

アンサンブル学習を利用した深層言語モデルによる マルチラベル分類手法の改良

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と、その精度を向上させることは大きな課題となっている。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) タスクの重要性が高まっている。そこで本研究では、アンサンブル学習を利用することで、マルチラベル分類の精度を向上させる深層言語モデルを提案する。特徴量の内積計算手法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と同様である。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。

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。アスペクトベースセンチメント分析におけるアスペクトとは、文章が書かれた対象 (entity) と、その属性 (attribute) によって定義される。

文章中に含まれるアスペクトを利用して、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。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

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

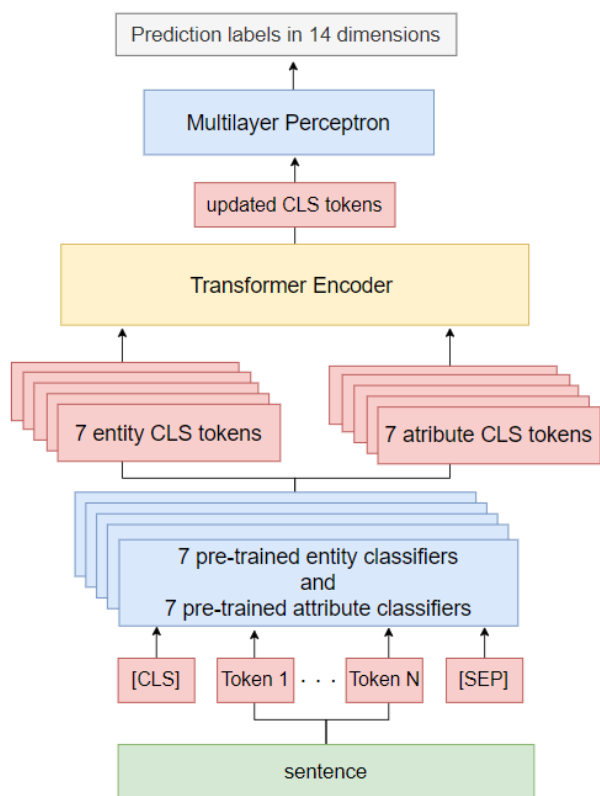


図 1: Multi pre-trained models Transformer (Mpm+T) のモデル概要図

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。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：アスペクト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、「朝食、夕食、風呂、サービス、施設、立地、部屋」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。全 76623 のデータから、ノイズを含むデータや日本語以外の言語で書かれたデータと、文章のトークン数が 101 以上のデータを除くことで、データ数は 73461 となった。この前処理後のデータセットの内訳は、14 のいずれのカテゴリにも属さないデータ数は 23429 であり、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数は 1858 である。これらの 2 種類のデー

タを取り除いてモデルの学習や評価に使用するデータ数は 48354 である。

5 提案モデル

5.1 Multi pre-trained 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容易ではない。そこで、アスペクトベースセンチメント分析におけるアスペクトに着目した。アスペクトは文章の書かれた対象と属性によって定義されることから、それぞれの対象と属性の分類に特化した小分類器を作成することを考案した。本研究では、マルチラベル分類手法を改良するためにアンサンブル学習を利用した深層言語モデルである、Multi pre-trained 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 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、entity CLS tokens と attribute CLS tokens の内積計算機構として Transformer Encoder を使用することは、先行研究 [1] で三浦らの提案モデルであるアスペクトベースセンチメント分析ネットワークに倣っている。本モデルを比較して 2 つの新規な点を挙げる。

- 1. マルチラベル分類をするのが単一の BERT モデルではなく、アスペクトにおける対象と属性についての小分類器を作成した点である。
- 2. Transformer Encoder への入力として、文書トークンの分散表現を使用せずに、CLS トークンのみを用いている点も挙げられる。一般的に BERT や Transformer を用いた文書分類タスクでは分散表現を各文書トークンの Attention の導出に使用している。しかし本研究では、各ラベルに対応する Attention を導出することを目地的として、小分類器から文書分類情報を保持した CLS トークンのみを Transformer Encoder に入力する。

続いて、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 pre-trained entity classifiers と pre-trained attribute classifiers に入力する。そして入力文書から各トークンの特徴量を抽出して、その先

表 1: pre-trained entity classifiers および pre-trained attribute classifiers 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
entity classifiers 作成に使用したデータ数	30000	30000	30000	30000	30000	30000	30000
entity classifiers の正解率	0.916 \pm 0.00146	0.923 \pm 0.00618	0.943(0.00598)	0.888(0.00538)	0.951 (0.000677)	0.868(0.00960)	0.919(0.000844)
attribute classifiers 作成に使用したデータ数	13000	10000	9000	18000	6400	13000	10000
attribute classifiers 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)

頭トークンである CLS トークンをそれぞれの小分類器において entity CLS tokens と, attribute CLS tokens として獲得する. 次に Transformer Encoder では entity CLS tokens と attribute CLS tokens 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多層パーセプトロン (Multilayer Perceptron, MLP) で MLC に使用する特徴量に次元圧縮をする.

表 1 に entity classifiers と attribute classifiers 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 pre-trained entity classifiers と entity CLS tokens

pre-trained entity classifiers 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 7 種類の小分類器である. 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. entity CLS tokens とは, 文書データが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, 2 値分類に用いる情報を含んでいる.

5.3 pre-trained attribute classifiers と attribute CLS tokens

pre-trained attribute classifiers 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, ポジティブまたはネガティブに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 7 種類の小分類器である. 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. attribute CLS tokens とは, 文書データが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, 2 値分類に用いる情報を含んでいる.

5.4 Transformer Encoder と MLP

Transformer Encoder では, 小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. MLP では Transformer Encoder 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, 30000 のデータで, 訓練データ 24000 件と検証データ 6000 件となる様な 5 分割交差検証をした. MLC の多くの関連研究と同様に評価指標として micro-F1 を用いることに加えて, 完全正解率 (Perfect Accuracy) と部分的正解率 (Partical Accuracy) を定義する. 完全正解率とは, 全データの中で正解データと予測データが全て一致した割合を指している. 部分的正解率とは, 全データの中で正解データと予測データが部分的に一致した割合を指す.

6.1 比較手法

2 種類の比較手法を紹介する. 従来手法として, BERT+MLP と, 三浦らのアスペクトベースセンチメント分析ネットワークの模倣モデルを選択する. mpm+T と同様のデータ数で 5 分割交差検証をした.

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
Mpm+T	0.848 \pm 0.0148	0.866 \pm 0.0133	0.851 \pm 0.00259
BERT+MLP	0.683 \pm 0.00468	0.772 \pm 0.00871	0.724 \pm 0.00462
三浦らのモデル	-	-	-

表 3: テストデータの完全 / 部分的正解数とマルチ / シングルラベルの完全正解率

評価指標	完全正解数	部分的正解数	マルチラベルの完全正解率	シングルラベルの完全正解率
Mpm+T	4228	1772	0.625	0.746
BERT+MLP	3188	2812	0.418	0.641
三浦らのモデル	-	-	-	-

6.2 実験結果

表 2 に 5 分割交差検証の結果を示す。Mpm+T が比較手法を上回る分類精度であることが確認できる。続いて、それぞれのモデルにおける micro-F1 が最良のモデルで 6000 件のテストデータの評価をする。6000 件に含まれるマルチラベルデータ数 2060 件であり、シングルラベルデータ数は 3940 件である。表 3 にテストデータの完全正解と部分的正解データ数および、マルチラベルデータとシングルラベルデータの完全正解率を示す。完全正解数と、マルチラベル / シングルラベルの完全正解率は高ければ良い値であり、表 3 から Mpm+T が比較手法を上回っていることが確認できる。特にマルチラベルの完全正解率は比較手法に比べて大きく向上していることが確認できる。マルチラベルの分類には詳細な分類情報を必要とすることから、アンサンブル学習を利用したことがこの結果に影響を与えていると考察する。

7 まとめと今後の課題

本研究では、アンサンブル学習を利用した深層言語モデルによるマルチラベル分類の改良手法について述べた。先行研究とは異なる点として、アスペクトに基づいた小分類器を導入することと、小分類器から得られる CLS トークンのみを Transformer Encoder の入力することが挙げられる。結果として、従来手法や先行研究の模倣モデルを上回る分類精度が確認できた。

今後の課題として、クレンジングで除いたデータなどを活用してモデルの汎用性を向上させることがあ

る。また、小分類器がモデル全体に与える影響や、アスペクト情報の抽出内容について確認する。

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, Łukasz 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.