

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

第 1 グループ 楠本 祐暉

1. はじめに

近年では、日々大量のテキストデータが執筆されている。それらを有効に活用するために、さまざまなシーンでテキスト分類タスクの需要が高まっている。テキスト分類タスクとは、事前に定義されたカテゴリまたはラベルをテキストに割り当てる処理である。特にビジネスシーンにおいては、作業の省人化とデータの有効活用を目的として、各テキストを自動的にカテゴリに分類すること、その精度を向上させることは大きな課題となっている。テキスト分類では 1 つのテキストに 1 つのラベルのみを割り当てる場合もあるが、データを有効活用するためにはテキストには 2 つ以上のラベルを割り当てる必要がある。このようにテキストに複数のラベルを割り当てるタスクはマルチラベル分類 (Multi Label Classification, MLC) と呼ばれ、自然言語処理 (Natural Language Processing, NLP) の分野では感情分析におけるタスクとして注目を集めている。その中でもアスペクトベースセンチメント分析 (Aspect-Based Sentiment Analysis, ABSA)[1] は文書から様々な観点を抽出し、その観点到感情極性を付与する研究である。アスペクト情報とは本研究の目的は NLP におけるテキスト分類の問題に従い、MLC の分類精度を向上させることに定めた。分類対象となるテキストデータはアスペクトベースで作成された日本語文章のカスタマーレビューであり、単一または複数のラベルが付与されている。ABSA における MLC の分類精度向上は大きな課題であり、多くの先行研究が存在するが、分類対象が日本語文章である研究は少ない。また、三浦らのモデルは精度改善の余地がある。そこで本研究では三浦らのモデルにアンサンブル学習を導入することで MLC の改良手法の有効性を検証をした。

2. データセット

2.1. 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

楽天グループ株式会社が事業等を通じて取得したデータから構築した「楽天データセット (コレクション)」[2] に含まれている「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」[3] 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 項目の観点に対して、感情極性としてポジティブまたはネガティブのタグが付与されている。つまり、全てのデータに付与されるクラス数は 14 であり、該当するクラスには 1 が、該当しないクラスには 0 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。この処理後のデータセットから更に 2 種類のデータを取り除いた。それら 2 種類のデータは 14 のいずれのカテゴリにも属さないニュートラルデータ 23428 件と、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ポジネガ両ラベルデータ 1858 件である。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 件である。14 クラスのうち、ラベル 1 が 1 つしか付与されていないデータの数は 件で、これをシングルラベルデータとする。また、14 クラスのうち、ラベル 1 が 2 つ以上付与されているデータの数は 件で、これをマルチラベルデータとする。

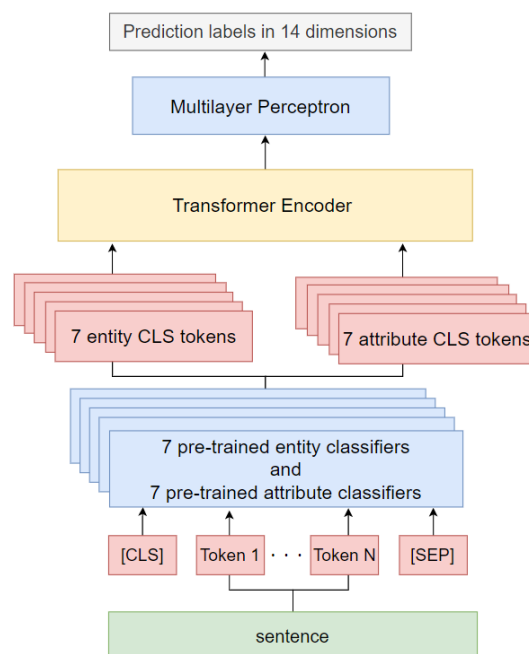


図 1: 提案モデル Multi pre-trained models Transformer (Mpm+T) の概要図

3. 提案手法

3.1. Entity Classifiers と Attribute Classifiers

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。Entity Classifiers は「朝食、夕食、風呂、サービス、施設、立地、部屋」のカテゴリごとに作成した 7 個の 2 値分類器である。文書トークンを入力することで文書の分散表現を得る。Attribute Classifiers は「朝食、夕食、風呂、サービス、施設、立地、部屋」のカテゴリごとの感情極性としてポジティブまたはネガティブ (ポジネガ) を分類する 7 個の 2 値分類器である。Entity Classifiers と同様に文書トークンを入力することで文書の分散表現を得る。

3.2. Multi pre-trained models Transformer (Mpm+T)

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。まず、はじめに文書から文書トークンを生成する。次にそれらのトークンを小分類器である entity classifiers と attribute classifiers に入力してそれぞれから CLS トークンを得る。得られたそれぞれの CLS トークンに位置情報を付与して Transformer Encoder へ入力する。そしてマルチラベル分類情報を持った CLS トークンを抽出し

表 1: Entity classifiers と Attribute Classifiers の正解率と標準偏差

	朝食	夕食	風呂	サービス	立地	施設	部屋
Entity Classifiers の正解率	0.916 \pm 0.002	0.923 \pm 0.006	0.943 \pm 0.006	0.888 \pm 0.005	0.951 \pm 0.001	0.868 \pm 0.01	0.919 \pm 0.001
Attribute Classifiers の正解率	0.936 \pm 0.001	0.922 \pm 0.005	0.917 \pm 0.001	0.926 \pm 0.001	0.925 \pm 0.003	0.908 \pm 0.094	0.905 \pm 0.007

表 2: 各 MLC モデルの評価指標の平均と標準偏差

Evaluation indices	Precision	Recall	micro-F1
Mpm+T	0.846 \pm 0.019	0.872 \pm 0.028	0.858 \pm 0.017
BERT+MLP	0.683 \pm 0.005	0.772 \pm 0.009	0.724 \pm 0.005
三浦らのモデル	0.741 \pm 0.015	0.804 \pm 0.006	0.773 \pm 0.013

表 3: 完全正解データを占めるマルチラベルデータとシングルラベルデータのデータ数と割合

データの種類	完全正解	
	マルチラベルデータ	シングルラベルデータ
Mpm+T	1298 (0.630)	3013 (0.764)
BERT+MLP	862 (0.418)	2326 (0.641)
三浦らのモデル	1097 (0.533)	2787 (0.706)

て、最後に Multilayer Perceptron (MLP) に入力することで予測ラベルを得る。本提案モデルの全体的な構造と、BERT[4] から得られる分散表現の内積計算機構として Transformer[5] を使用することは、先行研究 [6] 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。本モデルと比較した場合の 2 つの新規な点を挙げる。1 点目は、各アスペクトの分類に特化した小分類器である Entity Classifiers と Attribute Classifiers をアンサンブル的に導入したことである。2 点目は、Transformer Encoder での計算量削減を目的として、小分類器から得られる分散表現のうち、CLS トークンのみを入力とすることである。

4. 数値実験

4.1. 実験 1

実験 1 では Mpm+Tn におけるアンサンブル学習のための小分類器である Entity Classifiers と Attribute Classifiers を作成した。Entity Classifiers はテキストのエンティティがあるカテゴリに該当するならラベル 1 を、そのカテゴリに該当しないならラベル 0 を付与する。Attribute Classifiers はテキストのアトリビュートを分析してポジティブまたはネガティブを 2 値分類する。ポジティブならラベル 1 を、ネガティブならラベル 0 を付与する。Entity Classifiers と Attribute Classifiers の識別性能を検証するために 5 分割交差検証をした。表 1 に Entity Classifiers と Attribute Classifiers の正解率と標準偏差を示す。ほとんどのカテゴリにおいて小分類器は高い精度で 2 値分類ができている。しかし、Entity Classifiers のサービスと施設の正解率は他のカテゴリに比べて低くなっている。これは Entity Classifiers の予測が 0 に寄ってしまう傾向があることと、サービスと施設のカテゴリにおける正例と負例の割合に起因する。各カテゴリに付与されている正例の割合の平均は 0.237 となっている一方で、サービスと施設はそれぞれ 0.372, 0.272 と平均を大きく上回る割合である。そのため、予測が 0 に寄る傾向がある中で正例の多さが分類を難しくしたと考えられる。

4.2. 実験 2

実験 2 では Mpm+T 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 件となるような 5 分割交差検証をした。比較手法として、2 種類を紹介する。従来手法の 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 にそれぞれ

のテストデータ 6000 件でのモデルの評価指標と標準偏差を示す。関連研究 [6] と同様に各モデルの micro-F1 を比較すると、Mpm+T は 85% 程度と大きく向上した。ここで本研究の評価指標として、完全正解を定義する。完全正解とは、正解データと予測データの 14 個のラベルが全て一致した場合を指す。表 3 にテストデータ 6000 件において、各モデルの完全正解したマルチラベルデータとシングルラベルデータの数と割合を示す。Mpm+T の完全正解データ数はマルチラベルとシングルラベル共に比較手法より大きく向上していることが確認できる。特にマルチラベルデータの完全正解データ数の向上が著しく、データ数としては三浦らのモデルより 200 件多く正解している。このことから Mpm+T は比較手法よりも MLC に特化したモデルであるといえる。また、ABSA においてアンサンブル学習を導入することの有効性が確認できた。

5. 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。先行研究と比較した新規性として、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、それぞれの小分類器がアスペクト情報を抽出できていることを確認する必要がある。

参考文献

- [1] B. Liu. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1):1–167, 2012.
- [2] 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.
- [3] 楽天グループ株式会社. 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, feb 2021.
- [4] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.
- [5] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pp. 5998–6008, 2017.
- [6] 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.