

Neural Attention と RNN を利用した観点付き感情極性推定 Aspect-Based Sentiment Analysis Using Neural Attention and RNN

谷 健太郎[†] 新妻 弘崇[†] 太田 学[†]
Kentaro Tani Hirotaka Niitsuma Manabu Ohta

1. はじめに

文章には書き手の感情が含まれており、その観点を予測し、観点に対する感情極性を推定することは観点付き感情極性推定(ABSA: Aspect-Based Sentiment Analysis)と呼ばれる文書分析のタスクの一つである。この ABSA に対して Neural Network を利用する研究が多く行われている。

本稿では、Word Embedding により単語をベクトルに変換した後、RNN の一種である LSTM と Neural Attention を組み合わせたモデルを利用し、英語のレストランレビューの観点付き感情極性を推定する。また、学習の際に観点 (Aspect Category)、感情極性(Polarity)に加え、意見の対象となる表現であるターゲット表現(OTE: Opinion Target Expression)も利用する。

2. 研究背景

昨今、文書分類では離散的である単語を Word Embedding で連続的なベクトルに変換し、Neural Network を適用する研究が多く進められており、その代表的な手法としては、CNN (Convolutional Neural Network)や RNN (Recurrent Neural Network)などを組み合わせたモデルがある。これらはタスクやモデルによってうまく使い分けると、良い結果が得られることが多い。特に RNN の一種である LSTM (Long short-term memory)は内部のループで単純な RNN よりも長期の状態記憶ができるため、文から文脈などの前後情報を抽出できると考えられている。

文章分類における感情極性推定は自然言語処理における基本的なタスクで、文章に対して positive, negative, neutral の感情極性ラベルを付与するものである。しかし文章の感情極性推定だけでは不十分な場合がある。それは文章には一般に複数の異なる観点についての感情極性が含まれるからである。観点付き感情極性推定はこれに対処するためのフレームワークである。

我々の研究[1]では、朝倉らの観点付き感情極性推定モデル[2]で LSTM を用いた変換より恒等変換を用いたモデルが良い結果であったことから LSTM を用いながったが、朝倉らのモデルを修正、拡張することで[2]を上回る結果を確認した。

本稿では、LSTM を[1]での提案モデルに用い、さらに[2]とは異なる構成で LSTM を利用したモデルでも比較実験を行った。

3. 関連研究

文書分類において Convolutional Neural Network (CNN)は最も成功した手法の一つである。

Collobert ら[3]は単語ベクトル列である文章に対する

[†]岡山大学大学院自然科学研究科 Graduate School of Natural Science and Technology, Okayama University

Sequential Convolution が多くの文書分類のベンチマークで高い精度を達成したことを報告し、Kim[4]は Collobert らの手法を拡張することで同様の様々なベンチマークで既存手法を上回る精度を達成した。

ABSA では、Wang ら[5]と Yanase ら[6]が Neural Attention を感情極性抽出に利用した。これらはニューラル機械翻訳での利用法が元になっている。ただし、目的関数を含む Neural Attention の利用法、Aspect Category のベクトル表現の方法がニューラル機械翻訳と異なり、特に Yanase らは Aspect Category が Entity#Attribute で表されることから、Entity, Attribute をそれぞれ別のベクトルとして表現し、選ばれたベクトルの期待値もそれぞれ別に計算した。そしてそれらを連結し、感情極性を予測した。

Wang ら、Yanase らはともに単語ベクトルを LSTM でエンコードしたが、朝倉ら[2]は ABSA では局所領域への注目がより重要になっていると考え、Sequential Convolution を用いている。また Yanase らは Yelp Academic Dataset による事前学習を行ったが、朝倉らの研究[2]は同データセットで半教師あり学習を行うために Yelp Academic Dataset を教師なしデータとして用いている。

4. データセット

SemEval-2016 Task 5 Subtask 1 (SE16T5S1)は、英語の飲食店や商品レビューのデータセットの ABSA を行うタスクである。本研究ではこのうち飲食店のレビューを扱うレストランドメインのデータセットを用いる。以下はこれらのデータセットの例文の 1 つである。

"The sushi is amazing!!!"

```
<Opinion category="FOOD#QUALITY"
polarity="positive"
target="sushi" from="4" to="9" />
```

<Opinion>内に付与されているラベルが記述されており、Aspect Category は category、感情極性は polarity の属性に対応している。OTE は target, from, to で表される。1 レビュー内に複数の <Opinion> が存在する場合もある。category は Entity#Attribute の形で表現され、全部で 13 種が定義されている。polarity は positive, negative, neutral の 3 種類がある。target が OTE の単語もしくは熟語の文字列、from, to は OTE の開始と終了の文字位置を表す。

SE16T5S1 は以下の Slot1, Slot2, Slot3 のサブタスクから構成される。

Slot1 文中に含まれる 0 個以上の Aspect Category を抽出する。抽出の精度は F 値で評価する

Slot2 文中に含まれる OTE を抽出する、抽出の精度は F 値で評価する。OTE は単語かフレーズである。

Slot3 与えられたレビュー文と Aspect Category のペアに対して positive, negative, neutral のいずれかの感情極性を付与する。結果はテストデータ中のすべてのペアに対して正しく付与できた割合で評価する。本稿では、Slot3 に取り組む。

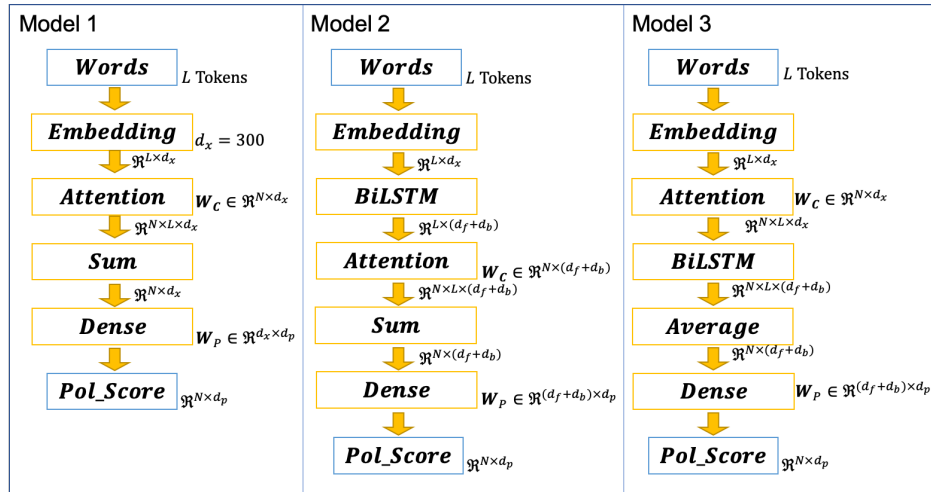


図 1 提案モデル

5. 感情極性推定モデル

朝倉ら[2]は Word Embedding により単語をベクトルに変換した後、Neural Attention を用いて観点付き感情極性推定モデルを構築した。最初に行う Word Embedding では単語に対して対応するベクトルを学習で最適化し用いる手法とそのベクトルを LSTM や CNN などに入力し変換したベクトルを用いる手法を比較した。しかし、朝倉らの実験では、文脈情報などを加味できる LSTM などの手法より単語のベクトルをそのまま恒等変換で用いた手法の方が良い結果を示した。そのため我々の研究[1]では、朝倉らの結果で最も良い結果であった恒等変換のみを実験で評価した。

本稿では、文脈情報を加味したベクトルを生成するために LSTM により[1]で提案したモデルを拡張する。Neural Attention と LSTM を組み合わせたモデルと Neural Attention のみを用いたモデルを実験により比較する。

5.1 RNN

RNN(Recurrent Neural Network)は、再帰的にニューラルネットワークを構成するモデルである。それにより入力データの情報をベクトルして再帰的に入力することで蓄積し、それを加味した学習を行えることから時系列データに適したモデルになっている。RNN は様々な拡張がされており、いくつかの派生したモデルが考案されている。基本的な RNN に加え、RNN を拡張した LSTM, GRU, 入力を逆順にした RNN と組み合わせた双方向 RNN などがある。

LSTM は長期的なデータの保持を目的としたモデルである。内部ループを持つことで再帰的な入出力とは別に情報をベクトルに抽出し保持することができ、これにより入力の前後情報(文脈情報)を学習できる。

本稿では、双方向 RNN を LSTM で構成した BiLSTM を用いる。

5.2 Neural Attention

Neural Attention は、入力の各ベクトルに対して注目度を計算しそれを入力ベクトルの係数として利用する手法である。本稿では Attention 層で各 Aspect Category(観点)における注目度を計算し、それを用いて入力ベクトルを重み付けする。

6. 提案モデル

本稿で提案するモデルは、朝倉らのモデル[2]を修正し、拡張したモデル[1]に BiLSTM を加えてさらに拡張したモデルである。図 1 に本稿で比較するモデルを示す。Model 1 は[1]と同じモデルで朝倉らのモデル[2]に活性化関数や目的関数を変更したものである。Model 2 は Model 1 で Neural Attention を導入する前に BiLSTM を用いるモデルである。Model 3 は Neural Attention を導入した後に BiLSTM を用いるモデルである。この 3 つのモデルで後述の目的関数のパラメータと LSTM の出力次元数 `rn_dim` をいくつかの組み合わせで実験した。

6.1 層(レイヤ)の説明

図 1 は、モデルを層(レイヤ)ごとに示したものである。それぞれの層について説明する。

Embedding 層では、長さ L の入力の文を計算で処理できる形にするためトークンに分割し、トークン列を d_x 次元のベクトル列 $\mathbb{R}^{L \times d_x}$ に変換する。

BiLSTM 層では、入力ベクトル列 (Model 2: $\mathbb{R}^{L \times d_x}$, Model 3: $\mathbb{R}^{N \times L \times d_x}$) に対して双方向 LSTM を利用し、それぞれの結果を結合し出力ベクトル列 (Model 2: $\mathbb{R}^{L \times (d_f + d_b)}$, Model 3: $\mathbb{R}^{N \times L \times (d_f + d_b)}$) とする。出力ベクトルの次元数は順方向 LSTM の d_f 次元と逆方向 LSTM の d_b 次元の和となる。本稿の実験では、 $d_f = d_b = \text{rn_dim}$ とする。

Attention 層では、入力ベクトル列 (Model 2: $\mathbb{R}^{L \times (d_f + d_b)}$, Model 1, 3: $\mathbb{R}^{L \times d_x}$) に対して重み付けを行うレイヤである。観点を表す N 通りの Aspect Category で重みの計算方法を用意するために N 行の行列 W_c と転置した入力ベクトル列の行列積を活性化関数の sigmoid 関数で 0~1 に射影したものを注目度 $Alignment \in \mathbb{R}^{N \times L}$ として用いて入力の各ベクトルに掛けることで N 通りの重み付きのベクトル列 (Model 1, 2: $\mathbb{R}^{N \times L \times (d_f + d_b)}$, Model 3: $\mathbb{R}^{N \times L \times d_x}$) とする。

Sum 層では、入力ベクトルを足し合わせる。図 1 の Attention 層と合わせて、入力ベクトルに対し Aspect Category ごとに重み付き和ベクトルを計算する。

Average 層は入力ベクトル列の平均ベクトルを計算する。

Dense 層は、入力ベクトル列 (Model 1: $\mathbb{R}^{N \times d_x}$, Model 2, 3: $\mathbb{R}^{N \times (d_f + d_b)}$) に対して重み W_p を掛けバイアスを足し、全結合を行った後、活性化関数を用いる。活性化関数には

softmax 関数を利用し出力を確率とする。出力ベクトルの次元を感情極性の数 d_p に合わせ、それぞれの Aspect Category における感情極性の確率を出力 $Pol_Score \in \mathbb{R}^{N \times d_p}$ とする。

推定は、推定対象の Aspect Category の id を i としたとき、入力文の正解ラベル $Pol_label \in \{0,1\}^{d_p}$ と推定した結果の $Pol_Score[i,:]$ を最大値のみ 1 それ以外 0 としたものと比較し F 値で計算する。ここで行列の要素の抽出は Numpy の表記法で示した。

次にモデルの変数等を最適化するための目的関数について述べる。

6.2 目的関数

学習によって最適化する目的関数は、それぞれのモデルに対して以下の式(1), (2), (3)を足し合わせたものを用いる。ここで推定対象の Aspect Category の id を i とし、行列の要素の抽出には Numpy の表記法を利用した。

$$Loss^P = CE(Pol_Score[i:], Pol_label) \quad (1)$$

$$Loss^C = CE(\widehat{Alignment}, Cat_label) \quad (2)$$

$$Loss^T = CE(\widehat{Alignment}[i:], Ote_label) \quad (3)$$

$$\widehat{Alignment} = \text{softmax} \left(\sum_{k=0}^{l-1} (\widehat{Alignment}[i:,k]) \right) \quad (4)$$

$Ote_label \in \{0,1\}^l$ は、入力文の OTE からトークンごとに OTE であるかを 2 値のラベルで表したもので、 $Cat_label \in \{0,1\}^N$ は入力文に付与されている Aspect Category を 2 値のラベルで表したものである。 $\widehat{Alignment} \in \mathbb{R}^N$ は、 $\widehat{Alignment}$ を Aspect Category ごとに足し合わせたものを softmax 関数で計算し確率とする。 $CE(A, B)$ は A と B のクロスエントロピーを表す。

全体の目的関数: $Loss$ は、 $Loss^P$ に $\alpha Loss^C$, $\beta Loss^T$ を足し合わせたものである。実験では係数 α , β のいくつかの組み合わせで比較する。

$$Loss = Loss^P + \alpha Loss^C + \beta Loss^T \quad (5)$$

7. 実験

前述の SE16T5S1 のデータセットを用いて Model 1, 2, 3 のパラメータを変えて、4 節で述べた Slot3 の実験をする。

実験は最大 epoch を 100, 試行回数 5 回とし、レビュー文のトークン化には NLTK [7] を用いる。トークン化後、記号を表すタグが付与されたトークンは削除し、その他の英字については、大文字を小文字へ変換した。また、Embedding 層で生成する単語ベクトルの初期値は Google News Corpus の 300 次元の単語ベクトルを用い、学習中にさらに最適化されるようにした。また LSTM の出力は順方向と逆方向を同じ rnn_dim としていくつかの値で比較する。モデルの出力ベクトルの次元である感情極性数 d_p は 3 で positive, negative, neutral を表す。

7.1 実験結果

6 節で述べた各モデルによる推定結果を F 値で評価する。本稿の実験では正解ラベル Pol_label も推定結果も one-hot

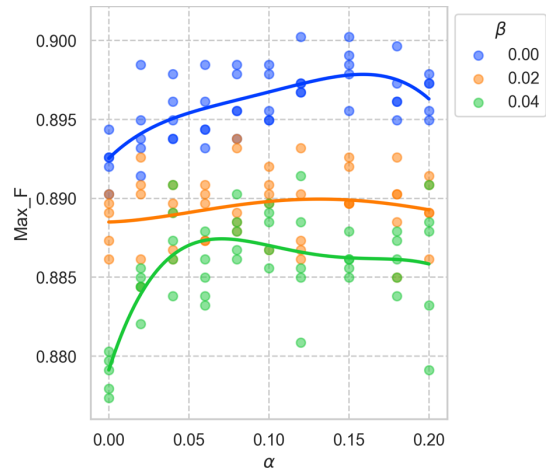


図 2 F 値の最大値の分布 (Model 1)

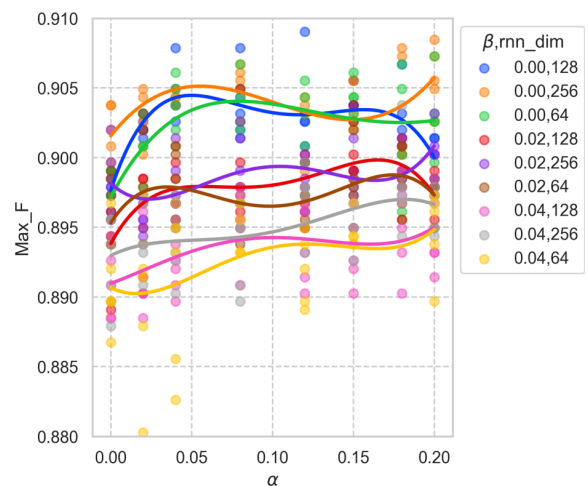


図 3 F 値の最大値の分布 (Model 2)

であるため、再現率と適合率(精度), 2 つの調和平均である F 値は同じ値になることから、以下は F 値で評価する。

7.1.1 F 値の最大値の分布

各モデルでパラメータによる影響を見るために 100epoch, 5 回の実行で実行ごとの F 値の最大値の分布と β , rnn_dim ごとの 4 次の回帰曲線を図 2, 3, 4 に示す。

図 2 は Model 1 における F 値を示したものである。この図から Model 1 では $\alpha = 0.15$, $\beta = 0.00$, のときに良い結果になることがわかる。

図 3 は Model 2 における F 値を示したものである。この図から Model 2 では $\beta = 0.00$, $rnn_dim = 256$ で $\alpha = 0.04 \sim 0.08$, 0.2 のとき結果が良いことがわかる。

図 4 は Model 3 における F 値を図に示したものである。この図ではパラメータごとに回帰曲線の傾向も様々で、値のばらつきも大きいことがわかる。

図 2, 3, 4 から Model 1, 2 において、 α を適当な値で用いた場合、 α を 0 としたときと比べて約 0.5 ポイント F 値が向上することを確認できる。ただし、Model 3 に関してはパラメータの影響が複雑で比較的 F 値も低いいため、推定に有効でないことがわかる。

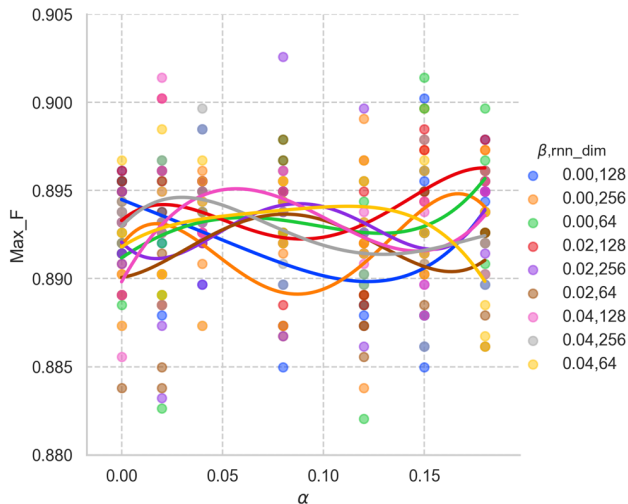


図 4 F 値の最大値の分布 (Model 3)

これらの結果から目的関数等のパラメータを次のように設定して比較する。Model 1, 2 は結果が良かったもの、Model 3 は比較的值のばらつきが少ない $\alpha = 0.04$ で、このとき高次回帰の結果が良い $\beta = 0.04$, $rnn_dim = 128$ を利用する。

Model 1 : $\alpha = 0.15$, $\beta = 0.0$

Model 2 : $\alpha = 0.04$, $\beta = 0.0$, $rnn_dim = 256$

Model 3 : $\alpha = 0.04$, $\beta = 0.04$, $rnn_dim = 128$

7.1.2 Epoch ごとの F 値の推移

各モデル、パラメータにおける Epoch ごとの F 値の推移を図 5 に示す。図の実線が 10 回の中央値、色付き領域が最大最小の範囲である。図から F 値が最大になるまで 15epoch 程度でモデルごとに大きな差がないことがわかる。

7.1.3 各モデルの F 値の最大値の比較

各モデルにおける前述のパラメータでの F 値の最大値の分布の箱ひげ図を図 6 に示す。Model attn は Neural Attention のみを利用したモデル (Model 1, $\alpha = 0$, $\beta = 0$) である。Attention 層の次に BiLSTM 層を用いた Model 3 よりも Attention 層のみの Model attn の方がわずかに F 値が高い。 α , β を適切な値としたとき BiLSTM 層の次に Attention 層を通す Model 2 が Model 1 よりも F 値が高いことが確認できた。

8. おわりに

本稿では、[1]で利用しなかった RNN を利用し、モデルの推定への影響を調査した。朝倉らの先行研究[2]では有効ではなかった LSTM だが改良を加えたことで本稿では有効に働くことを確認した。

また、LSTM の前に Neural Attention を導入した Model 3 は、LSTM を用いない場合と比べ F 値が低下することを確認した。

Model 1, 2 では目的関数に対して各 Aspect Category の Neural Attention の注目度を学習に利用した結果、適当な係数 α で学習することで、 $\alpha = 0$ の通常の学習を行った場合と比べ 0.5 ポイントほど F 値が向上することを確認した。しかし OTE を学習に利用する場合は係数 β を大きくするほど F 値が低下し有効でなかったため、今後の課題として OTE の活用法を検討していきたい。

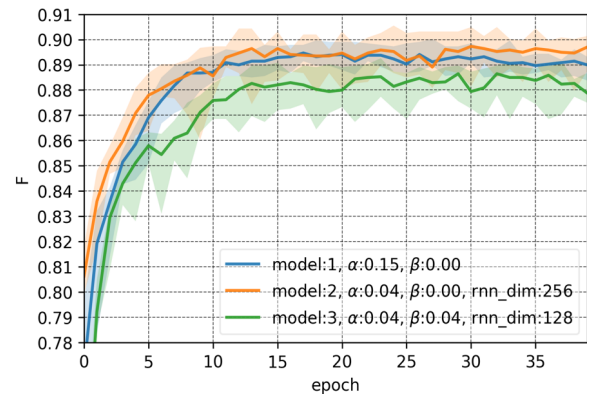


図 5 epoch 数と F 値

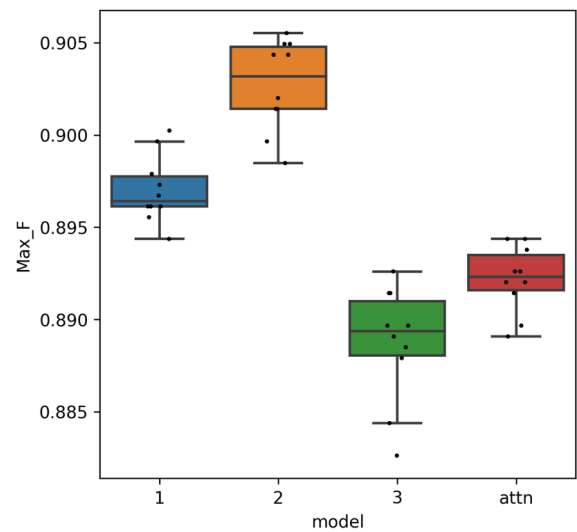


図 6 F 値の最大値の箱ひげ図

参考文献

- [1] 谷健太郎, 新妻弘崇, 太田学. “ターゲット表現と Neural Attention を利用した観点付き感情極性推定”. 第 11 回データ工学と情報マネジメントに関するフォーラム (DEIM2019), C4-3, 2019.
- [2] 朝倉遼, 新妻弘崇, 太田学. “Aspect-Based Sentiment Analysis における Neural Attention の効率的な利用方法”. 第 10 回データ工学と情報マネジメントに関するフォーラム (DEIM2018), G4-4, 2018.
- [3] Ronan Collobert, Jason Weston, Leon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. Natural Language Processing (Almost) from Scratch. J. Mach. Learn. Res, Vol. 12, pp. 2493–2537, 2011.
- [4] Yoon Kim. Convolutional neural networks for sentence classification. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, pp. 1746–1751, 2014.
- [5] Yequan Wang, Minlie Huang, Xiaoyan Zhu, and Li Zhao. Attention-based LSTM for Aspect-level Sentiment Classification. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pp. 606–615, 2016.
- [6] Toshihiko Yanase, Kohsuke Yanai, Misa Sato, Toshinori Miyoshi, and Yoshiki Niwa. bunji at SemEval-2016 Task 5: Neural and Syntactic Models of Entity-Attribute Relationship for Aspect-based Sentiment Analysis. In Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), pp. 289–295, 2016.
- [7] Edward Loper and Steven Bird. Nltk the natural language toolkit. Association for Computational Linguistics, Vol. 1, pp. 63–70, 2002.