

# SE-ResNet Attention モデルによる観点付き感情極性推定の一手法

久保田大貴<sup>†</sup> 新妻 弘崇<sup>††</sup> 太田 学<sup>†††</sup>

<sup>†</sup>, <sup>†††</sup> 岡山大学大学院自然科学研究科 〒700-8530 岡山県岡山市北区津島中三丁目1番1号

<sup>††</sup> 大阪大学データビリティフロンティア機構 〒565-0871 大阪府吹田市山田丘2番8号

E-mail: <sup>†</sup>kubotad@de.cs.okayama-u.ac.jp, <sup>††</sup>niitsuma@ids.osaka-u.ac.jp, <sup>†††</sup>tohta@cs.okayama-u.ac.jp

あらまし 感情極性とは文章に含まれる意見が好意的か否定的かを表す情報である。これは意見要約や可視化するために重要である。中でも Aspect-Based Sentiment Analysis (ABSA) は文章中に含まれる、観点に対する感情極性を予測するタスクである。本稿では、画像認識分野で高精度を達成した Squeeze-and-Excitation Residual Network (SE-ResNet) に Attention 機構を組み合わせたモデルを提案し、文章の感情極性推定タスクに適用して推定精度を評価する。

キーワード ニューラルネットワーク, 感情極性, 深層学習, SENet, ResNet, SE-ResNet

## 1 はじめに

感情極性推定とは感情極性ラベルが付与された文章を学習して未知の文章の感情極性を推定する自然言語処理のタスクである。感情極性はその文章が好意的か否定的かを表す情報であり、positive, negative などのラベルで表される。これらの感情極性は文章全体の要約に役立ち、意見の収集の効率化や可視化などに用いられている。近年ではニューラルネットワークでこれらを推定する研究が活発に行われている。

一方、実際の文章には複数の観点と相違なる感情極性が含まれるものもある。そのような文章を、文章全体の感情極性で表すのは不十分な場合がある。そこでこの問題に対応したのが、Aspect-Based Sentiment Analysis (ABSA) [4] と呼ばれるタスクである。ABSA は文章から観点を抽出し、それぞれの観点に対する感情極性を推定する。具体的には、観点毎の感情極性ラベルが与えられた教師データを学習することで推定する。このタスクは日本語では観点付き感情極性推定と呼ばれている。Niitsuma らは単純な Attention モデルである Simple Alignment Sentence Classification (SASC) を提案し、ABSA において、従来手法の Convolutional Neural Network (CNN) や Recurrent Neural Network (RNN) を分類精度、実行時間共に上回ったことを報告した [3]。朝倉らは Ladder Network [2] を拡張し、半教師あり学習を ABSA に適用する手法を提案し、更に高い推定精度を達成した [1]。

本稿では観点付き感情極性推定のため、画像認識の分野で高精度を達成した Residual Network と Squeeze-and-Excitation Network にカテゴリの Neural Attention を組み合わせたモデルを提案する。実験では SemEval2016 Task5 Subtask1(SE16T5S1)<sup>1</sup>の感情極性推定タスクにおける正解率で評価する。

## 2 関連研究

### 2.1 Neural Attention

Neural Attention は考慮すべき情報に高い重みを与えるニューラルネットワークである。Bahdanau ら [6] は翻訳モデルに Neural Attention を組み込むことで長文の入力に対しても精度の高い翻訳ができることを報告した。翻訳モデルは入力の特徴量に変換する Encoder と特徴量から出力を得る Decoder で構成されている。提案モデルは、重みとして入力文章全体の情報を Decoder に渡すが、その際 Encoder が現在処理している入力文章のスコアを大きくしてから用いる。Decoder は出力の際、過去の入出力と受け取った重みから翻訳単語を決定する。その結果、入力文章全体についての情報と現在処理している単語情報を持つため、長文であっても翻訳の精度を保つことが可能となった。観点付き感情極性推定では、柳瀬ら [7] は双方向 RNN から得られるベクトルに対して 2 つの Attention 層によりエンティティと属性について考慮するモデルを提案した。彼らは SemEval 2015 Task 12<sup>2</sup>のデータセットを用いて観点付き感情極性推定を行い、SemEval 2015 Task 12 に参加した最上位チームである NLANGP の手法 [8] 以上の精度を達成したことを報告した。

### 2.2 Residual Network

Residual Network (ResNet) とは、He ら [11] が提案し ILSVRC2015<sup>3</sup>の画像分類タスクで優勝したモデルである。ニューラルネットワークには、層を深くするにつれて認識精度が向上する傾向があるが、層が深いほど勾配を維持することができず消失してしまうため、学習が困難であった。しかし、ResNet では Residual module と呼ばれるショートカットを含んだ構造を利用することで、層が深くなっても勾配を次の層に伝えることを可能にしている。Residual module の内

<sup>1</sup> : <http://alt.qcri.org/semeval2016/task5/>

<sup>2</sup> : <http://www.alt.qcri.org/semeval2015/task12/>

<sup>3</sup> : <http://image-net.org/challenges/LSVRC/2015/>

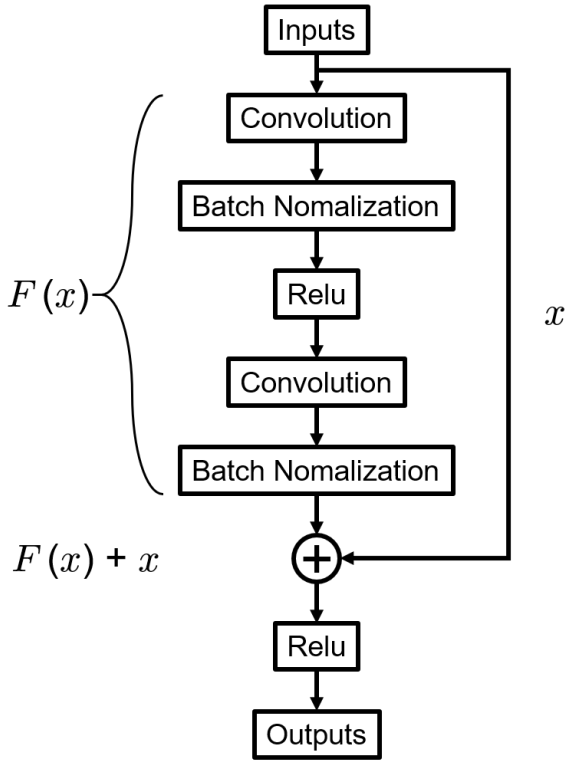


図 1 Residual module の内部構造 [11]

部構造を図 1 に示す。入力  $x$  に対する、Residual module による変換関数  $H(x)$  は、内部の層の変換関数  $F(x)$  にショートカットを通った入力を加えた  $H(x) = F(x) + x$  で表すことができる。He らは 16 個の Residual module を重ねることにより pooling 層と全結合層を合わせて 34 層の ResNet を提案した。また、計算コストを増やさずにより層を深くするために Bottleneck 構造も提案している。計算コストが等価である Residual module と Bottleneck 構造を比較したものを図 2 に示す。Residual module は Convolution 層が 2 層であるのに対し、Bottleneck 構造では 3 層ある。図 2 の Bottleneck 構造は 2 つの  $1 \times 1$  の Convolution が次元の縮小と復元の役割をしている。具体的には、1 目目の  $1 \times 1$  の Convolution が 256 次元を 64 次元に縮小し、2 目目で 64 次元で 256 次元に復元する。Bottleneck 構造のショートカットを projection に置き換えた場合、縮小前の入力部分と復元後の出力部分、つまり高次元の 2 か所に接続されるため、計算コストとモデルサイズが 2 倍となる。しかし、ショートカットによる計算はパラメータ数が増えないため、Bottleneck 構造を効率的なものにしている。

### 2.3 Squeeze-and-Excitation Network

Squeeze-and-Excitation Network (SENet) は ILSVRC2017<sup>4</sup> の画像分類タスクの優勝モデルで、Hu ら [12] が提案したものである。従来の手法である CNN は Convolution と活性化により、平面方向の特徴を抽出していたのに対し、SENet は各 channel の出力を重み付けすることにより、画像全体における channel 間の関係性を考慮する SE block と呼ばれる機構

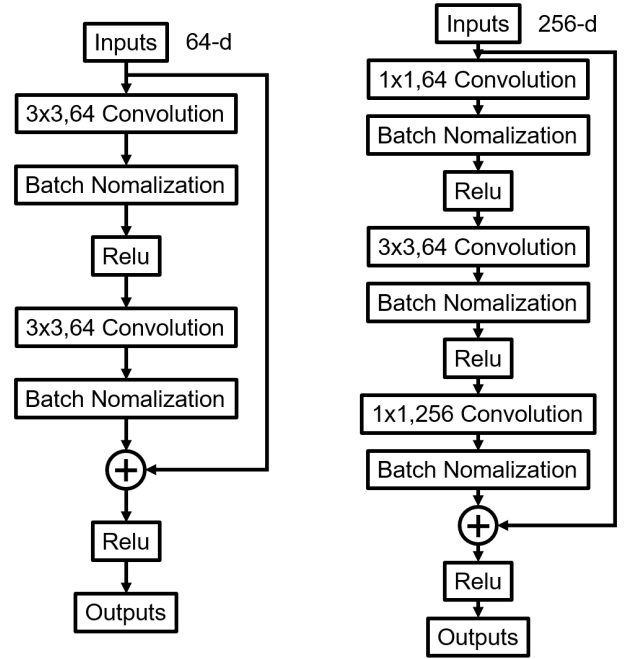


図 2 計算コストが等価な Residual module (左) と Bottleneck 構造 (右) [11]

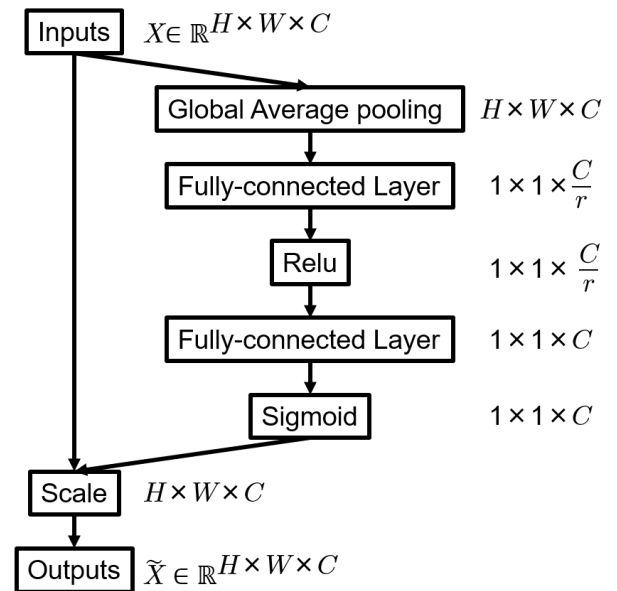


図 3 SE block の内部構造 [12]

を持つ。その構造を図 3 に示す。図 3 の  $X$  は変換する画像の channel 毎の画素値であり、画像のサイズ  $H \times W$  の数の画素値を channel 数  $C$  の数だけ含む。また  $r$  は次元の削減と増加の割合を表す Reduction ratio である。SE block は Convolution で得た特徴マップに対し、空間方向の画像全体の平均を取ることで、channel 次元のベクトルに変換する。そのベクトルを Sigmoid 関数で活性化させることにより、channel 方向に重み付けする。この機構はどのようなネットワークとも組み合わせることができ、Inception module [13] や ResNet, VGG-16 [14] に組み合わせて元のモデルと比較したところ、全てのモデルが元のモデルを上回ったことが報告されている。

4: <http://image-net.org/challenges/LSVRC/2017/>

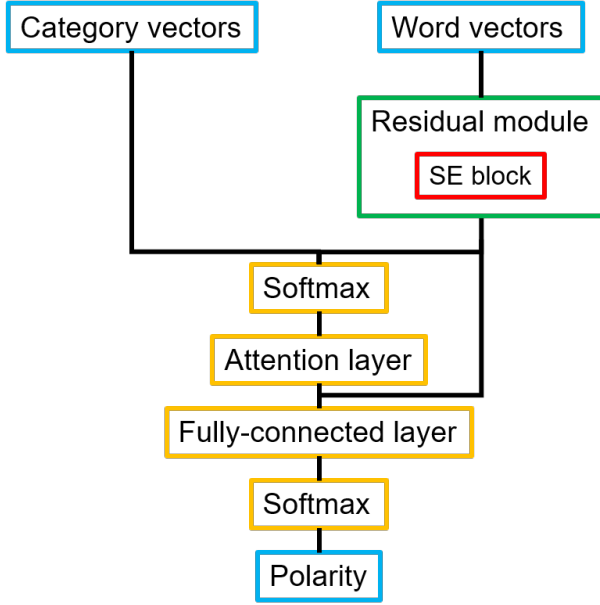


図 4 SE-ResNet Attention モデルの全体構造

### 3 提案手法

#### 3.1 SE-ResNet Attention モデル

本稿で提案する SE-ResNet Attention モデルについて説明する。このモデルは SE-ResNet にカテゴリを考慮させる Neural Attention を組み合わせたものである。モデル構造を図 4 に示す。本モデルの入力はカテゴリと文章の 2 つである。この 2 つの入力は embedding 層によってそれぞれベクトル化される。入力される文章の単語を embedding で変換した Word vector を  $X$ 、カテゴリを embedding で変換した Category vector を  $V$  とし、 $L$  は文の長さ (単語数)、 $n$  はカテゴリの種類とすると、 $X$  と  $V$  はそれぞれ以下の行列で表される。

$$X = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_L \end{bmatrix} \quad (1)$$

$$V = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \dots & \mathbf{v}_n \end{bmatrix} \quad (2)$$

$\mathbf{x}_i \in \mathbb{R}^d$ ,  $\mathbf{v}_i \in \mathbb{R}^d$  の次元数  $d$  は、embedding の際に定めるパラメータである。入力文章の単語を embedding で変換した Word vector を SE block に通し変換したベクトル  $\tilde{X}$  と Category vector  $V$  から Softmax 関数により Attention Weight  $\alpha \in \mathbb{R}^L$  を次のように計算する。この時  $\mathbf{v}_a$  は考慮させるカテゴリのベクトルである。

$$\alpha = \text{softmax}(\tilde{X}^T \mathbf{v}_a) \quad (3)$$

算出した Attention weight と Word vector から感情極性の確率分布  $\mathbf{y} \in \mathbb{R}^3$  を求めて感情極性を推定する。 $\mathbf{y}$  は positive, neutral, negative の 3 つの感情極性のどれに推定されるかの確率であり、以下の式で求められる。ここで  $P \in \mathbb{R}^{3 \times d}$  は推定する感情極性への射影行列である。

$$\mathbf{y} = \text{softmax}(P \tilde{X} \alpha) \quad (4)$$

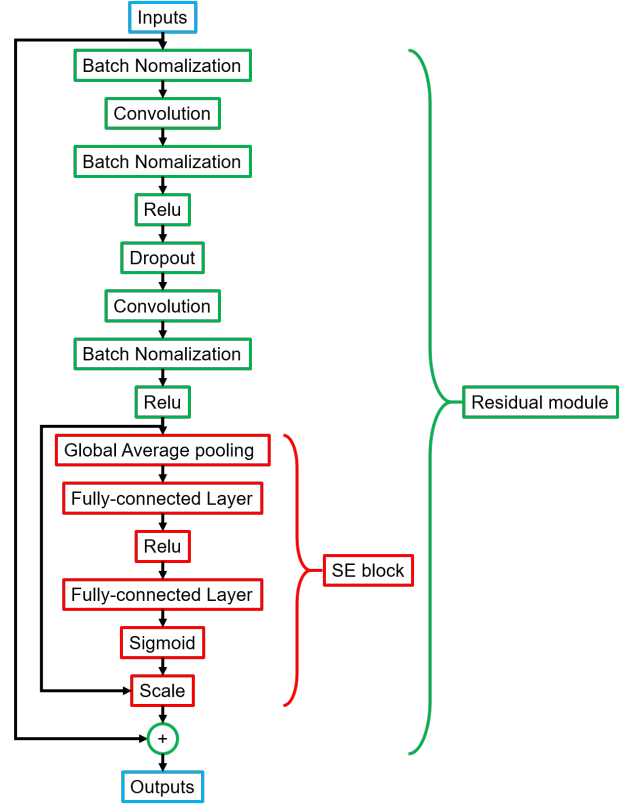


図 5 提案する Residual module の内部構造

#### 3.2 Residual module

本稿で提案する Residual module は、Hu ら [12] の提案した SE-ResNet module と同様に、Residual module に SE 機構を組み込んだものである。内部構造を図 5 に示す。Residual module 部分は Convolution, Batch Normalization(BN) [9], ReLU, Dropout からなる。具体的には BN-Conv-BN-ReLU-Dropout-Conv-BN-ReLU の順で適用する。

また、SE block は Residual module の処理のあと、何も処理をしないベクトルと Global Average Pooling, Fully-connected Layer, ReLU, Fully-connected Layer, Sigmoid の順に適用したベクトルを乗算する。その後 Residual module のショートカットを通した元の入力と加算する。

Word vector  $X$  から得られた特徴マップを  $\mathbf{U} \in \mathbb{R}^{H \times W \times C}$  とすると、Global Average pooling 層の変換  $F_{GAP}(\mathbf{x})$  で得られる  $\mathbf{z}$  は以下の式のように channel 毎に平均する。この時、 $H$ ,  $W$  は平面次元、 $C$  は channel 数を表す。

$$\mathbf{U} = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \dots & \mathbf{u}_C \end{bmatrix} \quad (5)$$

$$z_k = F_{GAP}(\mathbf{u}_k) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_k(i, j) \quad (6)$$

その後の処理は以下の式で表される。

$$\mathbf{s} = \sigma(\mathbf{W}_2 \delta(\mathbf{W}_1 \mathbf{z})) \quad (7)$$

この時、 $\sigma$  は Sigmoid 関数、 $\delta$  は ReLU 関数である。また  $\mathbf{W}_1 \in \mathbb{R}^{\frac{C}{r} \times C}$ ,  $\mathbf{W}_2 \in \mathbb{R}^{C \times \frac{C}{r}}$  は全結合層を表し、 $r$  は次元を削

減あるいは増加する割合を表す Reduction ratio である。また  $\mathbf{s} \in \mathbb{R}^C$  は各 channel の重みであり、 $\mathbf{U}$  に channel 毎に式 (9) のように乗算し、重みを付与した特徴マップ  $\tilde{\mathbf{U}} \in \mathbb{R}^{H \times W \times C}$  を算出する。

$$\tilde{\mathbf{U}} = \begin{bmatrix} \tilde{\mathbf{u}}_1 & \tilde{\mathbf{u}}_2 & \dots & \tilde{\mathbf{u}}_C \end{bmatrix} \quad (8)$$

$$\tilde{\mathbf{u}}_k = s_k \mathbf{u}_k \quad (9)$$

## 4 評価実験

評価実験として、提案モデルで英語のレストランレビュー文の観点付き感情極性推定を行い、その正解率を算出する。比較のため先行研究である Ladder Network [1], SASC [3] を用いた同様の実験も行う。

### 4.1 データセット

SemEval2016 Task5 Subtask1(SE16T5S1)<sup>5</sup>で提供されるレストランメインのデータセットを利用する。このデータセットは学習データ 2,000 件、テストデータ 676 件の英語のレストランレビュー文で構成されている。レビュー文の例を以下に示す。

```
"The wine list has many good values."
<Opinion category="DRINK#PRICES"
  polarity="positive" />
```

このように各レビュー文には category と polarity を含む Opinion タグが付与されている。本稿ではレビュー文と category から polarity を推定する。カテゴリは以下の 12 種類である。

- RESTAURANT#GENERAL
- RESTAURANT#MISCELLANEOUS
- RESTAURANT#PRICES
- SERVICE#GENERAL
- FOOD#PRICES
- FOOD#QUALITY
- FOOD#STYLE\_OPTIONS
- DRINKS#PRICES
- DRINKS#QUALITY
- DRINKS#STYLE\_OPTIONS
- AMBIENCE#GENERAL
- LOCATION#GENERAL

またこのデータセットの感情極性の分布を表 1 に示す。

### 4.2 ハイパーパラメータ

Word vector の初期値には 2016 年時点の Google News Corpus の 300 次元ベクトル<sup>6</sup>を用いる。このベクトルは学習中に最適化される。モデルは cross-entropy loss が最小になるように学習を進める。最適化には He ら [11] 同様に、Momentum 項を加

表 1 データセットの感情極性ラベル分布

データセット	Positive	Neutral	Negative
学習データ	1657	749	101
テストデータ	611	204	44

えた確率的勾配降下法を用いる。学習率は 0.005, Momentum は 0.5 とした。ハイパーパラメータである Residual module (図 5) に含まれる Convolution の filter size はどちらも 3, stride は 1, channel は 128, Dropout 率は 0.5 とし、また Reduction ratio  $r$  は Hu ら [12] の実験で最も Top1 error 率が小さい 4 とした。学習回数である epoch 数は 60 までとした。

### 4.3 実験結果

実験結果を表 2 に示す。提案モデルの正解率は 0.861 となり、先行研究のうち最も性能がよい Ladder Network を 0.7 ポイント上回った。また、単純な Attention のみのモデルである SASC と比べ、2.6 ポイント高い正解率となったことから、SE block や Residual module が英文の観点付き感情極性推定にも有効であることが分かる。表 2 には先行研究の ATAE-LSTM [16], CNN [5], Bi-LSTM の実験結果も示したが、提案モデルの正解率が最も高かった。CNN の filter size は 3, channel は 128, Bi-LSTM の出力次元は forward 層と backward 層どちらも 128 とした。

## 5 考察

提案モデルのモデル構造とハイパーパラメータは、Residual module の個数やパラメータを変更しながら実験を行い調整した。Residual module の内部構造については、Convolution 層の数、BN や Relu, Dropout の有無を比較した。1 つの Residual module に含まれる Convolution 層の数は 1 層または 2 層の正解率が高く、3 層以上にすると正解率が落ちることが分かった。画像認識の ResNet では層は深いほど良く、Convolution 層を 3 つ含む Bottleneck 構造 (図 2) が採用されていたが、提案モデル (図 5) では Bottleneck 構造を用いず、通常の Residual module の 2 層構造を採用した。

次に、図 5 の Convolution 層の前後での BN と Relu の有無と、Convolution 層間と Residual module 間での Dropout の有無について述べる。BN や Relu などの位置は He ら [15] は層が深いほど前につけることで精度が高くなることを示した。本稿では Convolution 層の前後とショートカットの加算後の計 4 か所でそれぞれ有無を比較した。挿入の検討箇所を図 6 に示す。その結果、1 つ目の Convolution 層の前で BN, 2 つの Convolution 層の後で BN と Relu, Convolution 層間で Dropout を挟むモデルが最も高くなることを確認した。そのため、Residual module (図 5) を提案した。

Dropout 率は最も高い結果は 0.5 のモデルだが、0.2 でも近い値を確認しており、誤差の範囲であった。

Convolution のハイパーパラメータについては、filter size は 1, 2, 3, 4, 5, channel 数は 64, 96, 100, 128, 200, 256, 300 の中の組み合わせで実験し、先行研究 [1] で用いられたパ

<sup>5</sup> : <http://alt.qcri.org/semeval2016/task5/>

<sup>6</sup> : <https://code.google.com/p/word2vec/>

表 2 観点付き感情極性推定の正解率

モデル	正解率
SE-ResNet Attention (提案モデル)	0.861
Ladder Network [1]	0.854
SASC [3]	0.835
ATAE-LSTM [16]	0.817
CNN [5]	0.828
Bi-LSTM	0.811

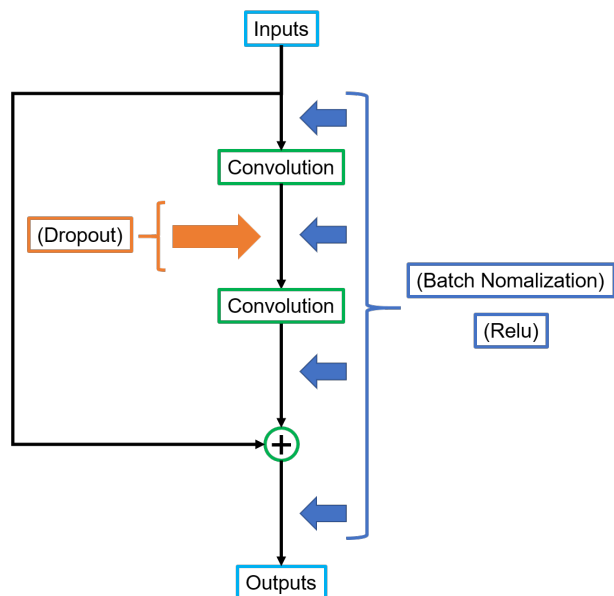


図 6 Residual module への BN, Relu, Dropout の挿入検討箇所

ラメータと同様に filter size が 3, channel 数が 128 のものが最も高い正解率となった。

図 4 の Residual module の数は 1 から 10 までの実装を比較したが、正解率に大きな差は出ず、1 つか 2 つで十分な効果が得られることが分かった。これらを比較したグラフを図 7 に示す。図 7 の Block は Residual module の数であり、Residual module の構造は図 5 の通りである。

最適化手法として用いた Momentum 項を加えた確率的勾配降下法の学習率と Momentum はそれぞれ 0.01 と 0.9 が TensorFlow<sup>7</sup>や Keras<sup>8</sup>の実装ではデフォルトとなっており、Qian [17] や Ruder [18] も推奨しているが、文章で学習を行う場合、学習が早く進みすぎるため、本実験では学習率を 0.005, Momentum を 0.5 と、それぞれ小さくすることで学習のピークを遅らせている。表 3 に正解率が 0.86 以上になったモデルとそのハイパーパラメータを示す。Reduction ratio に関しても、Hu ら [12] の実験から最も Top1 error 率が小さい 4 を選択したが、Top5 error 率とのバランスがよい 16 が推奨されている。実際に 16 でも差はなかった。また、2 つの Convolution 層で別々に filter size を定める組み合わせにも検討の余地があると考えられる。組み合わせでは (1,3), (3,3), (3,5) の組み合わせが比較的良好な結果となった。なお (1,3) は 1 つ目の Convolution 層の filter size

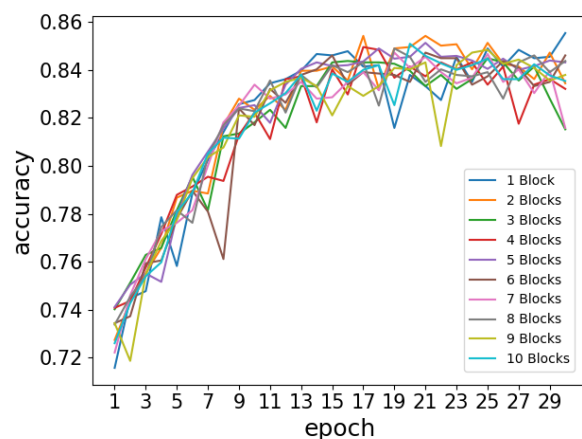


図 7 ResNet Attention モデルによる Residual module の数別の epoch 毎の正解率

表 3 観点付き感情極性推定の正解率が 0.86 以上だったモデル

Block 数	Reduction ratio	Dropout	filter size	channel	正解率
2	4	0.5	3,3	128	0.861
2	4	0.5	3,5	128	0.860
2	16	0.2	3,3	128	0.860

が 1, 2 つ目の Convolution 層の filter size が 3 という意味である。

## 6 ま と め

本稿では、英文レビュー文書に対する観点付き感情極性推定の精度向上のため、SE-ResNet Attention モデルを提案した。このモデルは Squeeze-and-Excitation block (SE block) と Residual module にカテゴリを Attention として組み合わせたモデルである。英語のレストランレビュー文を用いて実験し、感情極性の推定精度でモデルを評価した。提案モデルは Ladder Network や Simple Alignment Sentence Classification よりも高い正解率を達成し、提案手法が英文の観点付き感情極性推定において有効であることを示した。また、提案モデルの構造について考察し、SE block と Residual module は、層が比較的少なくても観点付き感情極性推定では十分な効果が得られることが分かった。

今後の課題としては、ベクトルの初期値を word2vec を用いて生成しているため、Transformer などを用いることでより高い精度が得られるかを検討したい。また Reduction ratio や filter size などのハイパーパラメータの調整に検討の余地がある。

## 文 献

- [1] 朝倉 遼, 新妻 弘崇, 太田 学: アテンションを利用した観点付き感情分析. 電子情報通信学会論文誌 (D), Vol.J102-D, No.4, pp.245-256. 2019. DOI: 10.14923/transinfj.2018DET0001
- [2] Antti Rasmus, Harri Valpola, Mikko Honkela, Mathias Berglund, Tapani Raiko: Semi-Supervised Learning with Ladder Networks. *arXiv preprint arXiv:1507.02672*. 2015.
- [3] Hirotaka Niitsuma, Ryo Asakura, Manabu Ohta: Simple

7 : [https://www.tensorflow.org/api\\_docs/python/](https://www.tensorflow.org/api_docs/python/)

8 : <https://keras.io/ja/>

Alignment Sentence Classification for Aspect-Based Sentiment Analysis. In: *Computational Linguistics and Intelligent Text Processing: Proceedings of the 19th International Conference*. 2018.

- [4] Bing Liu: Sentiment analysis and opinion mining. 2012.
- [5] Yoon Kim: Convolutional Neural Networks for Sentence Classification. *arXiv preprint arXiv:1408.5882*. 2014.
- [6] Dzmitry Bahdanau, KyungHyun Cho, Yoshua Bengio: Neural Machine Translation by Jointly Learning to Align and Translate. *arXiv preprint arXiv: 1409.0473*. 2016.
- [7] Toshihiko Yanase, Kihsuke Yanai, Misa Sato, Toshinori Miyoshi, Yoshiki Niwa: Aspect-based Sentiment Analysis by using Neural Attention Model. In: *The 30th Annual Conference of the Japanese Society for Artificial Intelligence*. 2016.
- [8] Zhiqiang Toh, Jian Su: NLANGP: Supervised Machine Learning System for Aspect Category Classification and Opinion Target Extraction. In: *Proceeding of the 9th International Workshop on Semantic Evaluation (SemEval-2015)*. 2015.
- [9] Sergey Ioffe, Christian Szegedy: Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv preprint arXiv: 1502.03167*. 2015.
- [10] Diederik P. Kingma, Jimmy Ba: Adam: A Method for Stochastic Optimization. *arXiv preprint arXiv: 1412.6980*. 2014.
- [11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun: Deep Residual Learning for Image Recognition. *arXiv preprint arXiv: 1512.03385*. 2015.
- [12] Jie Hu, Li Shen, Samuel Albanie, Gang Sun, Enhua Wu: Squeeze-and-Excitation Networks. *arXiv preprint arXiv: 1709.01507*. 2017.
- [13] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich: Going deeper with convolutions. *arXiv preprint arXiv: 1409.4842*. 2014.
- [14] Karen Simonyan, Andrew Zisserman: Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv preprint arXiv: 1409.1556*. 2014.
- [15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun: Identity Mappings in Deep Residual Networks. *arXiv preprint arXiv: 1603.05027*. 2016.
- [16] Yequan Wang, Minlie Huang, Xiaoyan Zhu, Li Zhao: Attention-based LSTM for Aspect-level Sentiment Classification. In: *Proceeding of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp.606-615. Association for Computational Linguistics, 2016.
- [17] Ning Qian: On the Momentum Term in Gradient Descent Learning Algorithm. *Neural networks : the official journal of the International Neural Network Society*, 12(1):145–151, 1999.
- [18] Sebastian Ruder: An overview of gradient descent optimization algorithms. *arXiv preprint arXiv: 1609.04747*. 2016.