

アスペクトベースセンチメント分析におけるマルチラベル分類手法の提案

1 はじめに

昨今の人工知能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。その一方で、機械学習手法はその判断根拠を示すことが難しいという問題点がある。そこで近年では人工知能に判断根拠を示させる「説明可能な人工知能」に関する研究が盛んであり、自然言語処理分野においても同様である。また、日々大量のテキストデータが執筆されており、それらの有効活用が求められている。特にビジネスなどにおいては、作業の省人化とデータの有効活用を目的とした各テキストに対するラベル付与の自動化は大きな課題である。そのため 1 つの文書に複数のラベルが付与されることを想定するマルチラベル分類 (Multi Label Classification, MLC) タスクの重要性も高まっている。そこで本研究では、各ラベルの分類器を作成して、それらの出力を統合するマルチラベル分類手法を提案する。出力を統合する層とモデルの全体的な構造は、先行研究 [1] で渥美らが提案したアスペクトベースセンチメント分析ネットワークのモデルを参考にした。本研究のモデルは MLC に有効な特徴量の抽出に加えて、文書の特徴量も抽出できるため、今後は自然言語処理の分野における説明可能な人工知能の研究への活用も可能と考えられる。検証実験の結果から、提案手法の有効性を確認した。

2 要素技術

2.1 BERT

BERT[2] 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によりさまざま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¹を使用した。

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

2.2 Transformer

Transformer[3] 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、デコーダ形式の再帰的ニューラルネットワーク (Recurrent Neural Network, RNN) や畳み込みニューラルネットワーク (Convolution Neural Network, 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。また、この Transformer のエンコーダ層は BERT に使用されている。

3 関連研究

3.1 アスペクトベースセンチメント分析

アスペクトベースセンチメント分析とは、自然言語処理の分野において MLC と人工知能の説明可能性の 2 つのタスクに焦点が当てられている研究である。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 (肯定的) かネガティブ (否定的) かを推定して分類する。自然言語処理技術の発展にともない、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。また、感情分析も注目されている。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 [1] がある。アスペクト情報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、エンティティが対象の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。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を分析する。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

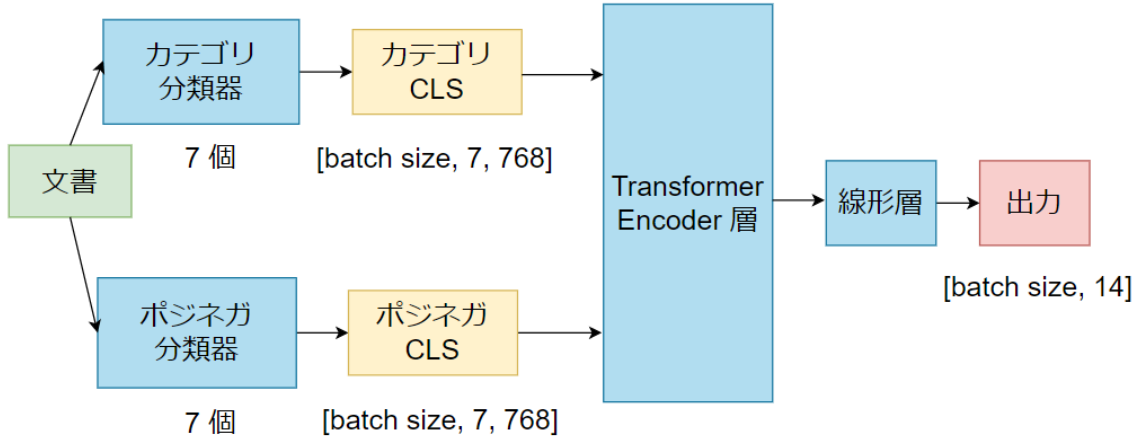


図 1: Multi pretrained models Transformer (Mpm+T) モデル概要図

表 1: カテゴリ分類器およびポジネガ分類器作成時に使用したデータ数と正解率とその標準偏差

	朝食	夕食	風呂	サービス	立地	施設	部屋
カテゴリ分類器作成に使用したデータ数	30000	30000	30000	30000	30000	30000	30000
カテゴリ分類器の正解率	0.916(0.00146)	0.923(0.00618)	0.943(0.00598)	0.888(0.00538)	0.951 (0.000677)	0.868(0.00960)	0.919(0.000844)
ポジネガ分類器作成に使用したデータ数	13000	10000	9000	18000	6400	13000	10000
ポジネガ分類器の正解率	0.9358(0.00993)	0.922(0.00542)	0.917(0.000487)	0.926(0.00668)	0.925(0.00337)	0.908(0.0943)	0.905(0.00746)

極性を分析する。そして全体としてその分析精度を向上させることを目指す。

4 データセット

データセットは楽天グループ株式会社が公開している「楽天トラベルレビュー：アспектセンチメントタグ付きコーパス」[4]を使用した。楽天トラベルの日本語レビュー文章とそれぞれの文章について、立地、部屋、食事等の7項目のカテゴリに対するポジティブまたはネガティブのタグが付与されている。「朝食、夕食、風呂、サービス、施設、立地、部屋」のポジティブ、ネガティブの14個のカテゴリに分類される。今回は14のいずれのカテゴリにも属さないデータを除くことで、少なくとも1つのラベルに属し、語彙数が10以下と100以上のデータを取り除いた。また、いずれのカテゴリにも属さないデータと、1つのカテゴリにポジティブとネガティブの両方が付与されたデータを取り除いた。総データ数は48354である。

5 提案モデル

5.1 Multi pretrained models Transformer (Mpm+T)

MLC タスクにおいて1つの文章から複数のラベル分類情報を持つ特徴量を抽出することは困難である。そこで本研究では、各ラベルごとに2値分類器を作成してそれらを統合することにより、複数のラベル分類情報を持つ特徴量をより高い精度で抽出できると仮定して、Multi pretrained models Transformer (Mpm+T) モデルを提案する。図1に本研究で作成したMpm+Tのモデル概要図を示す。モデルにおけるカテゴリCLSとポジネガCLSはBERTで使用されているタグの分散表現ベクトルである。本提案モデルの全体的な構造と、カテゴリCLSとポジネガCLSの内積計算機構としての使用は、先行研究[1]で渥美らによる提案モデルであるアспектベースセンチメント分析ネットワークを参考にした。本モデルを比較した時の新規な点は、マルチラベル分類を制御するのが単一のBERT層ではなく、各カテゴリごとに作成した小分類器がする点である。本モデルの全体的な処理の流れを説明する。最初に文書データをカテゴリ分類器とポジネガ分類器に入力す

る。そして入力文章から各トークンの特徴量を抽出して、その先頭トークンである CLS トークンをそれぞれの分類器においてカテゴリ CLS, ポジネガ CLS として獲得する。次に Transformer Encoder 層ではカテゴリ CLS とポジネガ CLS の内積を Attention 機構を使用して導出する。最後に、得られた導出結果を線形層でマルチラベル分類に使用する特徴量に変換する。表 1 にカテゴリ分類器とポジネガ分類器の 5 分割交差検証時に使用したデータ数と正解率を示す。以下ではモデル概要図の詳細について説明する。

5.2 カテゴリ分類器とカテゴリ CLS

カテゴリ分類器とは、BERT 日本語事前学習済みモデルに楽天トラベルレビューのデータセットを学習して作成した、カテゴリにおいて属しているか否かの 2 値分類に使用する特徴量を得るための学習済み分類器である。7 種類のカテゴリごとに作成したため、7 個のカテゴリ分類器が存在する。なお、作成時には訓練データ数が 24000 であり、検証データ数が 6000 となるような 5 分割交差検証をした。カテゴリ CLS とは、文書データがカテゴリ分類器によって変換された特徴量に含まれる CLS トークンを指しており、分類に用いる情報を含んでいる。

5.3 ポジネガ分類器とポジネガ CLS

ポジネガ分類器とは、BERT 日本語事前学習済みモデルに楽天トラベルレビューのデータセットを学習して作成した、カテゴリにおいて属しているか否かの 2 値分類に使用する特徴量を得るための学習済み分類器である。7 種類のカテゴリごとに作成したため、7 個のカテゴリ分類器が存在する。なお、作成時の 5 分割交差で各カテゴリごとにポジティブまたはネガティブのラベルが付与されているデータを使用する必要があったため、分類器によって使用したデータ数は異なる。表 1 に使用したデータ数を示す。ポジネガ CLS とは、文書データがポジネガ分類器によって変換された特徴量に含まれる CLS トークンを指しており、分類に用いる情報を含んでいる。

5.4 Transformer Encoder 層と線形層

Transformer Encoder 層では、カテゴリ分類器とポジネガ分類器で得られたそれぞれの CLS トーク

ンの内積を Attention 機構を使用して計算する。なお、Transformer Encoder 層の層数は 2 である。線形層では Transformer Encoder 層で得られた 768 次元の特徴量を 14 次元に圧縮して Sigmoid 関数に入力することで、正規化をする。

6 実験

提案モデルの有効性の検証のため、評価実験をした。MLC の多くの関連研究と同様に評価指標として micro-F1 を選択した。本実験においては Transformer Encoder 層のみを 25000 のデータで 5 分割検証をして、学習済みモデルであるカテゴリ分類器とポジネガ分類器は追加の学習はしなかった。また、Transformer Encoder 層における Query と Key と Value に割り当てる CLS トークンは、

- ・パターン 1「Query: カテゴリ CLS, Key, Value: ポジネガ CLS」と、
- ・パターン 2「Query: ポジネガ CLS, Key, Value: カテゴリ CLS」

の 2 つのパターンを用意して実験をした。いずれのパターンにおいても、Transformer Encoder 層の Multi-Head Attention 機構のヘッド数を 1, 4, 8 の場合で検証をした。

6.1 比較手法

3 種類の比較手法を紹介する。

1. 今回の実験におけるベースラインとして、RandomForest を使用して訓練データ数 20000 テストデータ 5000 で 14 クラスの MLC をする。
2. 単一の BERT のみで、データ数 25000 の 5 分割交差検証をする。
3. Transformer Encoder 層における Attention 機構のような発展的な内積計算手法を使用せずに、カテゴリ CLS とポジネガ CLS の内積を計算するために 6 層の多層パーセプトロン (Multi Layer Perceptron, MLP) を使用したモデル (Mpm+MLP) を作成して、訓練データ数 20000 テストデータ 5000 で MLC をする。

6.2 実験結果

表 2 に実験結果を示す。表 2 の結果から、提案モデルが RandomForest と単一の BERT を使用した

表 2: 実験結果の評価指標の平均と標準偏差

評価指標	Precision	Recall	micro-F1
RandomForest	0.292	0.908	0.442
BERT	0.579(0.0359)	0.775(0.0235)	0.679(0.0299)
Mpm+MLP	0.437	0.739	0.549
Mpm+T パターン 1 ヘッド数 1	0.803(0.0177)	0.824(0.0121)	0.813(0.00704)
Mpm+T パターン 1 ヘッド数 4	0.824(0.00625)	0.837(0.00672)	0.830(0.00251)
Mpm+T パターン 1 ヘッド数 8	0.811(0.00732)	0.850(0.0103)	0.830(0.00171)
Mpm+T パターン 2 ヘッド数 1	0.811(0.0105)	0.808(0.0202)	0.810(0.00750)
Mpm+T パターン 2 ヘッド数 4	0.827(0.00563)	0.832(0.00397)	0.830(0.00242)
Mpm+T パターン 2 ヘッド数 8	0.828(0.0143)	0.832 (0.0124)	0.832(0.00469)

MLC モデルを上回る精度で分類出来ていることが確認できた。また, Mpm+MLP モデルと Mpm+T モデルの比較から, カテゴリ CLS とポジネガ CLS の内積計算機構として MLP より Transformer Encoder 層が適切であることが確認できた。一方で, Transformer Encoder 層における Query, Key, Value の割り当てについてはパターン 1 とパターン 2 で大きな差異がないことが確認できた。Multi-Head Attention 機構のヘッド数を増やすことで分類精度が向上したが, ヘッド数が 4 と 8 では大きな差異が確認できなかった。つまり, ヘッド数は 2 または 3 でも同様の分類精度が確認できる可能性があると考えられる。

7 まとめと今後の課題

本研究では MLC において, 複数のラベル分類情報を持つ特徴量をより高い精度で抽出できるモデルを提案した。そして, 「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」を使用した評価実験において, BERT のみの MLC モデル上回る分類精度が確認できた。また, Mpm+T と Mpm+MLP の比較実験より, 内積計算機構として Transformer Encoder 層が適切であることが確認できた。続いて, Transformer Encoder 層における Query, Key, Value の割り当ての 2 パターンについては, 今回の実験では大きな差異が確認できなかったが, 今後のタスクにおいても比較の継続を検討している。最後に, Multi-Head Attention 機構のヘッド数は複数であれば分類精度は向上するが, ヘッド数が 4 と 8 では大きな差異が確認できなかった。今後の課題として, 実験時には使用しなかった 1 つのラベルにポジティブとネガティブの両方のラベルが付与されたデータも

使用して検証実験をすることや, 提案モデルのハイパーパラメータの最適化をすることなどが挙げられる。加えて, データセットを自然言語処理分野における説明可能な人工知能についての研究に使用するために拡張することが挙げられる。

参考文献

- [1] 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 (2021), pp. 2Yin507–2Yin507. 一般社団法人 人工知能学会, 2021.
- [2] 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.
- [3] 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.
- [4] 楽天グループ株式会社. 楽天データセット (コレクション), aug 2010.