

# 商品属性とレビューの対応関係学習に基づく言語モデル

河田友香<sup>1</sup> 山本岳洋<sup>2</sup> 大島裕明<sup>3</sup>  
藤田澄男<sup>4</sup>

本研究では、「画素数: 3,240万, 撮像素子: フルサイズ, 重量: 800g」のような商品属性と、「遊ぶ子どもの表情もくっきりと撮影できます」のような, 商品进行评估するテキストとの関連性を予測する言語モデルを提案する. モデルの学習では, 事前学習済みのBERTに, 商品属性とレビューを入力し, それらが実際に対応するものであるかどうかを予測する. モデルの学習における技術的課題としては, 属性値に含まれる量的データの扱いと, テキストと商品属性の対応の学習方法があげられる. これらの技術的課題を解決するために, 量的データのカテゴリ化と, 追加事前学習を行った. カメラの商品属性とレビューを用いて実験を行った結果, 追加事前学習により関連度を予測する性能が向上することが分かった.

## 1 はじめに

商品を購入する際に参考にする情報の一つとして, 商品属性が挙げられる. 商品属性とは, 商品の様々な属性名とそれに対応する属性値の集合である. 例えば, カメラであれば「撮像素子」という属性名に対して「フルサイズ」, 「APS-C」といった属性値があり, 「画素数」という属性名に対して, 「3,410万画素」, 「2,040万画素」といった属性値がある.

商品属性について詳しいユーザは, 「画素数が3,410万画素あれば風景写真も美しく撮影できる」や「フルサイズであれば色鮮やかに撮影できる」のように商品属性がどのような特徴に繋がるのかを理解している. そのため, 商品属性をもとに商品を探すことが可能である.

一方, あまり商品属性に詳しくないユーザには, 商品属性と関連づけて商品を探すことは難しい. 例えば, 画素数が風景写真を色鮮やかに撮影するために重要な属性なのか, 素早く動くスポーツ選手を撮影するために重要なのか理解できない. このような情報を, ユーザはレビューや商品説明文内から探すだろう. しかし, レビューに「色鮮やかに撮影できる」などと書かれていない場合にも, 商品属性に着目すると「色鮮やかに撮影できる」カメラであることがわかるような場合もあると考えられる.

このような問題を解決するために, 本研究ではレビューと商品

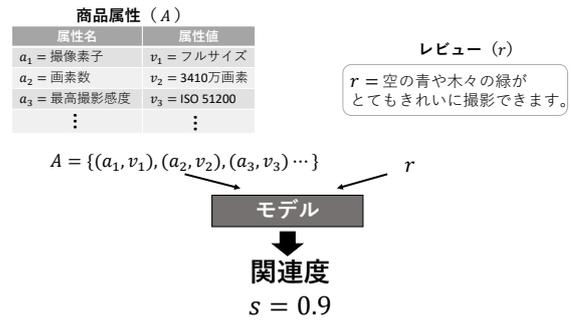


図1 レビューと商品属性を入力として, その関連度を求める手法の概要図.

属性の対応関係を学習し, これらの関連を理解した言語モデルの構築に取り組む. 提案手法の概要を図1に示す. このようなモデルを構築することができれば, ある商品属性に対して適切に対応するレビューがどのようなものか知ることができたり, あるレビューがどのような商品属性に対して適切かを知ることができる. これは, 多様な商品の探し方を可能にするということにつながる.

本研究では, レビューと商品属性の関連を理解したモデルをBERTを用いて構築する. 構築されたBERTにレビューと商品属性を入力すると, レビューと商品属性の関連度を予測することができる. 学習のために, ウェブ上の商品比較サイトに掲載されている商品属性とレビューから訓練データを構築する. 具体的には, ある商品の商品属性とそれに対して実際につけられたレビューには関連度があり, ある商品の商品属性とそれとは異なる商品のレビューには関連度が無いとするデータを用いて学習を行う.

商品属性とレビューの対応関係を学習したモデルを構築する際には, 追加事前学習によって精度が向上することが期待される. BERTにおける追加事前学習の有用性は, 様々な研究で示されており, 本研究のような商品属性とレビューの対応のような複雑な入力の場合にも有用に働くことを明らかにすることを目指す.

商品属性をBERTに入力する方法には, 様々な方法が考えられる. 特に, 量的データについて, 学習によってBERTにその大小関係を理解させることが必要である. 我々は, 量的データのカテゴリ化を行うことで, 商品属性とレビューの対応関係を予測する精度を向上させることを目指す. たとえば, 先行研究において, 商品属性とレビューの対応を学習したモデルを構築するものが存在するが, そこで対象としている商品属性は主としてカテゴリカルデータである. そのようなモデルに量的データを単に入力すると, 値の大きさを上手く学習することができないと考えられる. 本研究では, 量的データを含む属性値にカテゴリ化を行うことで, 商品属性とレビューの対応を予測する精度が向上するのかを明らかにする.

## 2 関連研究

### 2.1 商品属性やレビューを用いたテキスト分析

商品のレビュー, 商品の属性, ユーザ情報についてのテキス

<sup>1</sup> 非会員 兵庫県立大学大学院情報科学研究科  
berry04230317@outlook.jp

<sup>2</sup> 正会員 兵庫県立大学大学院情報科学研究科  
t.yamamoto@sis.u-hyogo.ac.jp

<sup>3</sup> 正会員 兵庫県立大学大学院情報科学研究科  
ohshima@ai.u-hyogo.ac.jp

<sup>4</sup> 非会員 LINEヤフー株式会社  
sufujita@lycorp.co.jp

ト分析については様々な研究が行われている。McAuleyらは、Yes/Noで判断できる質問に対して、レビューから回答を見つけ、答えを提示するモデルを提案している [12]。Royらは、商品の素材についてなどの商品属性から回答できる質問に対して、商品属性から自動的に回答を生成するモデルをXLNetやBERTをもとに構築している [15]。Zhangらは、ユーザが行った商品に関する質問から、その質問に対する答えを予測し、答えに関するレビューを提示している [19]。Chenらは、レビュー文から質問に関するスニペットを生成し、レビュー文に基づいて質問回答を生成するモデルRAGEを構築している [2]。

レビューと属性の両方についてテキスト分析を行う研究もある。Fengらは、商品の属性とレビューを関連づけるモデルを構築し、属性とレビューの関係を理解した上で回答を生成する手法を提案している [5]。Gaoらは、商品属性と属性値、過去にユーザによってつけられたレビューを用いて、質問に関する属性に言及したレビューから自動的に回答を生成している [7] [8]。Zhangらは、属性とBERTでベクトル化したテキストを結合するMA-Transformer Encodersを用いて、商品属性とテキストをベクトル化するMA-BERTを提案している [21]。

これらの研究と本研究は、商品属性とレビューの対応を学習するという点で類似している。これらの研究では、主にカテゴリカルデータを対象とした商品属性を扱っているのに対し、本研究では特に量的データを多く含む商品属性を扱う。量的データを多く含む商品属性をモデルに入力する際には、量的データに適切な処理を行う必要がある。本研究では、量的データのカテゴリ化を行うことで、商品属性とレビューの対応を学習することを目指す。

## 2.2 テキストからの属性抽出

テキストから属性を抽出する研究も数多く行われている。Iwanariらは、ブログ内の文章から、対象物とその対象物の評価を抽出し、属性に応じて対象物を順位づけた [9]。Fuchsらは、商品説明文と商品の属性名、属性値を与え、商品をあらわすために重要な属性値を抽出した [6]。内田らは、レビューでの相対的な評価や単体での評価から、属性ごとの順序関係を推定し、信頼性の評価を行った [23]。Zhaoらは、BERTの固有表現抽出を用いて、商品説明文やタイトルから商品属性を抽出した [22]。Wangらは、属性を問いとして与え、属性値を商品説明文から抽出するモデルをBERTをもとに構築した [18]。近年では大規模言語モデルを用いた属性抽出なども提案されている [4] 本研究では、既存の商品比較サイトに記載されている商品属性をモデルに対する入力に用いるが、こうした研究を用いて自動的に抽出された商品属性を用いることも可能だと考えられる。

## 2.3 言語モデルでの数値の処理

言語モデルで数値を扱う研究も取り組まれている。Naikらは、数値の大小関係や数え上げを単語の分散表現がどの程度捉えているか分析するフレームワークを提案した [13]。Chenらは、ニューラルネットワークモデルに数値を学習させることで、テキスト内に誇張して記述されている数値表現を検出できるか分析した [1]。Talmorらは数字の比較や質問応答などの推論タスクにおいて、異なる言語モデルでの推論能力を比較した [17]。Linらは、BERTなどの事前学習済みモデルでは、数値についての知識をど

の程度理解することができているのかを調査した [11]。Jinらは、数値を考慮した生成的言語モデルを構築するため、仮数と指数を表す埋め込み表現をGPTに組み合わせたNumGPTを提案した [10]。本研究では、商品属性に含まれる数値をカテゴリ化することで、商品属性とレビューの対応関係の学習を目指す。

## 3 問題定義

本節では、本研究で取り組む問題について定義する。

ある商品 $p$ の商品属性 $A$ を(1)のように表す。

$$A = \{(a_1, v_1), (a_2, v_2), \dots, (a_k, v_k)\} \quad (1)$$

ここで $k$ は属性の個数とする。また、 $a_i (1 \leq i \leq k)$ は属性名を表し、 $v_i (1 \leq i \leq k)$ は属性値を表す。カメラを例とすると、商品属性 $A$ は $A = \{(\text{画素数}, 3,410\text{万画素}), (\text{撮像素子}, \text{フルサイズ}), \dots, (\text{重量}, 800\text{g})\}$ のように表すことができる。商品 $p$ に対するレビューを $r$ と置く。レビュー $r$ は、単語の系列である。 $i$ 番目の単語を $w_i$ と置くと、(2)のように表すことができる。

$$r = w_1 w_2 \dots w_L \quad (2)$$

ここで $L$ はレビュー $r$ 中の単語の個数である。商品属性 $A$ とレビュー $r$ を入力として受け取り、商品属性 $A$ に対して、レビュー $r$ が書かれる可能性のある文章なのかを予測し、関連度 $s$ で表す。関連度 $s$ とは、商品属性 $A$ とレビュー $r$ がどの程度関連しているかの度合いを表す。関連度 $s$ は、レビュー $r$ と商品属性 $A$ に関連がなければ値が0に近くなり、レビュー $r$ と商品属性 $A$ に関連があれば値が1に近くなる。

具体的には、 $A = \{(\text{画素数}, 3,410\text{万画素}), (\text{撮像素子}, \text{フルサイズ}), \dots, (\text{重量}, 800\text{g})\}$ のような商品属性が与えられたとする。このカメラは撮像素子がフルサイズであり、画素数が3,410万画素であるため、美しく撮影することができると考えられる。このようなカメラの商品属性に対して、「空の青や木々の緑がとてもきれいに撮影できます。」のようなレビューは商品に対する評価としてあり得るテキストであるため、関連度 $s$ は1に近くなる。反対に「発色はきれいだが、画質は良いといえない。」のようなレビューは、その商品に対する評価とは考えにくいいため、関連度 $s$ が0に近くなる。このようなモデルを構築することで、商品属性 $A$ とレビュー $r$ が関連しているかを理解することができる。

## 4 提案手法

本研究では商品属性とレビューを入力し、それらの関連度を予測するBERTモデルの構築を行う。商品属性は、属性名と属性値の対応で表される。属性値には量的データを含むものも多く扱うこととする。そのような商品属性とテキストであるレビューの対応を学習する方法について述べる。

### 4.1 モデルの概要

本研究で提案するモデルの概要について述べる。提案モデルの概要を図2に表す。提案モデルでは入力として、レビュー、商品属性を受け取り、レビューと商品属性の関連度を算出する。

ある商品の商品属性とそれに対して実際につけられたレビューが入力された際には関連度が1となることが期待され、ある商品

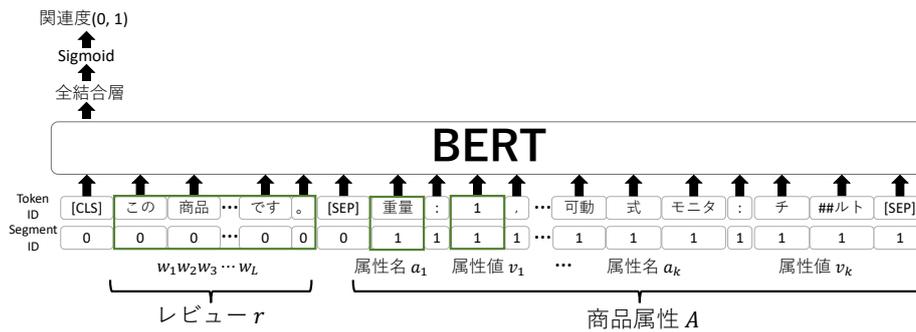


図2 提案モデル。レビューと属性名、属性値を入力として、入力した商品属性がレビューと対応しているかを判断する。提案手法は実際に入力した商品属性を持つ商品に付けられたレビューを正例、異なる商品属性を持つ商品に付けられたレビューを負例として学習する。

の商品属性とそれとは異なる商品のレビューが入力された際には関連度が0となることが期待される。そのようなデータを作成して、モデルの学習を行うこととなる。

## 4.2 BERT

本研究では、様々な自然言語処理タスクで使用されるBERT [3]を基に言語モデルの構築を行う。BERTはTransformer Encoderを用いたニューラルネットワークのモデルである。大量のテキストでファインチューニングを行うことで様々な自然言語処理タスクを精度よく解けることが知られている。

BERTでは入力されたテキストを分割し、トークンとして入力する。商品属性は、属性名と属性値を「画素数:3410,重量:800...」のように「:」と「,」を用いて結合し、テキストとしてBERTに入力することとした。ただし、属性値の表現方法については、4.3節にて説明する複数の方法を試す。レビューと商品属性の間には、レビューと商品属性を区別するために境界となるトークンとして [SEP] トークンを加える。

また、Segment IDを使用し、レビューに0、商品属性に1を割り当てた。Segment IDは、通常、BERTにテキストを2文入力する際に、入力するトークンが1文目のトークンなのか2文目のトークンなのかを判別するために使用されるデータ列のことである。本研究では、そのSegment IDをレビューと商品属性を判別させるために使用した。

## 4.3 量的データの処理

本節では、量的データをどのように処理するかについて述べる。属性値には、「3,410万画素」や「800g」のように量的データを含むものが多く存在する。これらの値をそのままのテキストで言語モデルに入力しても、数値の大小関係や規模感を理解することが難しいとされる [20]。商品属性に含まれる数値情報は商品を表す重要な情報であるため、適切に理解させることが必要である。そこで、属性値の量的データのカテゴリ化を行うことで、量的データの大小関係を理解させることを検討した。量的データのカテゴリ化とは、画素数であれば、例えば、4,000万画素のカメラは「画素数が非常に高い」や2,000万画素のカメラは「画素数はあまり高くない」のように、閾値を定め、その閾値に基づいて量的データをカテゴリに分割することを指す。カテゴリ化した量

的データを含む属性値の表現としては、以下の3つの手法を試すこととした。

- 属性ごとにカテゴリを表すトークンで表現
- カテゴリを数値で表現
- 属性間で共通のカテゴリを表すトークンで表現

表1では、量的データをそのまま入力する方法と、これら3つの手法について例示している。詳細について以下で説明する。

### 4.3.1 属性ごとにカテゴリを表すトークンで表現

これは、画素数や重量など、属性ごとにカテゴリの数、閾値を定め、それぞれに独自のトークンを用意する方法である。例えば、5段階のカテゴリで画素数と重量を表現することを考える。ある商品の画素数が5段階中、最も高いカテゴリ「5」だったとする。このとき、画素数の属性値を表す[*gaso\_5*]のようなトークンを用いる。一方、重量は5段階中、最も低いカテゴリ「1」だったとする。この重量のためには、[*weight\_1*]というトークンを用いる。このように、属性ごとにカテゴリを表現するトークンを準備し、それによって属性値を表現する方法である。属性ごとに値の大きさを表すトークンを用意することで、それぞれの属性における属性値の大小関係や意味を表現することができると期待される。

### 4.3.2 カテゴリを数値で表現

これは、数値を用いてカテゴリを表現する方法である。先ほどの例のように、ある商品の画素数がカテゴリ「5」だとすると、この時、画素数の属性値を「5」という数字で表現する。重量がカテゴリ「1」だとすると、その属性値を「1」という数字で表現する。この手法では、新たに追加するトークンがないため、事前学習済みモデルが新たなトークンについて学習する必要がないことが利点となる。1や5といった小さな数値については、事前学習におけるコーパスで頻出していることが期待され、大小関係についても学習できている可能性があるため、この手法が機能することが期待される。

### 4.3.3 属性間で共通のカテゴリを表すトークンで表現

これは、複数の属性で共通するカテゴリごとの大きさを表すトークンを追加し、属性値を表現する方法である。例えば、あ

表1 量的データを含む属性値の表現方法.

属性値の表現方法	入力形式
量的データをそのまま入力	画素数:3410,重量:800・・・
属性ごとにカテゴリを表すトークンで表現	画素数:[gaso_1],重量:[weight_3]・・・
カテゴリを数値で表現	画素数:1,重量:3・・・
属性間で共通のカテゴリを表すトークンで表現	画素数:[level_1],重量:[level_3]・・・

る商品の画素数がカテゴリ「3」だとすると、画素数の属性値を[level\_3]というトークンで表現する。重量がカテゴリ「3」だとすると、重量の属性値も[level\_3]というトークンで表現する。このように、異なる属性においても、共通のトークンで属性値を表現する方法である。このような属性値のカテゴリを表す数値トークンを導入することで、異なる属性間で共通のトークンを使って属性値を表現できる。新たに追加するトークンの数を抑えることができるという利点がある。

#### 4.4 追加事前学習

本研究では、BERTの事前学習済みモデルを用いて、それに追加事前学習を行う。追加事前学習は継続学習とも呼ばれ、タスクの性能が向上することが期待される [14] [16]。追加事前学習の概要を図3に示す。

本研究では、追加事前学習としてMasked Language Modelによる学習を行う。具体的には、入力において、レビューの文章中のトークンや、属性値のトークンに対してマスク処理を行い、その処理したトークンを予測するという学習を行った。属性名は入力において共通して使われるためマスク処理の対象としなかった。マスク処理では、まず、入力トークンの15%を選択して、そのうちの80%のトークンは[MASK]トークンに置き換え、10%のトークンは入力トークンとは異なるランダムなトークンに置き換え、10%のトークンでは元のトークンのままとするという方法を用いた。[MASK]トークンに部分について、元のトークンを予測することで学習を行う。このような追加事前学習により、トークン同士の関連が学習されることが期待される。

#### 4.5 正例と負例の構築

本節では、正例と負例をどのように構築するかについて述べる。モデルは、レビューに対して、入力された商品属性が関連しているかを当てるように学習を行う。この学習を行う上で、正例を入力された商品属性を持つ商品に実際につけられたレビュー、負例を異なる商品属性を持つ商品につけられたレビューと定義し、それらを当てるように学習する。

負例として用いる、異なる商品属性を持つ商品の選び方について述べる。異なる商品名を持つすべての商品からランダムに商品を選択し負例として用いると、後継商品やキットとして付いているレンズの違いなど、商品属性が類似している商品も多くあり、このような商品属性を負例として用いるのは望ましくない。そこで、商品属性のコサイン類似度を求め、そのコサイン類似度が一定の値以下の商品を負例として用いる。

コサイン類似度の求め方について説明する。属性値は、カテゴリカルデータと量的データの2種類の値がある。カテゴリカルデ

ータについては、ダミー変数に変換、量的データについては、正規化を行う。ダミー変数や正規化を行い置き換えた属性値を用いてコサイン類似度を求め、商品を選択する。本研究では負例となる商品は、コサイン類似度が0.95以下の商品をランダムで選択する。

#### 4.6 モデルの学習

提案モデルでは、入力された商品属性に対して、入力されたレビューがその商品属性を持つ商品に対するレビューかどうかを予測する。

正解ラベルとの誤差を計算し、損失値を求め、逆伝播を行うことで学習する。損失関数として、バイナリ交差エントロピーを用いた。今、 $o_i$ を訓練データ内の*i*番目の商品に対して、正例であるとモデルが予測した確率とする。また、 $y_i$ を*i*番目の商品の正解ラベルとする。ここで、商品の正解ラベルとは、入力された商品属性を持つ商品に付けられたレビューであるとき $y_i$ は1、そうでなければ $y_i$ は0をとる。訓練データの個数を $N$ 、*i*番目のデータに対して正例であると予測した確率を $o_i$ 、*i*番目のデータの正解ラベルが $y_i$ であるとき、以下のバイナリ交差エントロピー $L$ を最小化するように学習を行う。

$$L = - \sum_{i=1}^N (y_i \log o_i + (1 - y_i) \log(1 - o_i)) \quad (3)$$

#### 4.7 関連度の計算

入力された商品属性とレビューが関連している度合いを関連度と呼ぶこととする。関連度は、モデルに商品属性 $A$ とレビュー $r$ を入力することで求められる $(0, 1)$ の値である。関連度は、シグモイド関数を用いることで求める。ある商品の商品属性 $A$ とレビュー $r$ を用いて、関連度 $s$ を求める。関連度 $s$ は、分類ラベルを予測するために用いられる値であり、モデルからの出力値にシグモイド関数をかけたものであるため、 $(0, 1)$ の間で表される。

## 5 実験

### 5.1 データ

本研究では、デジタル一眼レフカメラを対象にし、価格.com<sup>\*1</sup>のレビューと商品スペック表を用いて実験を行う。デジタル一眼レフカメラは、様々な商品属性を持ち、スペック表を見ただけではどのような商品かを理解することが難しい場合がある。また、レビューで述べられていることがどの商品属性に関連するものかを判断することも難しい場合がある。そこで、本研究

\*1 <https://kakaku.com/>

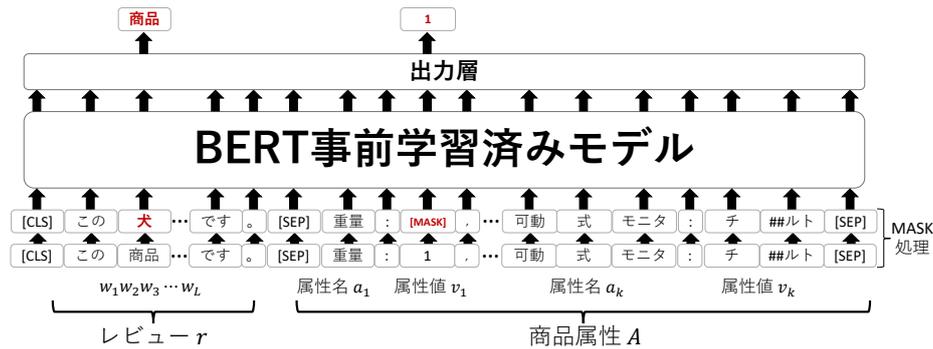


図3 追加事前学習の概要. レビューと属性値に対して, マスク処理を行い学習する.

表2 使用するカメラの商品属性とその形式, 属性値の例. 属性値の例として, “ $\alpha$ 7R III ILCE-7RM3A ボディ”の属性値を示す. カテゴリ数の列には, 各属性が何種類のカテゴリに分割されているかを示す.

属性名	形式	属性値の例	カテゴリ数
メーカー	文字列	SONY	-
タイプ	文字列	ミラーレス	-
画素数	数値	4240万	7
撮像素子	文字列	フルサイズ	-
連写撮影	数値	10	6
最低撮影感度	数値	100	4
最高撮影感度	数値	32000	5
最低シャッタースピード	数値	1/8000	4
最高シャッタースピード	数値	30	4
撮影枚数	数値	640	7
防塵・防滴	文字列	あり	-
幅	数値	126.9	8
高さ	数値	95.6	8
奥行	数値	73.7	8
可動式モニタ	文字列	チルト	-
重量	数値	572.0	8
カラー	文字列	ブラック系	-

ではデジタル一眼レフカメラを対象とする.

価格.com では, デジタル一眼レフカメラとして, 一眼レフ, ミラーレス一眼レフが含まれている. それらを合わせた対象商品数は458商品, また, 総レビュー数は23,246件であった.

### 5.1.1 商品属性

本研究で対象としたのは17種類の属性である. 提案手法は4.3節で述べたように, 属性値の形式が量的データである場合, カテゴリ化を行いモデルに入力する. カテゴリ数については, スタージェスの公式を使用して決定した. 実験に使用した商品と, その属性値の形式, 属性値の例, カテゴリ数について表2に示す. 属性値の例として, “ $\alpha$ 7R III ILCE-7RM3A ボディ”の商品属性を示す.

### 5.1.2 レビュー

レビューには, 「配送もスムーズで満足です。」や「セールだっ

たので購入しました。」のような商品属性とは関係のないテキストも存在する. これは商品属性に関するテキストではないので, 学習データとして用いるのは望ましくない.

そこで, レビュー中に評価項目を明記して書かれたレビューをデータとして使用した. 価格.comでは, 評価項目を明記して書かれたレビューが存在する. 例えばカメラであれば, 「デザイン」, 「画質」, 「操作性」, 「バッテリー」, 「携帯性」, 「機能性」, 「液晶」, 「ホールド感」という8つの評価項目で商品进行评估して, レビューが書かれることが通例となっている. 画質であれば, 撮影感度や画素数, 撮像素子などの属性が, 携帯性であれば, 幅や高さ, 重量などの属性が関連していると考えられる. 評価項目に基づくレビューは属性に関連するレビューであると考えられるため, 本研究ではこのようなレビューを学習や評価に使用する. 1つの評価項目について複数文にわたって記述されているレビューは, 1文ずつ切り出して用いる.

短すぎる, あるいは長すぎるレビューを除去するため, 本研究では, 全レビューの中で文書長が全体の25%から75%であるレビューを用いる. この範囲に含まれるレビューの文字数は19文字以上47文字以下となった.

### 5.2 実験設定

BERTの事前学習済みモデルは東北大学乾研究室の事前学習済みモデルを使用する<sup>\*2</sup>. 学習率は追加事前学習では $2.0 \times 10^{-5}$ , ファインチューニングでは $2.0 \times 10^{-6}$ とした. 訓練データを学習するエポック数については, 追加事前学習, ファインチューニングともに検証データに対して Patienceを10に設定し, 早期終了に基づき決定した.

入力するトークン数は, 256トークンと設定した. 訓練データと検証データには, 408商品のレビューを対象とし, 9:1にデータ分割を行った. その結果, 訓練データ86,514件, 検証データ9,613件となった. また, テストデータには, それらとは異なる50商品のレビューを対象とし, 11,067件のレビューを用いることとした. すなわち, 訓練データ, 検証データにはテストデータと同じ商品は含まれておらず, また, 商品属性についても同一のものがないようになっていた.

<sup>\*2</sup> <https://github.com/cl-tohoku/bert-base-japanese-v3>

ここで、1件のデータを具体的に説明する。データ1件はレビューと商品属性の組で表すことができる。例えば、価格.comに掲載されている、“α 7R III ILCE-7RM3A ボディ”を用いて、データの構成を説明する。“α 7R III ILCE-7RM3A ボディ”の商品属性は、表2に示す属性名と属性値の例のペアで表される。1つの商品は表2で示した商品属性を、必ず1つ持つ。“α 7R III ILCE-7RM3A ボディ”に対して、「軽くコンパクトなため、持ち運びに最適である。」といったレビューが記述されていたとする。このときデータ1件は、「軽くコンパクトなため、持ち運びに最適である。」といったレビューのテキストと、表2で示した商品属性の組である。

追加事前学習では、86,514件の訓練データと9,613件の検証データのレビューに対して、実際にそのレビューが記述された商品の商品属性を入力に用いた。ファインチューニングでは、同じ86,514件の訓練データと9,613件の検証データのレビューを正例と負例に分割する問題で学習を行った。訓練データとして使用する86,514件のレビューのうち、正例は43,271件、負例は43,243件だった。検証データとして使用する9,613件のレビューのうち、正例は4,793件、負例は4,820件だった。負例は、4.5節で述べた作成方法を用いて作成した。

実験では、4.3節で述べた3つの手法と、量的データをそのまま入力する手法を用いた。これらの手法における精度を検証することで、量的データの大小関係をBERTで理解させるための手法を明らかにする。

### 5.3 評価方法

本研究では3つの評価方法で、提案手法の有効性を評価する。

#### 5.3.1 関連度予測タスクによる評価

関連度を関連あり、関連なしの二値とみなし、それらを正しく予測できるかどうかを評価する。ここでは、関連あり、関連なしとは実際にその商品属性を持つ商品に与えられたレビューが付いたかどうかである。関連ありは入力した商品属性を持つ商品に付けられたレビュー、関連なしは異なる商品属性を持つ商品に付けられたレビューと定義する。

評価には、5.2節で述べたテストデータ50商品11,067件のレビューを用いる。11,067件のレビューのうち、5,534件のレビューを関連あり、5,533件のレビューを関連なしの判定として用いる。関連ありの判定として用いる5,534件のレビューは、レビューが記述された商品の商品属性とともに入力する。関連なしの判定として用いる5,533件のレビューは、レビューが記述された商品とは異なる商品の商品属性とともに入力する。関連なしとして使用する商品の選択は、テストデータとして用いる50商品の中から、4.5節で述べた負例の作成方法と同様の方法で行う。関連ありと関連なしを正しく当てられているかを正解率で評価する。このとき、モデルの関連度の出力が0.5以上であれば関連ありと予測したとみなし、そうでなければ関連なしと予測したとみなす。

#### 5.3.2 レビューランキングタスクによる評価

テストデータに含まれるレビューを商品属性との関連度が高い順にランキングを行い、実際に入力した商品属性を持つ商品に付けられたレビューが上位にランキングされているかを評価する。

ここでは、5.2節で述べたテストデータに含まれる50商品の商

品属性と、11,067件のレビューを使用する。ある商品の商品属性とともに、11,067件のレビューをそれぞれ入力し、関連度を求める。例えば、“PENTAX K-3 Mark III ボディ”の商品属性とともにレビューを1件ずつ入力していき、関連度を出力する。出力した関連度について、最も高い関連度を持つレビューを1位、最も低い関連度を持つレビューを11,067位となるようにランキングを行う。ランキングを行った11,067件のレビューについて、“PENTAX K-3 Mark III ボディ”に記述されたレビューが上位何件目に現れるかを評価する。他の商品属性についても、商品属性とレビューを入力し、同様に評価を行う。この評価により、実際にその商品に書かれたレビューについて、高い関連度で予測できているかを評価することができる。ランキングした結果について、平均逆順位 (MRR) を用いて評価する。

#### 5.4 類似商品のレビューを用いた関連度推定タスクによる評価

これまでの評価では、実際に入力した商品に記述されたレビューを判定できるかを評価していた。しかし、商品の中には類似した商品属性を持つ商品も多く存在する。本来であれば、そのような類似した商品属性を持つ商品に付けられたレビューも関連があると予測できることが望ましい。類似した商品属性を持つ商品として、例えば、シリーズ物の商品が挙げられる。このようなシリーズ物の商品では、先行して販売された商品と直後に発売された後継商品では、商品属性が大きく変わらないことが期待される。そのため後継商品に記述されているレビューは、先行して販売された商品のレビューと類似している可能性が高い。そこで、先行して販売された商品の商品属性に対して、後継商品に記述されたレビューを入力した際に関連ありと予測することができるのかを評価する。

実験に用いたデータについて説明する。はじめに、訓練データとして使用した408商品のうち、9商品を選択する。これらの訓練データとして使用した9商品の後継商品を、テストデータとして使用した商品から9商品抽出する。これにより、訓練データに含まれる先行して販売された商品である9商品と、テストデータに含まれる後継商品である9商品により、9商品ペアが作成できる。次に、後継商品である9商品に記述されたレビューを、テストデータから抽出する。レビューは、価格.comで評価観点として挙げられている、画質、携帯性、デザイン、操作性、バッテリー、機能性、液晶、ホールド感という8つの観点から1文ずつ抽出した。1商品につき8つのレビューを使用し、9商品ペア用いるため、合計72文のレビューについて実験を行った。今回抽出したレビューは、過去の商品と比較を行っている、商品名やメーカー名が含まれているものは除いた。先に発売された商品の商品属性と後継商品のレビューを提案モデルに入力し、関連度を二値で出力する。商品属性とレビューに関連があれば「関連あり」、商品属性とレビューに関連がなければ「関連なし」となる。今回は、類似した商品属性を持つ商品に記述されたレビューをすべて正解とするため、72件すべてのレビューの正解ラベルは「関連あり」とする。評価尺度には正解率を用いる。

表3 各モデルにおける関連度推定タスクの正解率。

量的データの入力方法	追加事前学習	正解率
量的データをそのまま入力	あり	0.736
量的データをそのまま入力	なし	0.694
属性ごとにカテゴリを表すトークンで表現	あり	0.728
属性ごとにカテゴリを表すトークンで表現	なし	0.711
カテゴリを数値で表現	あり	0.713
カテゴリを数値で表現	なし	0.703
属性間で共通のカテゴリを表すトークンで表現	あり	0.719
属性間で共通のカテゴリを表すトークンで表現	なし	0.705

表4 レビューランキングタスクによるMRR。

量的データの入力方法	追加事前学習	MRR
量的データをそのまま入力	あり	0.320
量的データをそのまま入力	なし	0.203
属性ごとにカテゴリを表すトークンで表現	あり	0.292
属性ごとにカテゴリを表すトークンで表現	なし	0.206
カテゴリを数値で表現	あり	0.343
カテゴリを数値で表現	なし	0.169
属性間で共通のカテゴリを表すトークンで表現	あり	0.315
属性間で共通のカテゴリを表すトークンで表現	なし	0.198

## 6 結果

### 6.1 関連度予測タスク

関連度予測タスクの結果について述べる。各手法における関連度予測タスクの正解率について表3に示す。

まず、追加事前学習なしの場合について正解率を比較する。このとき、ベースラインである「量的データをそのまま入力」の手法の正解率が最も低く、量的データの入力を工夫した提案手法はベースラインよりも正解率が高いことが分かる。このことは、追加事前学習を行わない条件においては、量的データの入力を工夫することにより、商品属性とレビューの関連度予測の精度が向上したことを示している。

つぎに、追加事前学習ありの場合となしの場合について正解率を比較する。このとき、どのモデルも追加事前学習を行うことで正解率が向上していることが分かる。このことは、レビューと商品の関連を理解したモデルを構築する上で、追加事前学習が有用であることを示している。

最後に、追加事前学習ありの場合について正解率を比較する。追加事前学習ありのモデルの中では、ベースラインが最も高い正解率となることが分かる。追加事前学習を行わない場合は、ベースラインは最も低い正解率であったが、追加事前学習を行うことで量的データの入力を工夫したモデルよりも高い正解率を達成できることが分かる。

これらの結果から以下のことが推察される。追加事前学習を行わない場合は、モデルの学習に用いるデータが比較的少ないため、ベースラインでは量的データを適切に扱うことができず、正解率が低くなってしまふ。一方、追加事前学習を行うことで、学習に利用するデータが比較的多くなるため、ベースラインでも商品属性と量的データの関連度を適切に推定できるようになったのではないかと考えられる。

### 6.2 レビューランキングタスク

レビューランキングタスクの結果を表4に示す。関連度推定タスクと同様に、追加事前学習を行うことでMRRの向上が見られた。まず、関連度予測タスクと同様に、どのモデルについても追加事前学習を行うことで、MRRが向上していることが分かる。一方で、追加事前学習なしの場合について比較すると、量的データをカテゴリを表すトークンを用いる手法がベースラインのMRRを上回っているものの、他の手法はベースラインを下回

表5 先行販売された商品の商品属性と後継商品のレビューを入力し予測を行った際の正解率。

量的データの入力方法	追加事前学習	正解率
量的データをそのまま入力	あり	0.736
量的データをそのまま入力	なし	0.361
属性ごとにカテゴリを表すトークンで表現	あり	0.736
属性ごとにカテゴリを表すトークンで表現	なし	0.625
カテゴリを数値で表現	あり	0.902
カテゴリを数値で表現	なし	0.805
属性間で共通のカテゴリを表すトークンで表現	あり	0.791
属性間で共通のカテゴリを表すトークンで表現	なし	0.666

っていた。追加事前学習ありの場合についてみてみると、量的データをカテゴリ化し数値で表現する手法が0.343とベースラインの0.320よりも高いMRRを達成していることが分かった。

### 6.3 類似商品のレビューを用いた関連度推定タスク

類似商品のレビューを用いた関連度推定タスクにおける各手法の正解率を表5に示す。先の2つのタスクと同様、追加事前学習により精度が向上していることがわかる。また、追加事前学習を行わない場合は、ベースライン手法は他の手法に比べて低い精度となっていることが分かった。

本タスクで正例として扱われたデータ、すなわち商品属性とレビューのペアは異なる商品に記述されたレビューのため、先の2つのタスクでは、評価を行う際には負例として扱われていた。表3と表5の追加事前学習を行なった場合のベースラインの精度から、類似した商品属性に記述されたレビューに対しても、同一の商品を入力した場合と同様の精度で関連度を予測できていることが分かる。

### 6.4 考察

3つの評価実験の結果から、追加事前学習を行うことで精度が向上することが分かった。マスクした属性値を当てたり、属性値をもとにマスクされたテキスト内のトークンを当てたりする学習を追加事前学習では行っていない。このような学習を行うことで、属性間の関係やテキストと属性の関係を学ぶことができたため、精度向上につながったと考えられる。追加事前学習が商品属性とレビューの関連度推定に有用であることが分かった一方で、量的データの入力方法を工夫することについては、今回の実験結果か

らは有用性を示す結果は得られなかった。

本研究では、類似した商品属性を用いた学習を行うことができていないという制約がある。学習において、正例を実際の商品に記述されたレビュー、負例を異なる商品に記述されたレビューを用いた。しかし、実際の商品に記述されたレビューがない場合には、同様の手法で学習を行うことはできない。その場合には、類似した商品属性を持つ商品のレビューを代替して用いるといった方法が考えられる。例えば、カメラにはシリーズ物の商品があり、多くの後継商品が販売されている。シリーズ物の商品では、先行に販売された商品と後継商品の商品属性は、類似している。十分にレビューがない場合でも、類似商品に記述されたレビューを用いることで、本研究で提案した手法により学習することが可能であろう。

しかし、このような類似商品の商品属性とレビューはその商品に記述されるレビューと必ずしも合致しているとは限らない。そのため、構築したモデルの精度が本研究で示した値よりも低くなってしまふことが考えられる。類似商品用いて学習する際には、どのようなレビューを学習として持ってくるのがふさわしいのか、また、学習に用いる類似商品の選択方法について検討する必要がある。

## 7 まとめ

本研究では、商品属性とレビューの関連を理解したモデルの構築に取り組んだ。そのために、商品属性とレビューをBERTに対して入力することで関連度を推定するモデルを提案した。さらに、追加事前学習と入力の商品属性に量的データのカテゴリ化を行い、どのような手法が関連度推定に適しているのかを評価した。関連度推定タスク、レビューランキングタスク、類似商品のレビューを用いた関連度推定タスクどちらにおいても、追加事前学習を行うことで精度の向上が見られ、追加事前学習の重要性を示すことができた。今後は、量的データの入力方法に関してより詳細な実験を行うとともに、類似商品の商品属性を用いた学習手法について検討する必要がある。

謝辞 本研究はJSPS科学研究費助成事業JP24K03228, JP22H03905, JP21H03554, JP21H03775による助成を受けたものです。ここに記して謝意を表します。

## 参考文献

- [1] Chung-Chi Chen, Hen-Hsen Huang, Hiroya Takamura, and Hsin-Hsi Chen. Numeracy-600k: Learning numeracy for detecting exaggerated information in market comments. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 6307–6313, 2019.
- [2] Shiqian Chen, Chenliang Li, Feng Ji, Wei Zhou, and Haiqing Chen. Review-driven answer generation for product-related questions in e-commerce. *arXiv preprint arXiv:1905.01994*, 2019.
- [3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.
- [4] Chenhao Fang, Xiaohan Li, Zezhong Fan, Jianpeng Xu, Kaushiki Nag, Evren Korpeoglu, Sushant Kumar, and Kannan Achan. LLM-ensemble: Optimal large language model ensemble method for e-commerce product attribute value extraction. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2910–2914, 2024.
- [5] Yue Feng, Zhaochun Ren, Weijie Zhao, Mingming Sun, and Ping Li. Multi-type textual reasoning for product-aware answer generation. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1135–1145, 2021.
- [6] Gilad Fuchs, Haggai Roitman, and Matan Mandelbrod. Automatic form filling with form-BERT. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1850–1854, 2021.
- [7] Shen Gao, Xiuying Chen, Zhaochun Ren, Dongyan Zhao, and Rui Yan. Meaningful answer generation of e-commerce question-answering. *ACM Transactions on Information Systems*, 39(2):1–26, 2021.
- [8] Shen Gao, Zhaochun Ren, Yihong Zhao, Dongyan Zhao, Dawei Yin, and Rui Yan. Product-aware answer generation in e-commerce question-answering. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, pages 429–437, 2019.
- [9] Tatsuya Iwanari, Naoki Yoshinaga, Nobuhiro Kaj, Toshiharu Nishina, Masashi Toyoda, and Masaru Kitsuregawa. Ordering concepts based on common attribute intensity. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, page 3747–3753, 2016.
- [10] Zhihua Jin, Xin Jiang, Xingbo Wang, Qun Liu, Yong Wang, Xiaozhe Ren, and Huamin Qu. NumGPT: Improving numeracy ability of generative pre-trained models. *arXiv preprint arXiv:2109.03137*, 2021.
- [11] Bill Yuchen Lin, Seyeon Lee, Rahul Khanna, and Xiang Ren. Birds have four legs?! numersense: Probing numerical commonsense knowledge of pre-trained language models. 2020.
- [12] Julian McAuley and Alex Yang. Addressing complex and subjective product-related queries with customer reviews. In *Proceedings of the 25th International Conference on World Wide Web*, pages 625–635, 2016.
- [13] Aakanksha Naik, Abhilasha Ravichander, Carolyn Rose, and Eduard Hovy. Exploring numeracy in word embeddings. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3374–3380, 2019.
- [14] Alexander Rietzler, Sebastian Stabinger, Paul Opitz, and Stefan Engl. Adapt or get left behind: Domain adaptation through bert language model finetuning for aspect-target sentiment classification. *arXiv preprint arXiv:1908.11860*, 2019.
- [15] Kalyani Roy, Smit Shah, Nithish Pai, Jaidam Ramtej, Prajit Nadkarni, Jyotirmoy Banerjee, Pawan Goyal, and Surender Kumar. Using large pretrained language models for answering user queries from product specifications. In *Proceedings of The 3rd Workshop on e-Commerce and NLP*, pages 35–39, 2020.
- [16] Chi Sun, Xipeng Qiu, Yige Xu, and Xuanjing Huang. How to fine-tune BERT for text classification? In *Chinese Computational Linguistics: 18th China National Conference*, pages 194–206, 2019.
- [17] Alon Talmor, Yanai Elazar, Yoav Goldberg, and Jonathan Berant. olmpics-on what language model pre-training captures. *Transactions of the Association for Computational Linguistics*, 8:743–758, 2019.
- [18] Qifan Wang, Li Yang, Bhargav Kanagal, Sumit Sanghai, D Sivakumar, Bin Shu, Zac Yu, and Jon Elsas. Learning to extract attribute value from product via question answering: A multi-task approach. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 47–55, 2020.
- [19] Shiwei Zhang, Jey Han Lau, Xiuzhen Zhang, Jeffrey Chan, and Cecile Paris. Discovering relevant reviews for answering product-related queries. In *2019 IEEE International Conference on Data Mining*, pages 1468–1473, 2019.
- [20] Xikun Zhang, Deepak Ramachandran, Ian Tenney, Yanai Elazar, and Dan Roth. Do language embeddings capture scales? In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 4889–4896, 2020.
- [21] You Zhang, Jin Wang, Liang-Chih Yu, and Xuejie Zhang. MA-BERT: Learning representation by incorporating multi-attribute knowledge in transformers. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 2338–2343, 2021.
- [22] Chen Zhao, Yuki Nakayama, and Koji Murakami. Multi-task learning for product information with fine-tuned BERT. 言語処理学会第26回年次大会発表論文集, pages 736–739, 2020.
- [23] 内田臣了, 山本岳洋, 加藤誠, 大島裕明, and 田中克己. 不確かな順序関係の分析に基づく経験的屬性によるオブジェクト検索. In 第8回データ工学と情報マネジメントに関するフォーラム, C2–4, 2016.