられたときに、それが商品 i のレビューに含まれると予想される確率であり、 $p(g(i)|s)$ はレビュー文 s が与えられたときに、それが商品 i のジャンル $g(i)$ に属する商品のレビューであると予想される確率である。本研究では、全ての商品 i について、その商品が所属するジャンルが一意に定まるものとし、 i が属するジャンルを $g(i)$ と表す。商品 i がジャンル g に属さないとき以下の式 (2) ようになる。

$$p(i|s, g) = 0 \quad (g \notin g(i)) \quad (2)$$

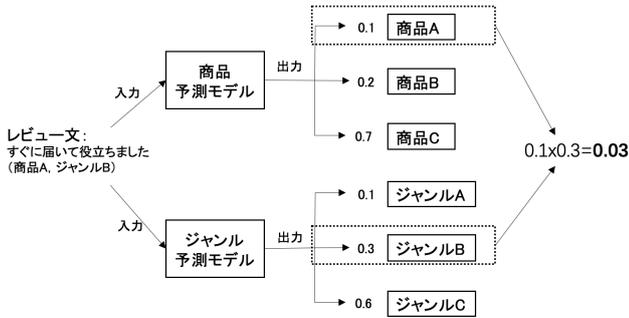


図2 結果の組合せ

4 実験と評価

4.1 データセット

本研究では評価実験のために国立情報学研究所 (NII) が提供している「楽天公開データ」[9]を使用する。楽天データセットは全商品データ約2億8300万商品、商品レビューデータ約7000万件、ショッピングレビューデータは約2250万レビューが含まれる。本研究の学習用データでは、「楽天市場: 商品情報, みんなのレビュー・口コミ情報」の2019年の「商品レビュー」である。使用したコラムは「商品名」「商品ID」「商品ジャンルID」「レビュー内容」である。

4.2 テキスト前処理

文書の前処理の手順を図3に示す。最初に入力された商品レビュー段落は記号「。?… () ~!」で文に分割する。その後、MeCabと辞書mecab-ipadic-NEologedで形態素解析し、名詞、動詞、形容詞と副詞を抽出する。最後にストップワード(例えば「し」「あと」「あり」などの共通のワード)を除去す

る。

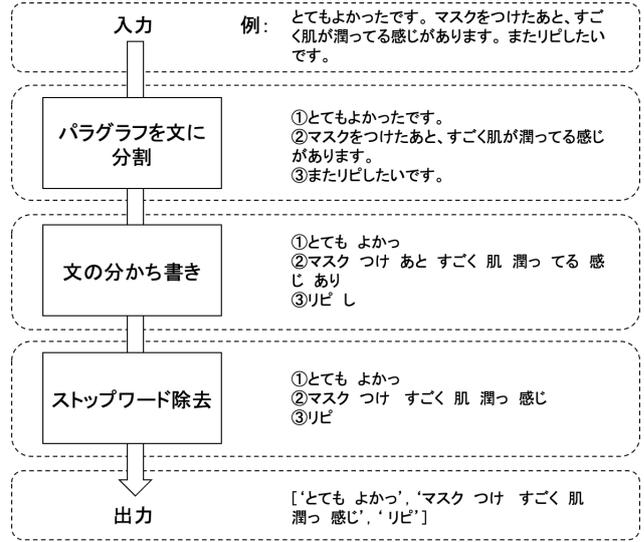


図3 文書の前処理の流れと例

機械学習は基本的には数値的な処理である。単語や文字などの記号をベクトルや行列で計算処理できる分散表現に変換する必要がある。本研究では、公開されている「日本語 Wikipedia エンティティベクトル」を使う。これは、日本語版 Wikipedia の本文全文から学習した、単語、および Wikipedia で記事となっているエンティティの200次元の Word2vec 分散表現 [10] である。学習済み Word2vec 分散表現を用いて、前処理した単語をベクトル化する。

4.3 比較手法

4.3.1 商品予測モデルだけにに基づく代表的なレビュー文の抽出

比較手法の1つは、商品予測モデルのみ利用して文を出力し、提案手法と比較する。LSTMの有効性を比較するために、全結合層のみから構成される Basic モデルを利用して商品予測モデルを訓練する。

4.3.2 tf-idfに基づく代表的なレビュー文の抽出

もう1つの比較手法は、文の tf-idf を利用する手法である。tf-idf は、単語の重要度を測る手法である。tf (term frequency, 単語の出現頻度) は各文書においてその単語がどのくらい出現したのかを表す。idf (inverse document frequency, 逆文書頻度) は単語がいくつかの文章内で共通して使われているかを表す。単語が他の文書にも多く出現しているほど idf 値は小さくなり、単語が他の文書にあまり出現していないほど idf 値は大きくなる。tf-idf は情報検索分野 [11] やテキストマイニング分野で広く使われている。

式 (3) に tf-idf に基づく文 s の重み $Score_{tfidf}(s)$ の定義を示す。 $N(s)$ は文の中の単語数、 $t(s)$ は文 s に含まれる単語集合を表す。単語集合の tf-idf の平均値から文の tf-idf を算出す

る。tf-idf 値が大きい文を代表的なレビュー文として選択する。

$$Score_{tfidf}(s) = \frac{\sum_{x \in t(s)} tfidf(x, s)}{N(s)} \quad (3)$$

4.4 実験方法

商品予測モデルの訓練は、キャリアバッグジャンルの中 15 件商品の約 3000 件レビューとフェイスパックジャンルの中 56 件商品の約 10000 件レビューを利用する。ジャンル予測モデルの訓練は、4つのジャンル (キャリアバッグ、エコバッグ、フェイスパック、化粧水) を使い、各ジャンルに対応して約 30000 文を利用する。

本研究の目的は、ある商品のレビュー文の集合から、その商品の特徴を表す文を抽出することであるため、モデルの精度を測るための検証データを設定せず、その商品のレビュー文の集合のすべてのデータをモデルのトレーニングデータとして使用する。レビュー文に対応する商品 ID ラベルにつける方法の欠点は、ノイズ文 (前述したように、どの商品のレビューの中にも出てきそうな記述) が多く存在する可能性が高いである。このようなノイズデータは学習されないため、モデルが予測する際に、ノイズ文に対応する商品予測の確率は低くなると考えられている。

提案手法と比較手法で抽出した文の有用性についてアンケート評価実験を行った。キャリアバッグジャンルとフェイスパックジャンルからそれぞれ3つの商品を選例としてランダムに選択する。評価される文では、各商品のスコアが上位5つレビュー文である。

被験者は、ネットショッピングの経験がある 20~30 代の 9 人である。アンケートは、まず、商品名を被験者に提示する。そして、各手法で抽出されたその商品のレビュー文を順番を変えて、被験者に提示する。被験者にこのレビュー文の内容が「商品の特徴に関する情報がどのくらい含まれているか」について五段階の評価をしてもらった。また「買い物するとき参考になるか」について回答してもらった。

4.5 実験結果

図 4 は 50epoch の LSTM キャリーバッグジャンル商品予測モデルの精度と損失、図 5 は 50epoch の LSTM フェイスパックジャンル商品 ID 予測モデルの精度と損失を示す。図 6 は 20epoch の LSTM ジャンル予測モデルの精度と損失を示す。

商品予測モデルだけにに基づく比較手法は、20epoch、50epoch の Basic 商品予測モデルと 20epoch の LSTM 商品予測モデルを訓練する。まだ、提案手法の 50epoch の LSTM 商品予測モデルも用いてレビュー文を抽出する。

評価項目は、評価値、参考になる数と適合率である。表 1 は各手法の各評価項目のアンケート平均値を示す。表 2 は提案手法と比較手法で抽出したキャリアバッグの3つ商品のスコアが一番高いレビュー文と被験者の評価値の平均を示す。図 7 示した評価値は、文に対する被験者の評価の平均値である。図 8 示

した参考になる数は、文に対する被験者が「参考になる」と思う文数の平均である。図 9 示した適合率は、被験者の評価値が 2 以上の文を正解として、各手法で抽出した文の正解の割合である。表と図の中で、P-LSTM*G-LSTM は提案手法、P-Basic は Basic 商品予測モデル、P-LSTM は LSTM 商品予測モデル、P-tfidf は tf-idf に基づく代表的なレビュー文の抽出手法である。

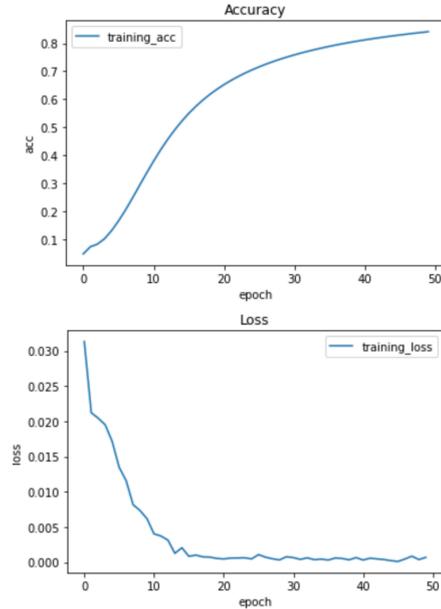


図 4 50epoch の LSTM キャリーバッグジャンル商品予測モデルの精度と損失

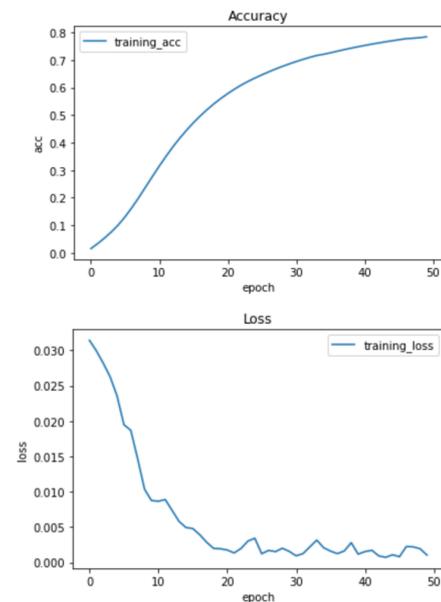


図 5 50epoch の LSTM フェイスパックジャンル商品予測モデルの精度と損失

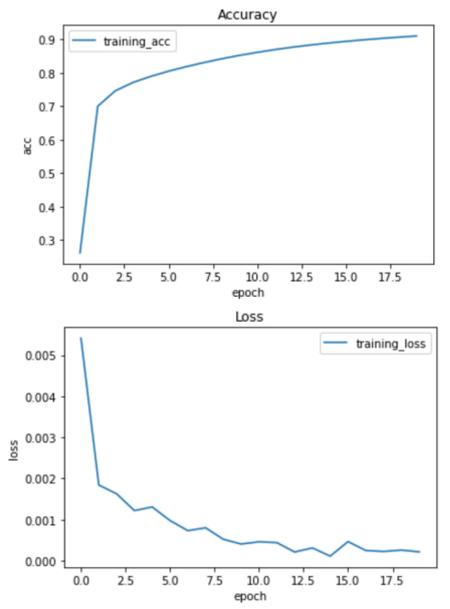


図 6 20epoch の LSTM ジャンル予測モデルの精度と損失

表 1 各手法の評価項目の平均値

	P-LSTM* G-LSTM	P-Basic (epoch=20)	P-Basic (epoch=50)	P-LSTM (epoch=20)	P-LSTM (epoch=50)	P-tfidf
評価値	3.56	2.31	2.67	2.94	2.91	1.58
参考になる数	4.12	2.75	2.89	3.51	3.29	1.06
適合率	0.84	0.45	0.56	0.65	0.61	0.17

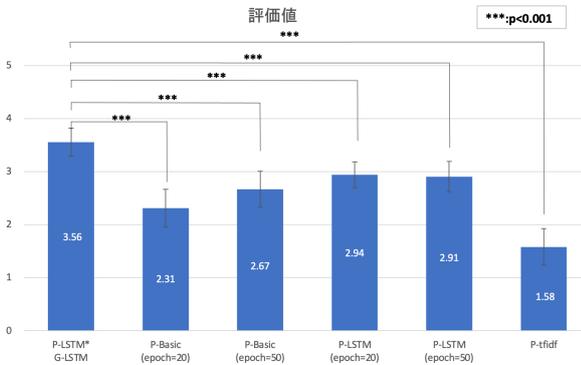


図 7 各手法に対する評価値

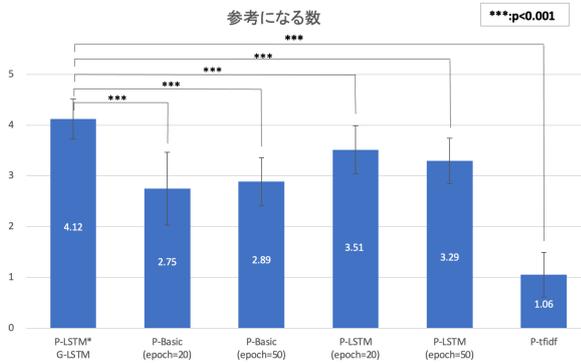


図 8 各手法で抽出した参考になる文の平均値

表 2 各手法で抽出したのキャリーバッグジャンルの商品の文の例

モデル	レビュー文	評価値
P-LSTM*	1. 作りもしっかりしていて、キャリーもくるくと滑らかに動きます	4.56
G-LSTM	2. 他の方のレビューにもある通り色味は基本的に写真より少し深いイメージを持たれると思います。	4.56
	3. 色も綺麗でとても喜んでいました。	3.22
P-Basic (epoch=20)	1. スクラッチガンメタの LM を購入しました。	1.33
	2. この度 S サイズを購入しました使い勝手は大変良いと思います。	2
	3. お値段を考えれば充分な商品でした。	1.78
P-Basic (epoch=50)	1. フィリピンへ 9 日間、大人一人子ども二人で行きました。	1.67
	2. しいていえば、ダイヤルの暗証番号があれば、音が少し違うので、バレるかも (笑)。	3.44
	3. でも 10 年以上使っていると、外側が汚くなってくる。	2.89
P-LSTM (epoch=20)	1. ロイヤルレッド、きれいな赤です。	3.56
	2. 汚れや傷の事も考え悩みましたが、色合いが気に入りブリリアントホワイトを購入しました。	4.33
	3. 高校生の娘が 8 日間の海外研修に持参しました。	2
P-LSTM (epoch=50)	1. 絶対無理と言われ、単品購入。	1.78
	2. グリーン系の S サイズ購入しました。	1.56
	3. まだ使用していないので、キャスターの動きがスムーズかは分かりませんが、ファスナーはスムーズに開閉できました。	4.67
P-tfidf	1. ジェムストーンパープルの LM サイズを購入。	1.44
	2. S サイズだと 1 - 2 泊用です。	2.44
	3. ブラックの LM サイズを購入しました。	1.11

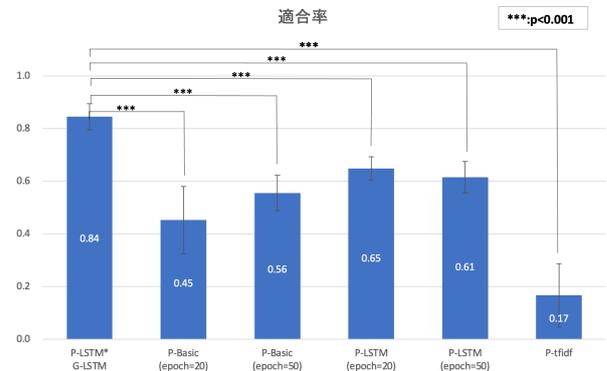


図 9 各手法に対する適合率

4.6 考 察

アンケートの結果を見ると、提案した商品予測モデルとジャンル予測モデルを組み合わせる手法の評価値、参考になる数と適合率が最も高い。一方、文の tf-idf に基づく結果が最も低いことが分かった。Basic モデルと比較して、LSTM モデルはより高い評価値、参考になる数と適合率を得ており、LSTM ネットワーク構造がより良い性能を持つことを示す。また、20epoch の LSTM と 50epoch の LSTM モデルの結果を比較すると、20epoch の LSTM の方が良い結果が得られた。図 4 と図 5 のモデルの損失からわかるように、20epoch 以降モデルの損失はほとんど減少せず、モデルは過学習状態にあり、過学習はモデルの結果に負の影響を及ぼしたことを考えている。

提案手法で抽出したレビュー文のすべてが高評価を得たわけではなく、レビュー文の中には除去しきれなかったノイズ文もあり、モデルの精度を上げていくことで改善できる可能性がある。また、提案手法の低評価のレビュー文の中には、特徴の記述が少ない文もあり、レビュー文の長さを制限することでこのような低評価のレビュー文削除される可能性があると考えている。

5 おわりに

本研究は、商品選択支援のため、レビュー中の文を用いて、商品予測モデルと商品ジャンル予測モデルを訓練し、2つモデルが予測したレビュー文に対応する商品の確率から商品の特徴を表すレビュー文を決定することを提案した。アンケート評価試験を経て、提案手法はより高い結果が得られることが分かった。今後の課題は、商品の特徴を最もよく表し、ユーザが商品を選択するのに役立つレビュー文を見つけるためには、商品の特徴に対する感情の分布や、レビューを支持しているユーザの数など他の指標を考える必要があると考えている。

謝 辞

本研究は JSPS 科研費 19H04219 の助成を受けたものです。また、本研究では、国立情報学研究所の IDR データセット提供サービスにより楽天グループ株式会社から提供を受けた「楽天データセット」(https://rit.rakuten.com/data_release/) を利用させていただきました。

文 献

- [1] 経済産業省：令和 2 年度電子商取引に関する市場調査、<https://t.hk.uy/aazn>. 2021.
- [2] 中野裕介, 湯本高行, 新居学, 上浦尚武ほか. 機械学習による商品レビューの属性-意見ペアの抽出. 研究報告データベースシステム (DBS), Vol. 2015, No. 14, pp. 1–8, 2015.
- [3] Y Wang and J Zhang. Keyword extraction from online product reviews based on bi-directional lstm recurrent neural network. In *2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, pp. 2241–2245. IEEE, 2017.
- [4] 立石健二, 福島俊一, 小林のぞみ, 高橋哲朗, 藤田篤, 乾健太郎, 松本裕治ほか. Web 文書集合からの意見情報抽出と着眼点に基づく要約生成. 情報処理学会研究報告情報学基礎 (FI), Vol. 2004, No. 93 (2004-FI-076), pp. 1–8, 2004.
- [5] 小橋賢介, 酒井哲也. 商品比較のための文脈つき評価軸抽出の検討. pp. G6–2. DEIM, 2020.
- [6] Debanjan Paul, Sudeshna Sarkar, Muthusamy Chelliah, Chetan Kalyan, and Prajit Prashant Sinai Nadkarni. Recommendation of high quality representative reviews in e-commerce. In *Proceedings of the Eleventh ACM Conference on Recommender Systems, RecSys '17*, p. 311–315, New York, NY, USA, 2017. Association for Computing Machinery.
- [7] 草刈祐子, 鈴木孝彦, 廣川佐千男ほか. 有用な商品レビュー抽出のための過賞賛レビュー除去フィルターの性能評価. 研究報告数理モデル化と問題解決 (MPS), Vol. 2018, No. 12, pp. 1–5, 2018.
- [8] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, Vol. 9, No. 8, pp. 1735–1780, 1997.
- [9] 国立情報学研究所. 情報学研究データリポジトリ：楽天データセット, <https://www.nii.ac.jp/dsc/idr/rakuten/>. 2021.
- [10] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pp. 3111–3119, 2013.
- [11] Juan Ramos, et al. Using tf-idf to determine word relevance in document queries. In *Proceedings of the first instructional conference on machine learning*, Vol. 242, pp. 29–48. Citeseer, 2003.