

日本語における文脈を考慮した観点感情分析の性能評価
Performance evaluation of Aspect-Based Sentiment Analysis
considering the context of Japanese datasets

大城 龍太郎¹⁾

Ryutaro Oshiro

1 はじめに

感情分析は主に文章から感情を抽出するタスクであるが、文章には様々な感情やその対象が含まれている場合がある。例えば、映画のレビュー文などのような比較的単語数の多い文章となると、感情を表現する単語・文章が多く含まれ、positive や negative などの極性の分類が難しくなる。近年では、文章に対する感情分析手法として観点情報を用いた観点感情分析が注目されている。観点感情分析では、文章中の単語やフレーズとその属性を観点として抽出し、極性を分類する。観点単位で感情分析を行うことで、複数の感情を持つ文章においても、細かく分析を行うことができるようになる。しかしながら、観点は様々あり、観点の推定やその対象語彙、フレーズの抽出精度に関して問題を残している。そこで文脈を考慮した言語モデルである BERT に注目する。文脈をもとに単語の分散表現を獲得することで、観点を的確に抽出することが可能になると考える。本研究では、日本語のデータセットを用いて、既存の SVM を用いた分類モデルに BERT を組み込み、文脈を考慮した観点抽出の有効性を検討する。

2 Aspect-Based Sentiment Analysis

2.0.1 定義

Aspect-Based Sentiment Analysis とは、感情分析手法の一つである。与えられたテキストから、その文章が何について述べているのか、どのような評価を受けているのかを Aspect 情報として抽出する。抽出された Aspect は、カテゴリや Aspect の対象語彙として文字列のラベルで表現する。それらのカテゴリ・対象語彙を用いて極性判定を行うことで、与えられたテキストを多角的に分析することができる。

2.0.2 主なタスク

国際会議である SemEval では、ABSA に関する 2 つのサブタスクが研究されている [1]. 事前にアノテーションされたデータを用いて学習を行う In-domain ABSA とトレーニングデータが存在しない Out-of-domain ABSA である. またどちらのタスクにも以下の 3 つの slot が定められており, SemEval 参加チームはサブタスクと slot を選び研究を行っている.

slot1 : Aspect カテゴリの抽出

テキストを入力し，そこからテキストが何について述べているのかを Aspect カテゴリとして抽出する（図 1）

slot2: Aspect カテゴリの対象語彙の抽出

各文に含まれる Aspect カテゴリ の対象になるもの、

また対象のテキスト中での位置を抽出する。単語またはフレーズと位置情報が抽出される。(図2)

slot3：極性判定

テキストと Aspect Category, 対象語彙を入力し, positive, negative, neutral のいずれかの極性を付与する

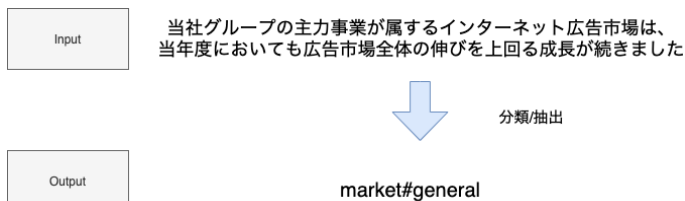


図 1 Aspect Category の抽出

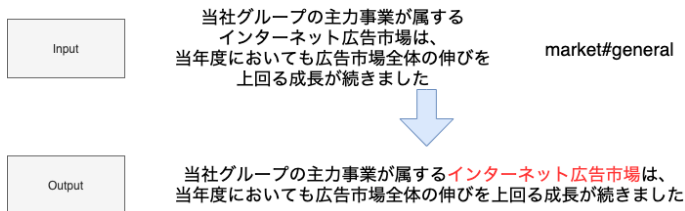


図 2 単語フレーズの抽出

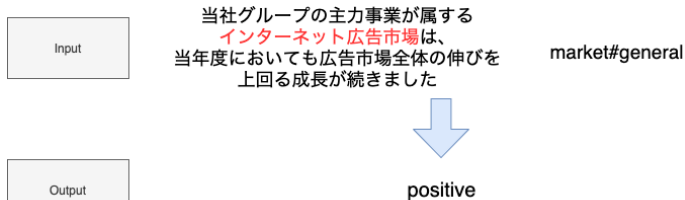


図3 極性の推定

3 BERT

BERT とは Bidirectional Encoder Representations from Transformers の略で, Google AI Language によって発表された言語処理モデルである [2].

3.1 BERT の事前学習

BERT は双方向から入力文を学習するために MLM, NSP の 2 つの事前学習を行う.

Masked Language Modeling (MLM)

入力文の一部をマスクトークンで隠し、元の文を推定するタスク

Next Sentence Prediction (NSP)

二つの文章を選び、連続している文かどうかを推定するタスク

1) <http://www.trueroad.jp>

2) ○○○○○○○○○○○○○○○○○○○○○○○○○○○○○○○
○○○○○○○. Foo bar baz boo bar baz Laboratories, Foo
Corporation.

これらの事前学習やファインチューニングを行うことで、様々な自然言語タスクにおいて精度の向上を達成している [3]。これは ABSA も例外ではなく、SemEval の各タスクで BERT を用いて実験を行い、高い成果を出している研究もある [4]。日本語への応用については、京都大学の黒橋研究室が日本語用に pretraining しているモデルがある [5]。本研究ではこの日本語 BERT を用いて各 slot の実験を行なう。

4 データセット

本実験で使用するデータセットは TIS 株式会社が感情解析を行うために開発し公開している「chABSA-dataset」を用いる [6]。chABSA-dataset は上場企業の有価証券報告書をベースに作られたデータセットであり、各テキストに対して、Aspect カテゴリ、カテゴリの対象語彙、極性などが付与されている。データセット数は約 2800 データあり、実験で用いるカテゴリは 19 種類定義されている。以下の図 4 に Aspect カテゴリを示す。

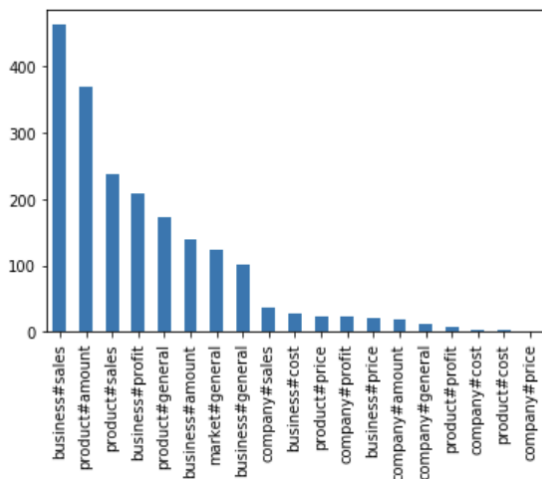


図 4 Aspect カテゴリ

slot2,3 ではアスペクトカテゴリを入力として使用するので、BERT モデル用に日本語への翻訳作業を行なった。以下のようにアトリビュート、エンティティを日本語に翻訳したうえでアスペクトカテゴリを日本語の組み合わせで表現する

エンティティ

- market : 市場
- company : 会議, グループ
- business : 部門
- product : 製品

アトリビュート

- sales : 売上
- profit : 利益
- amount : 販売数量
- price : 販売単価
- cost : 原価

5 先行研究の予備実験

先行研究として chABSA-dataset に SemEval の slot 1,2,3 を適用する実験 [6] があるので、各 slot が日本語ではどのように動作するのか検証を行い、精度の確認をするために予備実験を行なった。実験では SVM を用いて slot の適用、ラベルの予測を行なった。

5.1 予備実験の結果

結果を以下の表 1 に示す。表を見ると、特に slot2 の対象語彙・単語の位置情報の抽出に関しては他の slot に比べて精度が落ちている。これは、一つのテキストに含まれる対象語彙の数が多く、slot1 などのカテゴリの分類に比べて難易度が高いことが原因だと考えられる。前述した通り、BERT を利用して ABSA の精度改善を達成している研究もある。したがって本研究は、chABSA-dataset に BERT を組み込み各 slot の精度向上を行う。

表 1 各 slot の精度

	SVM
slot1	58.127
slot2	14.409
slot3	74.651

6 実験

6.1 実験の流れ

chABSA-dataset を用いた各 slot の実験に対し、BERT を適用することで精度向上を目指す。各テキストを単語に分割し、ID 化を行なった上で BERT に入力し性能評価を行う。

7 BERT モデルの説明

以下の図 5 に本研究で用いる BERT モデルの概要図を示し、各層の説明を行う。

文章の ID 群

テキストを単語ごとに分割し、ID 化をしたものをモデルへの入力とする。

Embedding 層

単語の ID を特徴ベクトルに変換

日本語 BERT エンコーダ

BERT モデルを使用して単語の分散表現の取得を行う。

ラベル分類器

Aspect カテゴリなどのラベルの予測を行う。

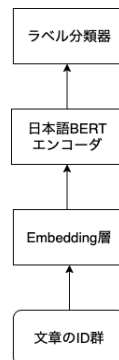


図 5 BERT モデルの概要

7.1 実験結果

各 slot に BERT を適用し、chABSA-dataset を用いて semEval の slot 1,2,3 を適用する実験をベースラインモデルとし、比較した結果を以下の表 2 に示す。比較すると slot1,slot3 に関しては精度の低下が見受けられるが、slot2 に関しては微小ながら精度向上を達成した。しかしながら、slot2 の精度は他の二つの slot に比べてかなり低い結果となった。これは、slot2 の実験が与えられた文章中からアスペクトカテゴリの対象語彙を抽出するため、他の二つの slot に比べて分類する対象がかなり多くなることが原因の一つである。

	chABSA	BERT
slot1	58.12	46.86
slot2	14.40	27.01
slot3	74.65	73.25

表 2 BERT モデルとの比較

7.2 slot2 の詳細比較

slot2 においてそれぞれのエンティティごとに精度の比較をした結果を以下の表 3 に示す。結果を見ると、エンティティの「market」、「product」では精度が著しく低いことがわかる。

ここで、「market」に関して、正しく対象語彙を予測できた例と間違えて予測した例を以下の図 6 に、「product」に関して、正しく対象語彙を予測できた例と間違えて予測した例を以下の図 7 に示す。図中の target_num は、sentence に含まれるアスペクトカテゴリの対象語彙の数である。それぞれ、target_num や Category で違いはあるが、はっきりとした大きな違いはないことがわかる。

	company	business	market	product
accuracy	0.39	0.37	0.06	0.12

表 3 slot2 の Entity ごとの accuracy

sentence
当業界におきましては、砂糖の国内消費が 8 年ぶりに前年比増加したものの、粗糖相場の上昇と円安の進行により原料コストが上昇しており、消費者の節約志向も勘案すると今後の国内消費の動向には引き続き留意すべき状況にあります

Category	True	Predicted	Target_num
market#general	国内消費	国内消費	3

sentence
当社グループの主力事業分野であります自動車業界においては、米国及びアセアン市場はほぼ横ばいでありましたが、中国並びに欧州市場が堅調に推移したことにより、日系自動車メーカーの海外生産全体は増加いたしました

Category	True	Predicted	Target_num
market#general	米国	欧州市場	5

図 6 market

sentence

プロテクトフィルムについては、台湾子会社で量産体制を整え、国内分も含め生産数量としては前年を上回ったものの、業界における販売単価下げの影響を大きく受けたことで売上が伸び悩み、事業全体としては売上微減の展開となりました

Category	True	Predicted	Target_num
product#price	プロテクトフィルム	プロテクトフィルム	2

sentence

都市事業セグメントにおいて投資家向けのビル等売却収益が減少したこと等により減収となったものの、住宅事業セグメントにおいて分譲マンションが増益となったこと、仲介事業セグメントやウェルネス事業セグメントにおける都市型ホテルの東急ステイが好調に推移したこと等により増益となりました

Category	True	Predicted	Target_num
product#general	分譲マンション	ビル等売却収益	3

図 7 product

8 まとめ

本論文では、chABSA-dataset を用いた slot1,2,3 の実験に BERT を組み込み、文脈を考慮した分散表現を用いることで各 slot の精度向上を目指す実験を行なった。その結果として、slot2 の文章中からアスペクトカテゴリの対象語彙の抽出精度を向上することはできたが、他 slot に比べると十分に予測できないことが判明した。その理由としては、対象語彙の数の多さもあるが、適切に入力文の特徴を捉えきれていないことが考えられる。アスペクトカテゴリの種類を対象語彙に合わせてさらに増やすことや、対象語彙の特徴を反映させたアスペクトカテゴリを構築することでさらに精度向上を達成することが可能と思われる。今後、より大きなデータセットを用いた実験を行い、入力文や対象語彙の特徴を適切に表現できるアスペクトカテゴリの構築、比較を行う予定である。

参考文献

- [1] Maria Pontiki, Dimitrios Galanis, Harris Papageorgiou, Suresh Manandhar, and Ion Androutsopoulos. SemEval-2015 Task 12: Aspect Based Sentiment Analysis. In *Proceedings of the 9th International Workshop on Semantic Evaluation*, 2015.
- [2] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Google AI Language 2018.
- [3] 曹 銳, 田中 裕隆, 白 静, 馬 ブン, 新納 浩幸. BERT を利用した教師あり学習による語義曖昧性解消. 言語資源活用ワークショップ発表論文集. 2019.
- [4] Chi Sun, Luyao Huang, and Xipeng Qiu. Utilizing bert for aspect-based sentiment analysis via constructing auxiliary sentence. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 380–385, 2019.
- [5] 柴田 知秀, 河原 大輔, 黒橋 禎夫. BERT による日本語構文解析の精度向上. 言語処理学会, 第 25 回年次大会, 発表論文集, 2019.
- [6] Takahiro Kubo and Hiroki Nakayama. chABSA: Aspect Based Sentiment Analysis dataset in Japanese. Strategic Technology Center, TIS, Inc., 2018.
<https://github.com/chakki-works/chABSA-dataset>