

Title	意見とその根拠を含む有用な商品レビューの分類
Author(s)	荘, 博閔
Citation	
Issue Date	2024-03
Type	Thesis or Dissertation
Text version	author
URL	<a href="http://hdl.handle.net/10119/18902">http://hdl.handle.net/10119/18902</a>
Rights	
Description	Supervisor: 白井 清昭, 先端科学技術研究科, 修士(情報科学)

## 概要

ネット上の商品レビューは、購入意思を決定する消費者と商品の評判を知りたい企業の双方にとって、重要な役割を果たしている。しかし、すべてのレビューが有益な情報を提供するわけではない。大量のレビューから有用なレビューを発見するのは多くの時間を要するため、レビューの有用性を自動的に判定する研究が注目を集めている。一方、レビューの有用性は様々な観点から定義される。レビュアーの意見に対する理由や根拠は重要な特徴の一つである。例えば「デザインがカッコいいので気に入った」というように、意見だけでなくその根拠も述べているレビューは、消費者と企業にとって有益である。しかし、このような観点からレビューの有用性を判定する研究は、これまで行われていなかった。

本研究は、多くの製品レビューから有益なレビューを選別するために、レビューにユーザの意見とその根拠が含まれているかどうかを判定することを目的とする。この目的を達成するために、以下の2つのタスクを実行する。1つ目は「根拠関係分類タスク」である。これは、入力として与えられたレビュー内の文や節の組がユーザの意見とその根拠を含んでいるかどうかを分類するタスクである。2つ目は「商品言及分類タスク」である。根拠を含むと判定した節の組に対し、それが商品に対する意見の根拠であるかを判定するタスクである。商品に関係のない記述に対する根拠が書かれていても有用なレビューとは言えないため、2つ目のタスクが必要である。

根拠関係分類タスクの提案手法は以下の通りである。まず、与えられたレビューから依存関係がある節のペアを抽出する。次に、それぞれの節のペアが「根拠関係-節ペア」（意見とその根拠が含まれる）であるか「非根拠関係-節ペア」であるかを分類する。根拠関係分類タスクを解く4種類の分類器を学習する。(1) ルールベースの方法。入力とした節の組のいずれかに談話標識「から」もしくは「ので」が出現していれば、その節ペアを根拠関係-節ペアと判定し、それ以外は非根拠関係-節ペアと判定する。(2) Bidirectional Encoder Representations from Transformers(BERT)。訓練データを使用してファインチューニングする。(3) BERTベースのIntermediate Fine-Tuning(IFT)。まず、比較的大きな関連タスクのデータセットを使用してBERTモデルをファインチューニングし、その後、対象タスクのラベル付きデータセットを使用して再びファインチューニングする。(4) ルールベースとBERTモデルを組み合わせたハイブリッド手法。最初に、ルールベースの手法を用いて入力の節ペアが根拠関係-節ペアか否かを判定する。非根拠関係-節ペアと判定されたとき、BERTによる根拠関係分類器を用い、その判定結果を最終の判定結果とする。

根拠関係分類器を学習するために3つのデータセットを用いる。1つ目は既存の談話関係解析データセットである京都大学ウェブ文書リードコーパス(KWDLC)である。KWDLCにおいて、「原因/理由」タグでラベル付けされた文の組を正例(根拠関係-節ペア)として抽出する。また、「原因/理由」以外の談話関係タグが付与された文の組を負例(非根拠関係-節ペア)として抽出する。KWDLCは文間の談話関係が人手で付与されたデータセットであるが、本研究が対象とする根拠関係

分類タスクに必ずしも適しているわけではない。また、KWDLICと商品レビューはドメインが異なる。KWDLICでは様々なウェブ文書に対して注釈付けされているが、根拠関係分類タスクのドメインは商品レビューである。2つ目は談話標識によるデータセットである。ラベルなしのレビューから節の組を抽出し、もし節に「から」または「ので」の談話標識が含まれていれば、意見とその根拠が含まれた正例(根拠関係-節ペア)として抽出する。また、これらの談話標識を含まない節の組を負例(非根拠関係-節ペア)として抽出する。KWDLICとは異なり、談話標識を手がかりに構築したデータセットはイン・ドメインのデータである。しかし、正例と負例は自動抽出されているため、誤りを含む可能性があることに注意する必要がある。3つ目はChatGPTによるデータセットである。意見とその根拠の両方を含む節をシード節として用意する。それぞれのシード節に対し、ChatGPTにより、それと類似しかつ意見とその根拠を含む新しい節を40個生成する。ChatGPTを用いて生成した節は、同じジャンルのレビューからランダムに選ばれた別の節と結びつけられ、正例の根拠関係-節ペアが作られる。また、談話標識によるデータセットと同じ方法で負例を作成する。なお、IFTでは、最初の学習段階でアウト・ドメインのデータセットであるKWDLICを用い、第2の学習段階でイン・ドメインの2つのデータセットを用いる。

商品言及分類タスクに対する提案手法は、キーワードマッチングに基づく単純な手法である。まず、商品カテゴリに関連するキーワード集合をあらかじめ構築する。ここでの商品カテゴリとは、「食品」「本」「家電」など、商品の種類を分類したものである。キーワードとは、そのカテゴリの製品に関するレビューで頻繁に使用される重要な単語を指す。本タスクの分類器は、入力された節ペアに商品カテゴリのキーワード集合のいずれかのキーワードが出現するときには、その節ペアの意見は商品に関連があるとみなす。逆に、節ペアに商品カテゴリのキーワードがひとつも出現していないときは、その節ペアの意見は商品に関連がないとみなす。

商品カテゴリに関するキーワードを抽出するために5つの手法を用いる。(1)TF-IDF. TF-IDFスコアの高い単語が抽出される。(2)TF-ENT. TF-ENTスコアの高い単語が抽出される。ここでのENTとは、単語の商品カテゴリに関する顕現性(単語が1つの商品カテゴリのみで使用される傾向がどれだけ強い)を評価するスコアであり、単語が商品カテゴリのレビューに現れる確率の確率分布のエントロピーによって算出される。(3)YAKE!. 既存の教師なしキーワード抽出手法YAKE!によって抽出された単語。(4)TF-IDF+YAKE!. TF-IDFとYAKE!の組み合わせによって抽出された単語。(5)TF-ENT+YAKE!. TF-ENTとYAKE!の組み合わせ手法である。各手法ごとに候補単語のスコアを計算し、その上位200件の単語をキーワード集合として取得する。

提案手法を評価するためにいくつかの実験を行った。根拠関係分類タスクのテストデータは、506個の節ペアに正解ラベルを人手で付与して作成した。また、商品言及分類タスクのテストデータは、上記のデータセットに含まれる186の根拠

関係-節ペアに対して、正解ラベルを人手で付与して作成した。

根拠関係分類タスクの実験結果から、自動的に構築した2つのデータセットを用いることで根拠関係分類タスクの性能が大きく改善することを確認した。最高のF1スコアはIFTが達成した0.71であり、既存のデータセットKWDLICで学習したBERTモデルより0.09ポイント高かった。これにより、異なる種類のデータセットを組み合わせる我々のアプローチが意見に対する根拠関係の分類に有効であることが示された。さらに、これらの異なる種類のデータセットを利用する際には、データセットの単純な結合よりもIntermediate Fine-Tuningの方が適していることがわかった。一方、ルールベースの手法とBERTモデルを組み合わせたハイブリッド手法は、BERTモデルを上回らなかった。これは、ルールベースの手法は評価データの23%しか分類できないが、BERTベースのIFTはこれらのデータに対する分類の精度が高い(0.79)ためである。

商品言及分類タスクについては、TF-IDFとTF-ENTを比較すると、TF-IDFの方が再現率とF1スコアが高かった。また、YAKE!はTF-IDFとTF-ENTよりも高い性能を示した。特に再現率の差が大きかった。最後に、TF-IDFとYAKE!を組み合わせるキーワードを抽出する手法が最も有効であることがわかった。この手法の再現率は0.91、F1スコアは0.89であり、5つの手法の中で最も高く、精度も他の手法と比べて大きな差はなかった。また、2つのキーワード抽出手法を組み合わせるTF-IDF+YAKE!は、ひとつの抽出手法だけを用いる場合と比べて、より多様なキーワードを獲得していることを確認した。このことが、TF-IDFとYAKE!を組み合わせることによって商品言及分類タスクの性能が向上する主な原因であると考えられる。