

## 1 はじめに

昨今の AI の発展は目まぐるしく、様々な応用分野でその有用性が示されている。しかしその一方で、機械学習手法はその判断根拠が不明な点が問題視されているため、近年では人工知能に判断根拠の説明を示させる「説明可能な人工知能」に関する研究が盛んである。自然言語処理の分野においても先述の研究は盛んであるため、このような研究に取り組むことを目標とし、本研究では複数のラベルが付与された日本語レビュー文章データセットの多値分類を通して、BERT の入力する文章のトークン数を限定した時の分類精度の上昇を確認した。また、そのデータセットを用いて多値分類をした時に最も注目されたトークンの抽出などを目的として、関連研究を参考にして新しいモデルの作成に取り組んだ。

## 2 要素技術

### 2.1 BERT

BERT[1] は 2018 年 10 月に Google の Devlin らの論文で発表された自然言語処理モデルである。Transformer というアーキテクチャを組み込み、文章を双方向から学習することによって文脈を読み取ることが実現された事前学習モデルである。ファインチューニングをすることで様々な自然言語処理タスクに対応することが出来る汎用性の高さが注目を集めた。本研究の実験では、東北大学の乾・鈴木研究室によって公開されている BERT 日本語 Pretrained モデル<sup>1</sup>を使用した。

### 2.2 Transformer

Transformer[2] は 2017 年に発表された深層学習モデルである。発表以前までは、翻訳などの入力文章を別の文章で出力するというモデルは Attention を用いたエンコーダ、デコーダ形式の RNN や CNN が主流であった。しかし、当モデルは再帰や畳み込みは行わず、Attention のみを用いていることから並列化も可能であるため、訓練時間の削減が実現された。具体的には自己注意機構 (self - attention) が

RNN や CNN と大きく異なる特徴であり、読み込んだデータの距離に関わらず、どのデータの値を読み込むかどうかを制御することが出来る。また、この Transformer のエンコーダ層は BERT に用いられている。本研究では Transformer のエンコーダ層を BERT のエンコーダ層に接続して用いた。

## 3 関連研究

### 3.1 アスペクトベースの感情分析

感情分析はテキストから意見や感情、態度を分析することを目的とするタスクであり、あるコンテンツの極性がポジティブかネガティブかを分類する。自然言語処理技術の発展に伴い、SNS 上での書き込みやネット上のレビューサイトにおいて、ユーザーの意見を定量的に評価することの需要が高まっている。感情分析のタスクに、文脈情報の解析に重点を置いたアスペクトベースの感情分析 [3] が存在する。文章中に含まれるアスペクト情報を利用して、その文章がどのような事柄について書かれたものかを分析する。アスペクト情報とはその文章のカテゴリを表しており、文章が何を対象として言及しているかを示すエンティティと、エンティティがどその対象のどのような側面について言及しているかを示すアトリビュートによって定義される。一般にアスペクトベースの感情分析では 3 つのステップで分析をする。最初に与えられたアスペクトカテゴリに文章を分類する。次に文中に含まれるアスペクトカテゴリに対するフレーズの位置を推定する。最後にフレーズの極性を分析する。全体としてその分析精度を向上させることを目指す。

## 4 提案手法

アスペクトベースの感情分析において渥美らが提案したモデルを参考に BERT エンコーダ層に Transformer エンコーダ層を接続したモデルを作成した。BERT エンコーダ層に Transformer エンコーダ層を接続した目的は、Transformer エンコーダ層でアスペクトカテゴリと各トークンの関連度を計算するためである。図 1 に作成したモデル図を示す。

<sup>1</sup><https://www.nlp.ecei.tohoku.ac.jp/news-release/3284/>

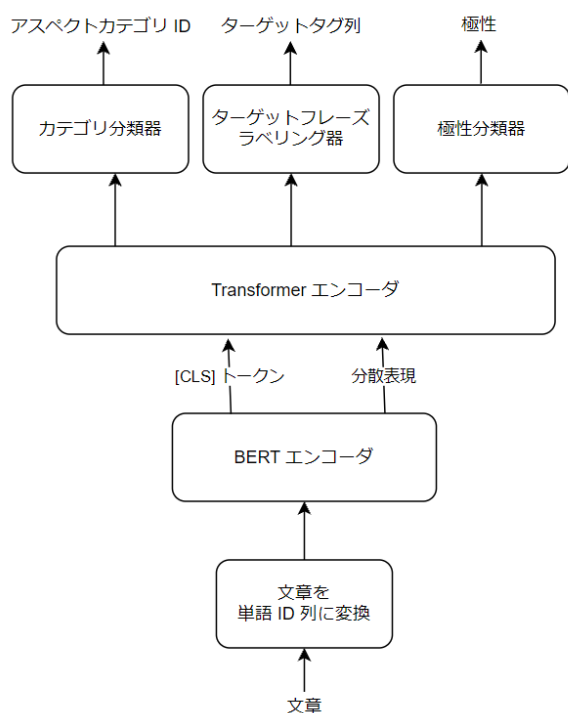


図 1: アスペクトセンチメント分析セットのモデル図

まず、文章を東北大学の乾・鈴木研究室によって公開されている BERT 日本語 Pretrained モデルを用いてトークン情報を獲得し、BERT エンコーダ層に入力する。次に、BERT エンコーダ層の出力として文章データの 1 単語目 "[CLS]" と長さ 512 のトークンの特徴量を得る。それらを Transformer エンコーダ層に入力し、Attention を計算した結果をカテゴリ分類器、ターゲットフレーズラベリング器、極性分類器に入力する。カテゴリ分類器では入力文章がどのラベルに属するかを、1 目目のトークンの "[CLS]" を用いて分類する。ターゲットフレーズラベリング器ではそれぞれの単語トークンがターゲットフレーズ（アスペクトカテゴリと類似度が高いと推定された単語）か否かの 2 値分類をする。極性分類器はターゲットフレーズを、ポジティブ、ネガティブ、ニュートラルに分類する。

## 5 データセット

### 5.1 楽天トラベルレビュー：アスペクトセンチメントタグ付きコーパス

実験 2 では日本語レビュー文章とそれぞれの文章のラベルが 14 個与えられたデータを用いた。総デー

タ数は 76624 で、そのうち全ラベルが 0 であるデータは 28255 であり、今回はこれらを除くことで少なくとも 1 つのラベルが付与されたデータのみを用いた。その結果として総データ数は 53192 となった。

## 6 実験

本実験ではラベルが複数あるデータに対して、提案手法で述べたモデルでの実験を想定してシンプルな BERT での多値分類をした。モデルの改良ではなく、データセットの調整をすることで分類精度の向上を目指した。以下にデータセットの調整内容について示す。

- 全文が英語で書かれたレビュー文を取り除いた。
- 日本語文字を全角にし、英単語を小文字に統一した。
- 具体的な数値は今回のタスクにおいて有用ではないと判断し、全てを 0 に変換し、各種記号については取り除いた。

以上の手順を踏まえた後にポジネガ両方のラベルが立っているデータと語彙数が 15 以下と 100 以上のデータを取り除いた。これらの処理を終えると楽天トラベルレビューのアスペクトセンチメントタグ付きコーパスの 53192 個のデータが 43920 になった。BERT モデルの末尾にネガポジ分類のための全結合層を追加し、出力として 14 クラス分類 [ネガティブ (0) or ポジティブ (1)] を出力するモデルを用いた。クラス分類には入力した文章データの 1 単語目 [CLS] の特徴量を利用した。また、BERT Layer の最終層と全結合層のみ fine-tuning をした。

表 1 に実験で用いたデータの内訳を示す。

表 1: 多値分類タスクに用いたデータの内訳		
総データ数	訓練データ数	テストデータ数
21000	16800	4200

## 7 今後の課題

提案手法にて示した BERT - Transformer モデルでマルチラベルデータセットを用いた実験を始める。また、分類精度の向上を図りながらモデルの改良を目指す。

## 参考文献

- [1] 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.
- [2] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pp. 5998–6008, 2017.
- [3] 三浦義栄, 渥美雅保. 事前学習言語モデルを用いたアスペクトベースセンチメント分析ニューラルネットワーク. 人工知能学会全国大会論文集 第 35 回全国大会 (2021), pp. 2Yin507–2Yin507. 一般社団法人 人工知能学会, 2021.