

# An improved method for multi-label classification methods using deep neural language models with ensemble learning

Yuuki Kusumoto<sup>1†</sup> and Makoto Okada<sup>2</sup> and Naoki Mori<sup>3</sup>

<sup>1,2,3</sup> Osaka Metropolitan University, Japan

(<sup>1,2,3</sup>Tel: 072-254-7511; E-mail: <sup>1</sup>kusumoto@ss.cs.osakafu-u.ac.jp, <sup>2</sup>okada@omu.ac.jp, <sup>3</sup>mori@omu.ac.jp)

**Abstract:** 本稿では旅行会社が公開している日本語レビュー文章を、複数のラベルに分類する従来手法を改良するための深層言語モデルを提案する。昨今のビジネスシーンでは大量のテキストデータを活用する必要性が高まっており、各テキストに対する複数のラベル付与を自動化すること、その精度を向上させることは大きな課題となっている。しかし、複数のラベルに属するテキストを分類することは容易ではないため、従来手法の改良が求められている。そこで我々は、アンサンブル学習を利用して各ラベルの分類に特化した小分類器を作成してそれらを統合することで、複数のラベルが付与されるテキストデータの分類手法を改良する深層言語モデルを提案し、その有効性を確認した。

**Keywords:** natural language processing, multi-label classification, ensemble learning

## 1. INTRODUCTION

昨今では、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複数の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。現在は、文書がポジティブな内容であるかネガティブであるかを分類する2値分類は高い精度で解くことができるタスクである。しかし、その様なシンプルな分類手法ではテキストデータを有効に活用することは難しいため、テキストが持つ複数の意味に応じてデータを分類する必要がある。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[?]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ベースとする。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。検証実験をして、提案手法の有効性を確認する。

## 2. 要素技術

### 2.1. BERT

BERT[?] は (Bidirectional Encoder Representations from Transformers) は Transformer を用いた双方向エンコーダで構成された自然言語モデルである。文章分類、質問応答、固有表現抽出等の多様なタスクで公開当時の最高性能を達成するといった大きな成果が報告されている。BERTは事前学習したものを転移学習および、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に注目が集まった。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[?] は Transformer は自然言語処理の分野で提案された Attention 機構を備えるエンコーダとデコーダをもつ深層言語モデルである。系列データに対して Recurrent Neural Network (RNN) を用いずに学習および推論ができるため、並列化による学習速度の向上が期待できる。また、性能の高さから自然言語処理の分野においてさまざまなタスクに用いられている。

### 2.3. 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かネガティブ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[?]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。

## 3. 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」[?]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

<sup>†</sup> Yuuki Kusumoto is the presenter of this paper.

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

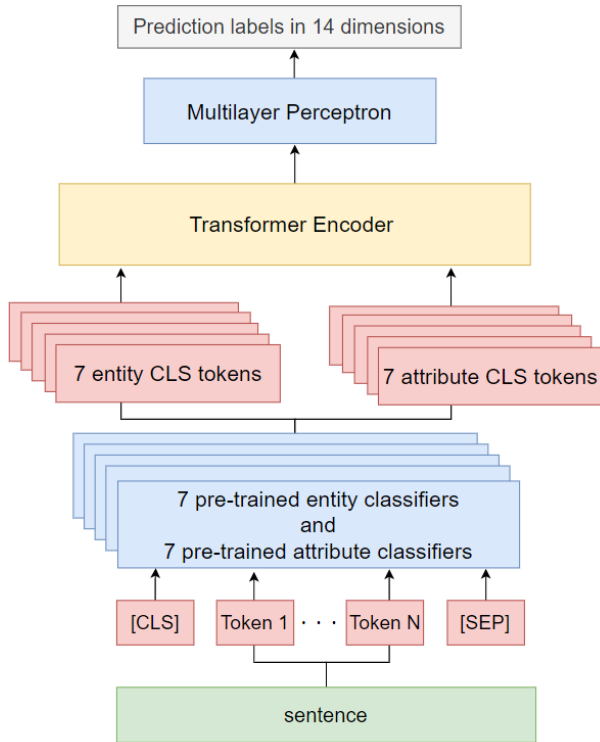


Fig. 1. Schematic diagram of the proposed model Multi pre-trained models Transformer (Mpm+T)

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。

#### 4. 提案モデル

##### 4.1. 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[?]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。

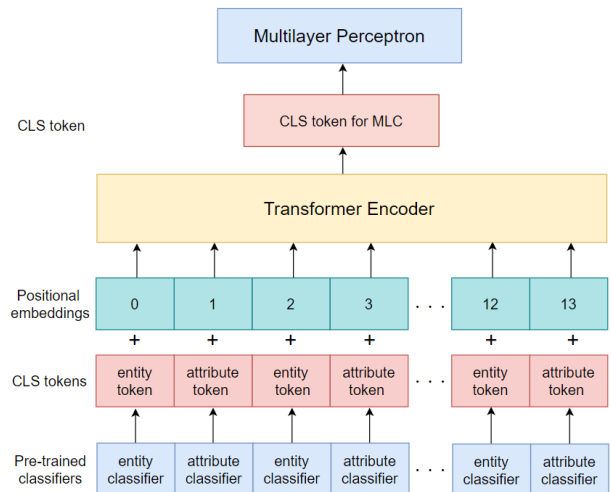


Fig. 2. 提案モデル Multi pre-trained models Transformer (Mpm+T) の概要図

- 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。これにより、三浦らのモデルや一般的な MLC よりも Transformer Encoder に入力する情報量の削減を実現した。

続いて、本モデルの全体的な処理の流れを説明する。図 2 にモデルの学習の様子を示す。最初に pre-trained entity classifiers と pre-trained attribute classifiers に共通の文書データを入力する。そして入力文書から各トークンの特徴量を抽出する。それらの先頭トークンである 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 として獲得して、それらの CLS tokens に位置情報を付与する。次に Transformer Encoder では CLS tokens の内積を self Attention 機構を使用して導出する。最後に、その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。以下ではモデルの詳細について説明する。図 3 に小分類器モデルの概要図を示す。

##### 4.2. pre-trained entity classifiers と entity CLS tokens

pre-trained entity classifiers は、図 3 に示すような 2 値分類モデルである。これは 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、カテゴリに属しているか否か 2 値分類する特徴量を得るための 7 種類の小分類器である。作成時には 30000 データで 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。

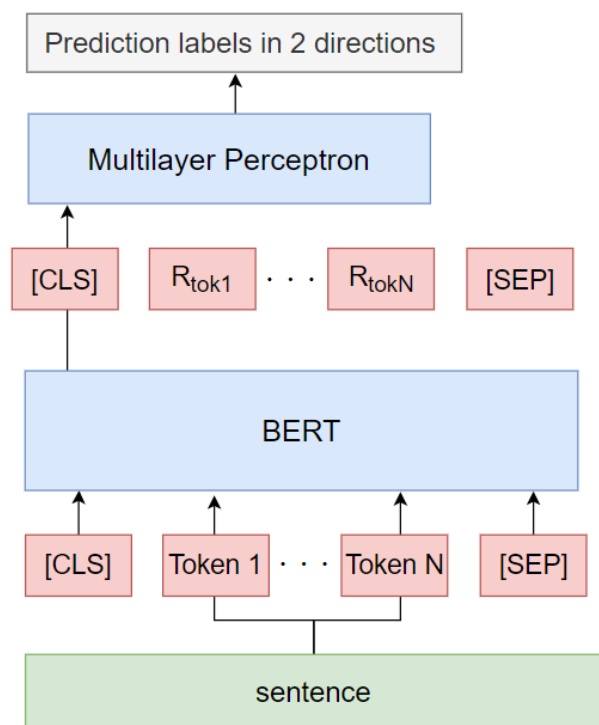


Fig. 3. 提案モデル Multi pre-trained models Transformer (Mpm+T) の概要図

#### 4.3. pre-trained attribute classifiers と attribute CLS tokens

pre-trained attribute classifiers は、図 3 に示すような 2 値分類モデルである。これは 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 2 値分類する特徴量を得るための 7 種類の小分類器である。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。

#### 4.4. 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。

### 5. 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 件となるような 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。

#### 5.1. 比較手法

2 種類の比較手法を紹介する。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。

### 6. 実験結果

表 2 に 5 分割交差検証の結果を示す。Mpm+T が各評価指標において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数を示す。完全正解数と、マルチラベル/シングルラベルの完全正解数は高ければ良い値であり、表 3 からは 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数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た。マルチラベルの分類には詳細な分類情報が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。

**Table 1.** 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 $\pm$ 0.015	0.866 $\pm$ 0.013	0.851 $\pm$ 0.003
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

**Table 2.** 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数	シングルラベルの完全正解数
Mpm+T	4228	1772	1288	2940
Mpm+T	4228	1772	1288	2940
BERT+MLP	3188	2812	862	2326
三浦らのモデル	3884	2116	1097	2787