

製品の部品別意見収集のための ラベル抽出とデータ拡張について

按田 将吾^{1,a)} 菊地 真人^{1,b)} 大園 忠親^{1,c)}

概要：複数部品からなる製品のレビューは、各部品への意見が内部に混在し、それらは多くの表現揺れや表記揺れを含む。本研究は、レビュー中の文に対し文中で言及されている部品とアスペクトのラベルを付与することで部品別の意見として収集することを目指している。本稿では、BERT に基づくマルチラベル分類器による部品別意見収集手法および、製品種類毎に異なる部品とアスペクトを考慮するためのパターンマッチングに基づくラベル抽出手法、WordNet を用いた類似語入れ替えによるデータ拡張手法について述べる。評価実験では、部品別意見収集手法およびラベル抽出手法、データ拡張手法を評価した。実験結果として、部品別意見収集手法のマクロ F1 値がロードバイクに関しては 0.86 まで上昇したが、ノート PC に関しては 0.69 であった。

キーワード：意見抽出、情報抽出、BERT、パターンマッチング、データ拡張、WordNet

1. はじめに

オンライン上での製品購入検討時に EC サイト上のレビューを用いて情報収集する場合がある。特に複数部品から構成されている自転車や PC などの製品の購入検討時は、製品中の各部品ごとの情報を収集し、他製品との比較や関連製品の購入検討を行う。しかし、複数部品から構成されている製品のレビューは、内部に製品中の各部品に対するレビュー投稿者の意見が混在し、部品別の効率的な意見収集は妨げられる。

この問題を解決するために、先行研究として BERT に基づく分類器を用いてレビューから部品別の意見を収集するシステムを開発した [1]。先行研究では、レビュー中の文（以下、レビュー文）に対して、文中で言及されている部品（以下、言及部品）と言及部品に対する“見た目”や“耐久性”等の評価の視点（以下、アスペクト）を BERT に基づく分類器によって自動で付与し、部品別意見としての収集を試みた。論文では、データセット作成のために分類ラベルを入手する必要があるが、分類ラベルは未知であるため 1) データセットに用いる言及部品とアスペクトを意味するラベルをレビューを手がかりに抽出した（以下、ラベル

抽出手法）。ラベル抽出手法を用いて作成したラベルをレビュー文へと人手で付与することでデータセットとした。2) レビューから製品中の部品別の意見を収集する手法（以下、部品別意見収集手法）を考案した。この手法では分類器によって、レビュー文に分類ラベルを付与する。作成したデータセットを用いて分類器の性能を検証した。また、部品別意見収集手法内の分類器を訓練する際、言及部品とアスペクトの組み合わせによって生じた訓練データ内の偏りは分類性能を低下させた。従って、3) 訓練データ内の言及部品とアスペクトの組み合わせによる偏りを低減するために訓練データを拡張した（以下、データ拡張手法）。

しかし、先行研究では、ラベル抽出手法の有効性、レビュー文中に部品名が明示されないレビュー文に対する部品別意見収集手法の分類性能、論文で具体例として扱ったロードバイク以外の製品種類に対する部品別意見収集手法の分類性能の 3 つの点を評価していない。

本稿では、本研究のラベル抽出手法の有効性、部品名を含まないレビュー文に対する部品別意見収集手法の分類性能、部品別意見収集手法およびラベル抽出手法、データ拡張手法のロードバイクではない他の製品種類における分類性能について述べる。評価実験として、先行研究で用いたロードバイクのラベル付きレビュー文のデータセットに加えて、新たに作成したノート PC のデータセットを用いて、ラベル抽出手法によって抽出したラベルと EC サイト上の部品やアスペクトを意味する指標との一致割合の評価、部

¹ 名古屋工業大学
Nagoya Institute of Technology

a) anda@ozlab.org

b) kikuchi@nitech.ac.jp

c) ozono@nitech.ac.jp

品別意見収集手法の分類性能に対するデータ拡張の効果を評価した。また、部品別意見収集手法の部品名が含まれないレビュー文への分類性能を評価した。

本稿の以降の構成を以下に示す。2.では、本研究の関連研究について述べる。3.では、データセット作成に用いたラベルをパターンマッチングをもとにレビューから抽出するラベル抽出手法について説明する。4.では、BERTに基づく分類器によるラベル付与を用いた部品別意見収集手法について説明する。5.では、部品別意見収集手法内で用いた類似語入れ替えによるデータ拡張手法について説明する。6.では、部品別意見収集手法を評価した3つの評価実験について説明する。7.では、考察と今後の展望について説明する。8.では、本稿についてまとめる。

2. 関連研究

本研究の関連研究である、レビュー閲覧支援に関する研究、レビュー中の表現パターンに着目した評価抽出についての研究、WordNetを用いたデータ拡張の研究について紹介し、本研究との関連点について説明する。

新里ら [2] は、レビュー中の意見が製品自体に対してだけでなく、配送過程や購入時の店舗の対応等に向けられていることに着目し、アノテーションデータを学習することで、レビューを意見の対象別に分類した。新里らは、レビュー中で言及されている対象が製品自体か、その購入過程の他の要素かの分類を試みた。本研究では、その言及対象が製品自体であった場合に、製品中のどの部品であるかの分類を試みた。

小林ら [3] は、レビュー中の意見が特定のパターンに基づいて表現されることに着目し、レビュー中の意見を収集するための辞書をパターンマッチングを用いた手法で半自動的に作成した。本研究のラベル抽出手法では、小林らの研究をもとに、パターンマッチングを用いて、レビューからの部品別意見収集のためのラベルを抽出した。

Vosoughi ら [4] は、Twitter 上のツイートの意味と内容に基づいたベクトル表現の生成について述べた。論文では、CNN-LSTM によるエンコーダ・デコーダモデルによって実現を試みた。学習の際、訓練データとして用いたツイート中の語を、WordNet による類似語と入れ替えることで訓練データを拡張した。本研究での研究対象であるレビューは、Vosoughi らの扱ったツイートと同じ自由記述のテキストであり、長文と短文が混在していることや、不完全な文が多いことが特徴である。本研究では、Vosoughi らの手法をもとに、レビュー文から類似レビュー文を生成する手法を用いて訓練データ内の偏りを低減した。

3. ラベル抽出手法

レビュー中の言及部品およびアスペクトを抽出するためのパターンマッチングによるラベル抽出手法について説明

する。本研究では、言及部品を意味するラベル (以下、言及部品ラベル) およびアスペクトを意味するラベル (以下、アスペクトラベル) を、分類器を用いてレビュー文へと自動で付与することでレビュー文を部品別意見とした。分類器訓練のためのデータセットを作成の際、製品種類毎に異なる部品とアスペクトを考慮したラベルをレビュー文へと人手で付与しデータセットとするために、レビューから言及部品ラベルおよびアスペクトラベルを抽出または作成した。

2種類のラベルを抽出する前段階として、パターンマッチングによって2種類のラベル候補を抽出した。言及部品ラベルの抽出元として言及部品ラベル候補を、アスペクトラベルの作成元としてアスペクトキーワードを抽出した。

関連研究である小林らは論文中で、レビュー中で製品に対する評価表現が以下のように出現するとした。

- 〈対象〉の〈属性〉は〈評価値〉

以下に示すように、レビュー中で部品名は〈対象〉、アスペクトは〈属性〉または〈評価値〉として出現する。

- キーボードの反応が悪い
- タッチパッドが壊れた
- 画面が大きい

よって、本研究では以下のようなパターンを定め、レビューから言及部品ラベル候補とアスペクトキーワードを抽出した。

- “名詞 A” の “名詞 B”
- “名詞 A” が “動詞 B”
- “名詞 A” が “形容詞 B”

名詞 A として出現した語を言及部品ラベル候補、名詞 B・動詞 B・形容詞 B として出現した語をアスペクトキーワードとした。また、以下のような語はストップワードとした。

- 文字数が 1 文字の語
- MeCab*1による解析の品詞分類 1 が一般では無い語ラベルの抽出は以下の手順で行う。

手順 1 同種類の製品に対するレビューを複数収集。

手順 2 手順 1 で収集したレビューを “。” と、改行によってレビュー文へと分割。

手順 3 パターンマッチングにより言及部品ラベル候補とアスペクトキーワードを抽出。

手順 4a 言及部品ラベル候補から言及部品ラベルを人手で抽出。

手順 4b アスペクトキーワードをもとにアスペクトラベルを人手で作成。

ノート PC の言及部品ラベルとアスペクトラベルをレビューから抽出した例を説明する。表 1 は Amazon*2 上の複数のノート PC のレビューから手順に沿って収集したレビュー文 9000 件をもとに、抽出した言及部品ラベル候補の出現回数上位 10 語である。言及部品ラベル候補には、製

*1 <https://taku910.github.io/mecab/>

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

表 1 抽出した言及部品ラベル候補の例 (出現回数上位 10 語)

全 525 種類	出現回数
パソコン	142
画面	117
キー	81
値段	79
キーボード	70
バッテリー	52
ソフト	50
電源	47
オフィス	44
価格	43

表 2 ノート PC の言及部品ラベル

キーボード	ディスプレイ	フレーム	ファン
ポート	バッテリー	マウス	CPU
スピーカー	ストレージ		

表 3 抽出したアスペクトキーワードの例

言及部品ラベル	アスペクトキーワード
キーボード	反応, 小さい, 位置, 左右
ディスプレイ	綺麗, キレイ, フリーズ, ノイズ
フレーム	材質, 歪む, サイズ, 薄い
バッテリー	持つ, 長持ち, 寿命
ポート	少ない, もう少し

表 4 ノート PC のアスペクトラベル

アスペクトラベル	詳細
耐久性	部品の欠損に関する意見
機能性	部品の処理性能に関する意見
使い心地	触覚や違和感などに関する意見
サイズ	大きさや厚さに関する意見
重量	重量に関する意見
存在	部品自体の存在や個数に関する意見
見た目	視覚的な情報に関する意見

品自体とは直接関係のない“値段”や“オフィス”などの語や、ノート PC の部品名ではない“パソコン”や“ソフト”などの語が多く含まれるため、人手でこれらの語を削除した。表 1 をもとに人手で抽出した言及部品ラベルを表 2 に示す。ノート PC の言及部品ラベルとして、“キーボード”や“ディスプレイ”などの 10 種類の語を人手で抽出した。

表 3 は先述の言及部品ラベル候補と同条件で抽出したアスペクトキーワードの一部である。アスペクトキーワードには、“キレイ”と“綺麗”などの表記揺れや“持つ”と“長持ち”などの表現揺れが多く含まれる。アスペクトキーワードをもとに人手で数個のアスペクトラベルを作成する。表 3 をもとに作成したアスペクトラベルを表 4 に示す。ノート PC のアスペクトラベルとして、“耐久性”や“機能性”などの 7 種類の語を人手で作成した。

4. レビューからの部品別意見収集手法

BERT に基づくマルチラベル分類器を用いたレビューか

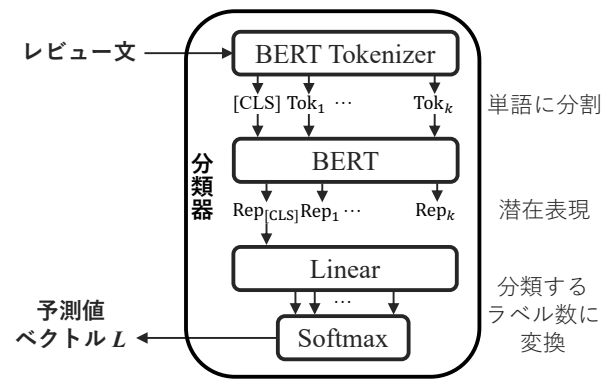


図 1 分類器モデル

らの部品別意見収集手法について説明する。

本研究では、レビュー文に対して、言及部品ラベルとアスペクトラベルを付与することで、製品に対する部品別の意見としてレビュー文を収集した。例えば、ノート PC に対する“画面が大きくとても綺麗”というレビュー文の場合、言及部品ラベルとして“ディスプレイ”を、アスペクトラベルとして“サイズ”と“見た目”を付与し、以下のような形で部品別意見とする。

- 部品別意見

- 製品名：ノート PC の製品名
- 言及部品：ディスプレイ
- アスペクト：サイズ・見た目
- レビュー文：“画面が大きくとても綺麗”

2 種類のラベルは日本語学習済み BERT モデル^{*3}に基づくマルチラベル分類器をファインチューニングしたものによって付与した。言及部品ラベルを付与するための言及部品分類器と、アスペクトラベルを付与するためのアスペクト分類器を作成した。

2 つの分類器で用いた分類器モデルを図 1 に示す。それぞれの分類器では、入力としてレビュー文を入力し、BERT Tokenizer によってトークン単位に分割したのちに BERT に伝播させ、Linear 関数、Softmax 関数によって事前に定義されたラベルの数に次元に変換した。出力として予測値ベクトル L を出力する。

次に、部品別意見収集手法全体のモデルを図 2 に示す。レビュー文を 2 種類の分類器に入力し、言及部品分類器では、言及部品ラベル数の次元を持った予測値ベクトル L を、アスペクト分類器では、アスペクトラベル数の次元を持った予測値ベクトル L' を出力する。 L および L' の各成分は $[0,1]$ の実数値を取る。従って、 L および L' を結合し、一定の閾値を設けて、全ての値を 0 または 1 変換し予測値ベクトル L_{bi} とする。 L_{bi} 内の値が 1 である要素に対応するラベルをレビュー文へと付与し、部品別意見とする。

^{*3} <https://huggingface.co/cl-tohoku/bert-base-japanese-whole-word-masking>

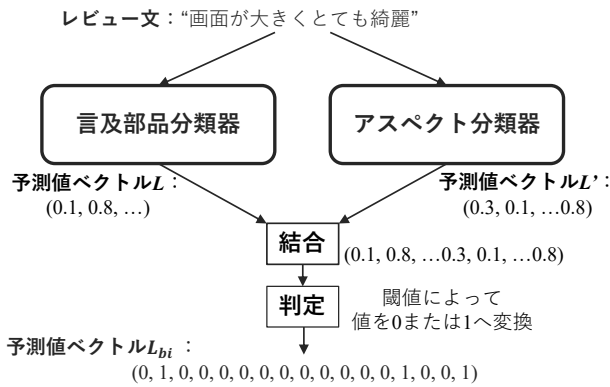


図 2 部品別意見収集手法のモデル

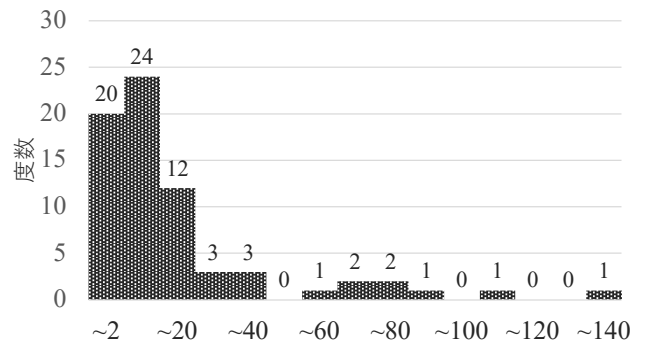


図 3 ラベル付きデータのヒストグラム

表 5 ラベル付きデータ内訳

	耐久性	機能性	使い心地	サイズ	重量	存在	見回目	合計
キーボード	2	9	109	85	7	54	10	173
ディスプレイ	8	40	33	131	26	7	63	169
フレーム	24	15	19	74	19	5	8	112
バッテリー	73	9	16	11	19	5	8	90
ファン	0	2	6	1	1	12	2	12
ポート	9	2	8	11	2	15	2	32
マウス	4	7	32	5	1	24	2	54
CPU	5	69	18	20	9	6	18	75
スピーカー	1	5	7	1	1	2	3	11
ストレージ	2	4	2	2	1	10	2	15
合計	87	93	196	243	69	113	87	

5. データ拡張手法

部品別意見収集手法で用いた類似レビュー文生成によるデータ拡張手法について説明する。

表 5 は第 3 節で説明した手法によって抽出した 2 種類のラベルをノート PC のレビュー文 500 件に対し人手で付与したデータの内訳である。本稿内で、表 5 中にある“キーボード”と“耐久性”などの言及部品ラベルとアスペクトラベルの組み合わせを項目と呼ぶ。表 5 中の各項目に属するレビュー文の数のヒストグラムを図 3 に示す。グラフ中の横軸がラベル付きデータ内のレビュー文の数、縦軸が該当するレビュー文の数となる項目の数である。図 3 の通り、属するレビュー文の数は 10 付近に偏った。

例えば、属するレビュー文の数が 5 の項目 k の場合を考える。この場合、先述のレビュー文 500 件を訓練データとしてマルチラベル分類器を作成すると、正のデータ 5 件、負のデータ 495 件によって項目 k の学習が行われる。このような訓練データで生成された分類器は、項目 k に属するレビュー文をレビューから分類する性能が低下する [5]。

本研究では、訓練データ内の偏りを解消するために類似レビュー文生成によるデータ拡張を施した。類似レビュー

文生成は WordNet による類似語生成を用いた。WordNet では入力された語に対して、その語の同義語を複数出力する。例えば、“効く”という語の場合、“作用”、“効果”、“利く”が出力される。このようにして出力される複数の同義語を“効く”の類似語として類似レビュー文生成に用いた。

類似レビュー文生成およびデータ拡張は、Algorithm 1 の手順で行う。 K は表 5 中の各項目に属するレビュー文が格納された項目データセット、 N は事前に設定した最低データ長であり、全ての項目のデータ長が N となるまでデータを拡張した。また、内容語の入れ替え処理時に用いた確率 p は、元となるレビュー文中の内容語の総数の逆数とした。

WordNet による出力によって、1 つの単語から複数の類似語が得られる。例えば、“画面が大きくとても綺麗”というレビュー文には、内容語として“画面”、“大きい”、“綺麗”の 3 つの語が含まれる。“大きい”、“綺麗”は複数の類似語をもつ。従って、これらの組み合わせをもとに元のレビュー文から多数の類似レビュー文が生成可能であり、類似レビュー文を訓練データに追加することで、ラベル付きデータ内の偏りを低減した。

6. 評価実験

部品別意見収集手法およびラベル抽出手法、データ拡張手法を評価するために行った 3 つの実験について説明する。ラベル抽出手法によって抽出されたラベルと EC サイト上の部品やアスペクトを意味する指標の一致割合、部品別意見収集手法の分類性能に対するデータ拡張の効果、部品名を含まないレビュー文に対する部品別意見収集手法の分類性能を評価するために 3 つの実験を行った。

評価実験では、先行研究で対象としたロードバイクのレビュー以外に、ノート PC のレビューを利用した。本研究の研究対象は、複数部品からなる製品である。自転車やパソコンは、Amazon 上に部品別の商品カテゴリー（以下、パーツカテゴリー）が存在するなど部品別の意見が求められていると思われるため、評価実験時の対象として選んだ。

Algorithm 1 データ拡張手法の擬似コード

```

 $\mathcal{K} \leftarrow \{K_1, K_2, \dots, K_n\}$  ▷ 項目データセット
 $\mathcal{K}' \leftarrow \{\{\}, \{\}, \dots, \{\}\}$  ▷  $\mathcal{K}$  と同じ次元の空集合の集合
 $|K_1|$  ▷  $K_1$  に属するデータの数を指す
function DATAUGMENTATION( $\mathcal{K}, N$ )
  for each  $K \in \{K \mid 2 < |K| < N, K \in \mathcal{K}\}$  do
    ▷ レビュー文の数が 2 より多く,  $N$  より少ない項目を選ぶ
    while  $(|K| + |K'|) < N$  do
      ▷  $K'$  は  $K'$  中の  $K$  に値する項目を指す
       $s \leftarrow \text{RANDOMSELECT}(K)$ 
      ▷  $K$  中のランダムなレビュー文を取得
       $s' \leftarrow \text{MAKESIMILARSSENTENCE}(s)$ 
       $K' \leftarrow K' \cup \{s'\}$ 
    end while
  end for
  return  $\mathcal{K} \cup \mathcal{K}'$ 
end function

function MAKESIMILARSSENTENCE( $s$ ) ▷  $s : [t_1, t_2, \dots, t_m]$ 
   $C \leftarrow \text{GETCONTENTWORDS}(s)$  ▷  $s$  中の内容語を取得
   $|C|$  ▷  $C$  中の語の数を指す
   $p \leftarrow 1/|C|$  ▷ 各内容語を入れ替える確率
  for  $i \leftarrow 1, 2, \dots, m$  do
     $t \leftarrow s[i]$ 
    if  $t$  is a content word then
      if  $p > \text{RANDOM}()$  then ▷  $\text{RANDOM}()$  は  $[0,1]$  の乱数
         $w \leftarrow \text{GETSIMILARWORDS}(t)$  ▷ 類似語を取得
         $s[i] \leftarrow \text{RANDOMSELECT}(w)$  ▷ ランダムな類似語と入替
      end if
    end if
  end for
  return  $s$  ▷ 類似レビュー文を返却
end function

```

6.1 ラベル抽出手法の評価

ラベル抽出手法では、製品種類毎に異なる部品とアスペクトを考慮するために、パターンマッチングを用いてレビューから言及部品ラベルおよびアスペクトラベルを抽出した。ラベル抽出手法を評価するために、抽出したラベルと EC サイト上のパーツカテゴリーまたは評価項目との一致割合を評価した。ロードバイクとノート PC のそれぞれの製品種類において、言及部品ラベルと Amazon 上のパーツカテゴリー、アスペクトラベルと価格.com^{*4} 上の評価項目の一致割合を評価した。

結果を図 4 に示す。グラフ中の縦軸が EC サイト上のパーツカテゴリーまたは評価項目が各ラベルと一致した割合、横軸は左から、ロードバイクの言及部品ラベル、ロードバイクのアスペクトラベル、ノート PC の言及部品ラベル、ノート PC のアスペクトラベルである。グラフ中の凡例のうち、文字列の一致は、抽出したラベルと EC サイト上のパーツカテゴリーが文字列として一致したものを示す。また、自転車のパーツカテゴリーの中には、タイヤとタイヤチューブ、チェーンとチェーンリングなど、部品の配置

^{*4} <https://kakaku.com>

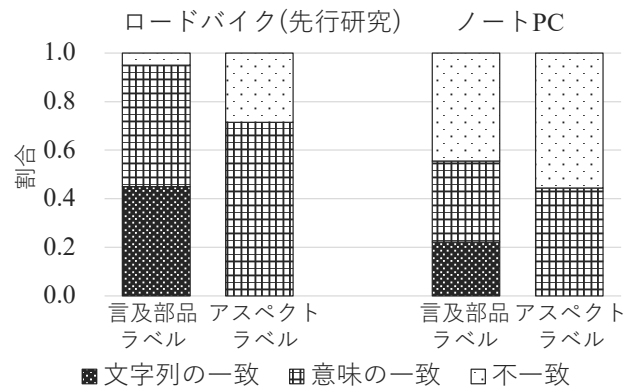


図 4 抽出したラベルと EC サイト上の指標の一致割合

場所や、機能の類似度などからレビュー中で分けて言及されない部品がある。そのような区別されない部品を cycle base asahi^{*5} 内の商品カテゴリー中の自転車部品を元に統合した結果、言及部品ラベルと一致したパーツカテゴリーを意味の一致とした。同様に、ノート PC でも、ディスプレイとグラフィックボード、ストレージと HDD や SSD などのパーツカテゴリーは意味の一致とした。また、アスペクトラベルでは、ラベルとして特定の文字列を抽出したわけではないため、文字列の一致はないとし、使い心地と使いやすさ、見た目とデザインなどの表記揺れや表現揺れを意味の一致とした。

先行研究にあるロードバイクの言及部品ラベルでは、パーツカテゴリー中の 45 % と文字列が一致した。また、意味の一致を含むと、パーツカテゴリー中のスタンド以外の全ての語と一致した。アスペクトラベルでは、評価項目中の 71 % と意味が一致した。不一致となった評価項目は、走行性能と使用パーツであった。

ノート PC の言及部品ラベルでは、パーツカテゴリー中の 22 % と文字列が一致した。また、意味の一致を含むと、パーツカテゴリー中の 55 % と一致した。不一致となったパーツカテゴリーは、メモリやインターフェースカード、ネットワークカード等であった。アスペクトラベルでは、評価項目中の 44 % と意味が一致した。不一致となった評価項目は、グラフィック性能、拡張性、バッテリー、画面、コスパであった。

6.2 データ拡張手法の評価

データ拡張手法では、訓練データ内の偏りを低減するために、WordNet による類似語をもとにデータを拡張した。データ拡張手法を評価するために、データ拡張時に設定する最低データ長 N と、拡張した訓練データによって生成される分類器の分類性能の関係を評価した。

評価実験に用いたデータセットとして、ロードバイクでは先行研究で用いたレビュー文 1,000 件のデータセット、

^{*5} <https://ec.cb-asahi.co.jp/category/cat1/自転車部品>

表 6 評価指標

予想ラベル \ 正解ラベル	True	False
True	TP	FP
False	FN	TN

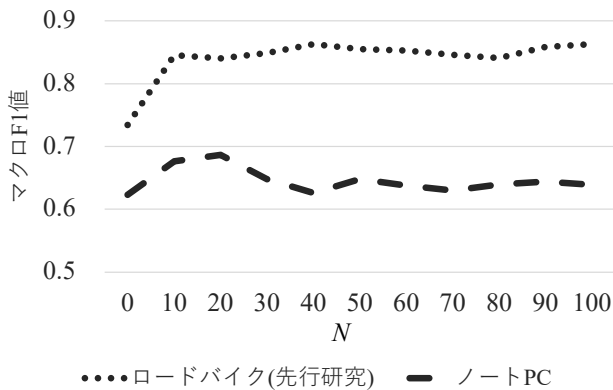


図 5 N とマクロ F1 値との関係

ノート PC では、500 件のレビュー文に人手でラベル付けし、偏りのないように選んだ 57 件を評価データ、ランダムに選んだ 50 件を検証データ、残りの 393 件を訓練データとした。

評価指標は、表 6 内の TP , FP , FN を用いて以下のよう
に、言及部品ラベルとアスペクトラベルを区別せずに算出したあるラベル x に対する適合率 (P) と再現率 (R), および調和平均である F1 値と、そのマクロ平均であるマクロ F1 値を用いた。

$$P = \frac{TP}{TP + FP} \quad R = \frac{TP}{TP + FN}$$

$N = 0, 10, \dots, 100$ とし、訓練データ内の各項目に対し属するレビュー文の数が N となるまでデータを拡張した。それぞれの拡張した訓練データによって 20epoch ずつファインチューニングを行い、最も損失の少ないモデルを N でのモデルとした。各モデルの出力となる予測値ベクトル L_{bi} は、閾値をもとに各値が 0 または 1 に変換されるベクトルである、従って、検証データに対して最もマクロ F1 値が高くなるように閾値を設定し、設定した閾値と評価データを用いて各 N でのモデルのマクロ F1 値を算出したものを図 5 に示す。グラフ中の縦軸がマクロ F1 値、横軸が最低データ長である N を示す。

ロードバイクでは、 $N \neq 0$ の全てで $N = 0$ の値を 0.1 以上上回った。しかし、ノート PC では、 $N = 10, 20$ の場合は 0.05 程度上回ったものの、 $N > 20$ では 0.02 程度の上昇となった。

それぞれの製品種類においてマクロ F1 値が最も高くなったロードバイクの $N = 40$ のモデルと、ノートパソコンの $N = 20$ のモデルのラベル毎の適合率および再現率、F1 値を図 6 および図 7 に示す。グラフ中の縦軸が適合率、再現率、F1 値、横軸が各ラベルを示す。ロードバイクで

の各ラベルの値に比べて、ノート PC での値は低い傾向となった。言及部品ラベルにおいては、ロードバイクは全ての F1 値が 0.8 を超えているものの、ノート PC においては、“マウス”ラベルを除く全てのラベルで F1 値が 0.8 を超えなかった。アスペクトラベルにおいて、ロードバイクとノート PC は、“耐久性”ラベルの F1 値が最も高い、再現率の著しく低いラベルが 1 つ存在するなどの共通点が見られた。また、ノート PC の言及部品ラベルにおいて、主に訓練データ中のレビュー文の数が少ないファンやスピーカー、ストレージの再現率が著しく低くなった。

6.3 部品名を含まないレビュー文の分類性能の評価

部品別意見収集手法は、レビュー中にある表現揺れや表記揺れによって、キーワード検索等による情報収集が困難であるため、BERT に基づく 2 種類の分類器を用いてレビューから意見を収集するものである。部品別意見収集手法を評価するために、部品名が文中に含まれないレビュー文の分類を行なった。2 つ目の評価実験の設定と同じデータセットによってファインチューニングおよび閾値設定を行い、評価データとして、文中に部品名を含まないレビュー文 30 件を用いて、マクロ F1 値を算出した。算出したものを表 8 に示す。

ロードバイクでは、全ての N で F1 値の大きな上昇は見られなかった。ノート PC では、 $N \neq 0$ の全てで $N = 0$ の F1 値を上回った。それぞれの製品種類において、各モデルのマクロ F1 値と N との間に比例関係等は見られなかった。

7. 考察と今後の展望

部品別意見収集手法内のラベル抽出手法、データ拡張手法、それらを実験した 3 つの実験についての考察と今後の課題について述べる。

7.1 評価実験結果に対する考察

1 つ目の評価実験において、ロードバイクの言及部品ラベルは、パーツカテゴリーの 95 % と意味が一致し、不一致となったパーツカテゴリーであるスタンドのレビュー中での出現頻度の小ささなどから、重要視されている部品を十分に言及部品ラベルとして定められたと考えられる。

ロードバイクのアスペクトラベルで不一致となった 2 つの評価項目は、“性能”についての項目である。“性能”は、“耐久性”、“操作性”、“使い心地”などのラベルの意味を包含しており、それらの意味を持つ語は、ラベル抽出時に細分化され、他のラベルとして定められたため、“性能”自体をアスペクトラベルとして定めることができなかった。しかし、評価項目の走行性能は製品全体に対する評価、使用パーツは各部品に対する総合評価であり、本研究における部品別意見収集のためのラベルとして適切とは言えない。

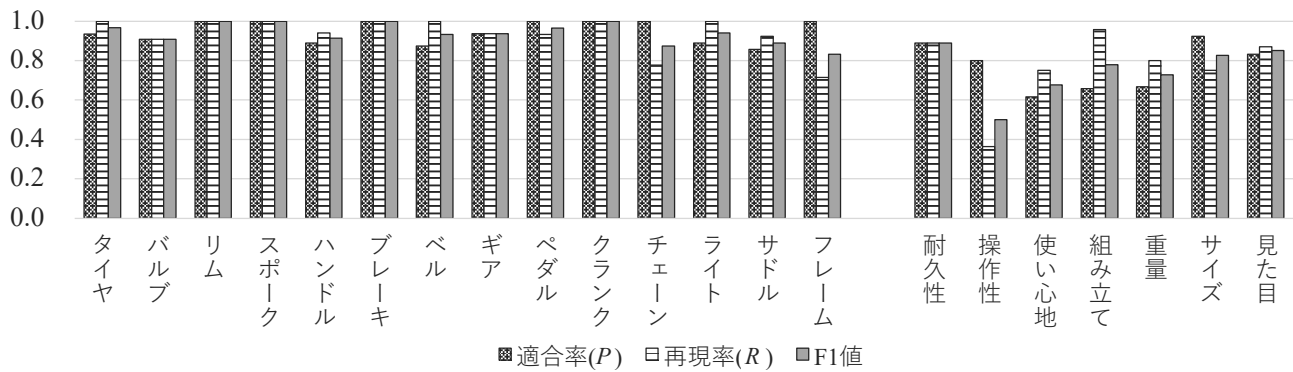


図 6 ロードバイクでの $N = 40$ 時のラベル毎の適合率と再現率と F1 値

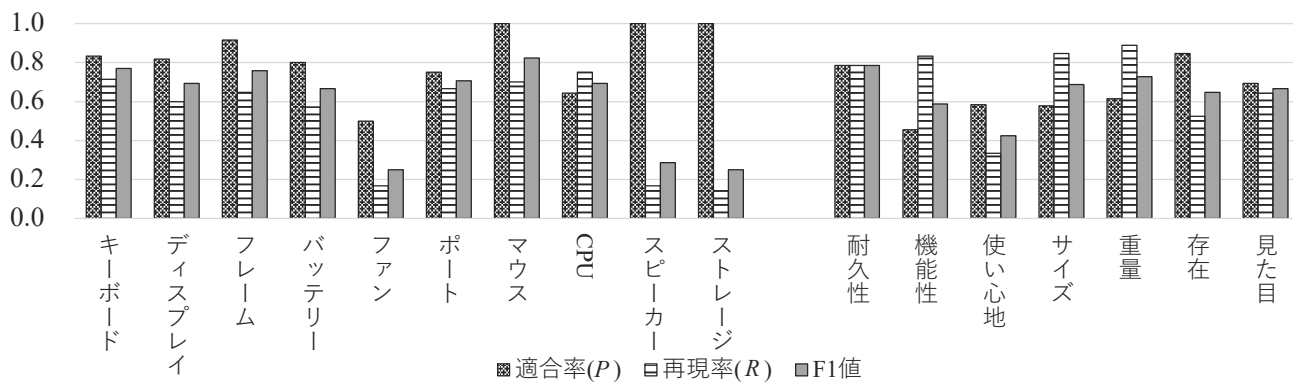


図 7 ノート PC での $N = 20$ 時のラベル毎の適合率と再現率と F1 値

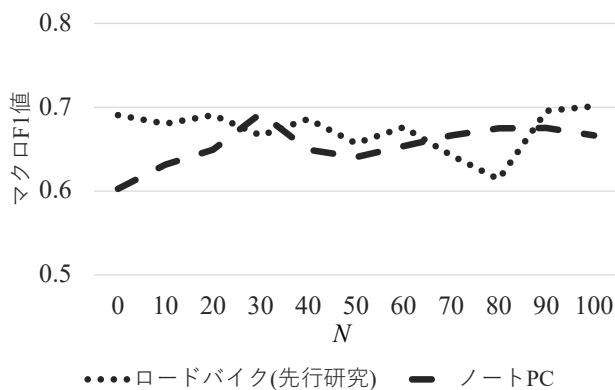


図 8 N とマクロ F1 値の関係 (部品名を含まない)

従って、ロードバイクでは、レビュー中で重要視されている部品およびアスペクトを、言及部品ラベルおよびアスペクトラベルとして定められた。

ノート PC の言及部品ラベルで不一致となったパーツカテゴリーには、メモリやマザーボードなどのパソコンには必須のパーツが含まれる。これらをラベルとして抽出できなかったことは、ノート PC と他の PC との間に、レビュー投稿者の着目点の違いがあることが考えられる。デスクトップ PC 等のレビューにおいては頻出するが、ノート PC では言及されにくいマザーボードなどのパーツカテゴリーを言及部品ラベル候補として抽出することができ

ず、ロードバイクの場合に比べて一致割合が低くなった。
ノート PC のアスペクトラベルで不一致となったグラフィック性能、画面、バッテリー、拡張性は、製品中の特定の部品自体または特定の部品の機能に対する評価項目である。よって、これらはロードバイクでの 2 つの評価項目と同様に、部品別意見収集のためのラベルとして適しているとは言えない。不一致となった評価項目のうち、コストパフォーマンスは、ロードバイクの場合と同様に“性能”に対する評価であり、アスペクトキーワードとして抽出できなかった。このように、ノート PC では、レビュー中で重要視されている部品およびアスペクトをラベルとして定められたとは言えない結果となった。

2 つ目の評価実験において、ノート PC でマクロ F1 値の上昇があまり見られなかったことの原因として、ロードバイクに比べてノート PC のレビューに、より専門性の高い複合語が多く含まれることが考えられる。例えば、“M1 チップ”や“外付け HDD”、“テンキー付きキーボード”などがある。これらの語によって文中の内容語の数が大きくなり、類似レビュー文生成時に元のレビュー文と類似度が高すぎる文や、複合語内の単語が入れ替わってしまった類似レビュー文が生成されることで、 $N > 20$ の範囲でデータ拡張手法によるマクロ F1 値の大きな上昇が見られなかったと考えられる。また、スピーカーやストレージなどのレ

ビュー文の少ないラベルに対する再現率が低いことなどから、ノート PC に対してデータ拡張手法は、偏った訓練データによる分類器への影響を低減できていない。

3 回目の評価実験において、ロードバイクにおけるマクロ F1 値と N の間に比例関係は見られなかったが、ノート PC においては N の増加とともにマクロ F1 値が上昇する傾向にあった。ロードバイクでは $N = 0$ のマクロ F1 値を下回ったモデルも多くあるが、ノート PC では全ての $N \neq 0$ が $N = 0$ を上回ったことから、部品名の含まないレビュー文への部品別意見収集手法およびデータ拡張手法の有効性は更なる検証が必要である。

7.2 ラベル抽出手法およびデータ拡張手法に対する考察

ラベル抽出手法では、評価実験の通り、抽象的なアスペクトをラベルとするために、ラベル間の包含関係を考慮した抽出が必要である。ロードバイクにおける“性能”と“耐久性”や“操作性”，ノート PC における“コスパ”と“機能性”や“使い心地”などの他のラベルとの包含関係を考慮したラベルによって、より抽象的な“良い”などの意見を部品別意見として収集することができる。また、言及部品ラベルとアスペクトラベルの因果関係も考慮が必要である。現在はレビュー文に対し、それぞれのラベルを複数ずつ付与しているため、アスペクトの要因となった部品が特定できない。例えば、“画面が大きく、充電も持つ”というレビュー文を部品別意見として収集する場合に、言及部品ラベルとして“ディスプレイ”と“バッテリー”が、アスペクトラベルとして“サイズ”と“耐久性”が付与される。しかし、どちらの言及部品がどちらのアスペクトで述べられているかを取得することができない。このような言及部品ラベルとアスペクトラベルとの因果関係、またはそれぞれのラベル内の包含関係を考慮するためには、ラベル付与時の閾値設定方法を改良する必要がある。現在では全てのラベルに対して一様に閾値を定めているが、ラベル間の包含関係や因果関係を考慮した閾値設定によって、より複雑なラベル付与を実現できる可能性がある。また、今回のパターンマッチングを用いた手法ではノート PC のメモリなどの重要視されると考えられる部品をラベルとして抽出できなかった。パターンマッチングによる抽出では、複合語や固有名詞によって頻繁に表現される部品を言及部品ラベル候補として抽出できなかった。形態素解析時の複合語の考慮や、頻出語の辞書作成等を用いて、より複雑なパターンを抽出することで改善が見込める。また、言及部品ラベル候補およびアスペクトキーワード抽出時に、主語と述語の関係を取得可能な構文解析器を用いるなどの方法による改善が見込める。

データ拡張手法では、類似度を考慮した類似レビュー文の生成および追加順の調整が必要である。現在は、ランダムなレビュー文内の、ランダムな内容語に対して、ランダム

な類似語と入れ替え類似レビュー文としている。しかし、今回のノート PC のように、内容語を抽出する際、名詞・動詞・形容詞だけでなく、複合語や製品種類毎の特徴表現を考慮することで、不適切な語の抽出および入れ替えを避けることが期待できる。また、類似レビュー文生成の元となるレビュー文選択時に、該当する項目中の“最も該当項目らしい”文をベクトル表現等によって選択することや、類似レビュー文の訓練データへの追加時に、文同士の類似度等を考慮した順で追加することでデータ拡張手法の汎用性の向上が見込める。

7.3 今後の課題

今後の課題として、部品別意見収集手法の汎用性の向上のために、ラベル抽出手法において、ラベル候補抽出時のパターンの改良、ラベル間の包含関係や因果関係の考慮、データ拡張手法において、類似語入れ替え時および類似レビュー文追加時の類似度による優先順の導入などがある。

8. おわりに

本稿では、BERT に基づくマルチラベル分類器を用いたレビューからの部品別意見収集手法について説明した。また、製品種類毎に異なる部品とアスペクトを考慮するために、パターンマッチングを用いてレビューからラベルを抽出する手法、訓練データ内のレビュー文に対し、WordNet を用いた類似語入れ替えによるデータ拡張手法について説明した。評価実験において、ラベル抽出手法、データ拡張手法、部品別意見収集手法について評価した。実験結果から、マクロ F1 値はロードバイクに関しては 0.86 まで上昇したが、ノート PC に関しては 0.69 であった。今後の課題として、ラベル抽出手法内での複合語の考慮や、データ拡張手法内での類似レビュー文追加順の変更などがある。

謝辞 本研究の一部は JSPS 科研費 JP19K12266 の助成を受けたものです。

参考文献

- [1] 按田将吾, 菊地真人, 大園忠親: AR に基づくレビュー閲覧支援システムのための製品の各部品に関する意見抽出機構, 信学技報, Vol. 121, No. 298, AI-2021-8, pp. 40-45 (2021).
- [2] 新里圭司, 小山田由紀: 店舗レビューには何が書かれているか?—調査及び自動分類システムの開発—, 自然言語処理, Vol. 25, No. 1, pp. 57-79 (2018).
- [3] 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).
- [4] Vosoughi, S., Vijayaraghavan, P. and Roy, D.: Tweet2vec: Learning tweet embeddings using character-level cnn-lstm encoder-decoder, *SIGIR 2016*, pp. 1041-1044 (2016).
- [5] López, V., Fernández, A., García, S., Palade, V. and Herrera, F.: An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics, *Information Sciences*, Vol. 250, pp. 113-141 (2013).