

# ツイートおよびレビューの極性判定における 畳込みニューラルネットワークの性能評価

Evaluation of Convolutional Neural Networks in Polarity Classification of Tweets and Reviews

博士前期課程 1 年 201720879 山本航平

筑波大学大学院 システム情報工学研究科 知能機能システム専攻

Kohei Yamamoto

Graduate School of Systems and Information Engineering, University of Tsukuba

(指導教員: 宇津呂武仁, 中内靖, 丸山勉)

In this study, we propose new CNN models that classify positive and negative sentiment of the sentences, and compare the performance of each model. We tested 12 types of models; the basic model, the multi-channel model, two types of models having an additional convolution layer, two types of models having additional input features on location, and these models having two fully connected layers. The benchmark results showed that the model with multi-channel having one fully connected layer exhibited the best performance, and in the most of tested models, the correct classification rate is better in models with one fully connected layer than that in models with two fully connected layers.

## 1. はじめに

現在, 人工知能は日常生活やニュースで良く目にするようになった. その中心にある手法は, 深層学習であり, 画像分野や音声分野など様々な分野において優れた成果を挙げ続けている. 自然言語処理の分野も例外ではなく, 深層学習の手法の 1 つである回帰型ニューラルネットワーク (RNN) を用いて機械翻訳を行う研究 [6] や, 同じく RNN を用いた言語モデルの研究 [4] など, 様々な研究がおこなわれている. また最近では, 画像分野で良く用いられる畳込みニューラルネットワーク (CNN) を用いた文書分類の研究がおこなわれ, 従来手法を上回る精度が得られている. しかし, CNN を用いた文書分類の研究は, 最近注目されたこともあり, 研究事例が少なく, 提案モデルの数が少ないのが現状である. そこで我々は, 新たなモデルを提案し, 実際にモデルを作成し文書の positive, negative といった正負感情の分類を行った. そして各モデルの正答率を比較し, どのモデルが最も正答率が良いのかを検証した.

## 2. 関連研究

関連研究として, Socher [5] と Wang [7] が本研究で用いる映画レビューをデータセットとして極性判定を行っている. Socher は再起型ニューラルネットワーク (recursive neural network) を用いたモデルを, Wang はニューラルネットワークの学習時に過学習を防ぐ手法の 1 つであるドロップアウト (Hinton [1]) を高速で行うアルゴリズムを用いて極性判定を行い, 従来手法を上回る評価結果を得ている. Kim [2] は CNN を用いたモデルを提案し, 様々な文書分類のタスクに提案モデルを用いて評価を行っている. この研究によると幾つかのタスクにおいて, Socher や Wang が提案したモデルを含む従来手法より優れた評価結果を得た事を示している. 本研究では Kim が提案したモデルを元に幾つかのモデルを提案し評価を行う.

## 3. 畳込みニューラルネットワーク (CNN)

本研究で用いる畳込みニューラルネットワーク (CNN) は畳込み層とプーリング層という特別な層を追加した順伝播型のニューラルネットワークである. 今回, 実験で使用した CNN は畳込み層, プーリング層, 全結合層の 3 種類の層から構成されている (図 1).

畳込み層は, 入力データにフィルタを重ねた時, 互いの重なり合う値どうしの積を求め, その和を出力する. この一連の計算を畳込みと呼び, この層の名前の所以となっている. また, 出力された値の集合を特徴マップと呼ぶ.

プーリング層は, 通常は畳込み層の直後に設置する. プーリング層の働きは, 畳込み層で抽出された特徴の位置感度を若干低下させることであり, 種類が幾つか存在する. 代表的なものでは最大プーリングと平均プーリングであり, 最大プーリングはフィルタが重なった部分の最大値を, 平均プーリングは平均値を出力する. 本研究では全モデル共通で最大プーリングを行った.

全結合層は, ある層のユニットが隣接層の全ユニットと結合している層である. 本研究では, 中間層と出力層からなる 2 層の全結合層を用いた場合と, 出力層のみから構成される 1 層の全結合層を用いた場合の 2 通りで正答率を求めた.

## 4. CNN を用いた文書の感情分類

### 4.1 Kim [2] のモデル

Kim [2] のモデルは畳込み層, プーリング層, 全結合層を各 1 つずつ持つ非常にシンプルなモデルである. 以後, 本論文ではこのモデルをベースラインとする (図 2).

入力層では文書上の単語を分散表現から 300 次元のベクトルとして表現し, そのベクトルを並べる事で文書を行列としてモデルに入力する. 行列のサイズは, 行はベクトルの次元数, 列は文書の単語数となる. この時, 文書行列のサイズは全てのデータにおいて不変でなくてはならない為, 各文書の長さを

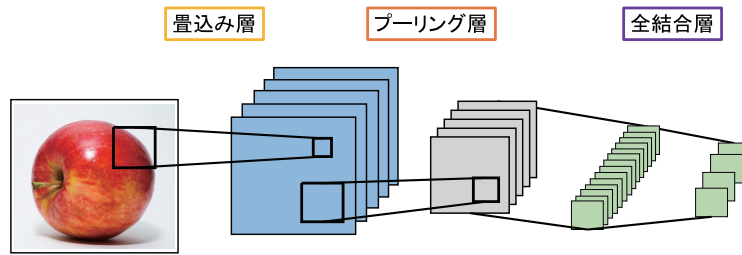


図 1: CNN の全体構造の略図

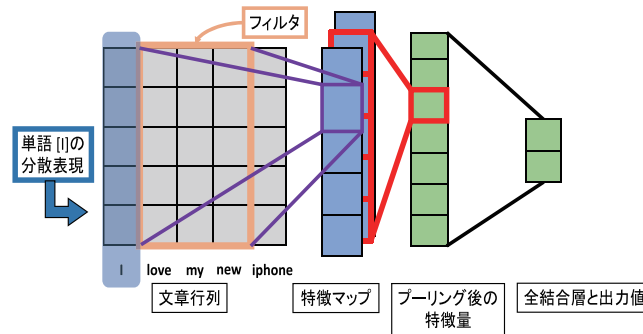


図 2: Kim [2] のモデルの略図

データセットの中で最長の長さの値に統一し、文書の長さから余った行列の成分には 0 を与えた。

畳込み層では、単語ベクトルごとにスライドするフィルタを用いて畳込みを行う。例えば、300 次元の分散表現から構成された入力行列に対して、2 単語毎にスライドするフィルタの場合、サイズは  $300 \times 2$  となる。  $h$  単語毎にスライドするフィルタの係数を  $w$ 、バイアスを  $b$ 、活性化関数を  $f()$ 、あるタイミングでフィルタと重なる単語ベクトルを  $x_{i:i+h-1}$  とすると、畳込み後の特徴量  $c_i$  は次の式で表せる。

$$c_i = f(W \cdot x_{i:i+h-1} + b) \quad (1)$$

文書上の単語全てに畳込みを終えた後は、各タイミングで生成された特徴量を並べて特徴マップ  $c$  を生成する。

プーリング層では、畳込み層で獲得した特徴マップに対して最大プーリングを行う。プーリング後の特徴量を  $\hat{c}$  とすると数式は次の通りとなる。

$$\hat{c} = \max\{c\} \quad (2)$$

全結合層ではプーリング層から出力された特徴量を入力として、正負感情の分類を行う。また、過学習の対策としてドロップアウト及び重み上限を行う。全結合層の入力の値を  $z$ 、重みを  $w$ 、バイアスを  $b$ 、活性化関数を  $f()$  とすると次の数式で表せる。

$$y = f(w \cdot z + b) \quad (3)$$

活性化関数は softmax 関数を用いた。

ドロップアウトは、全結合層 2 層の場合、入力層及び隠れ層で確率 0.2、0.7 で行い、全結合層 1 層の場合、入力層で確率 0.7 でドロップアウトを行った。重み上限は、全結合層 2 層の場合、隠れ層と出力層の間の重みで行い、全結合層 1 層の

表 1: 本研究で用いたモデル

| モデル名      | モデル紹介               |
|-----------|---------------------|
| ベースライン    | Kim が提案したモデル        |
| 多チャンネルモデル | 文書行列を複数用意するモデル      |
| 畳込み層追加モデル | 畳込み層を 1 つ追加するモデル    |
| 位置情報追加モデル | 分散表現を元に位置情報を追加するモデル |

場合は、入力層と出力層の間の重みで重み上限を行った。共に重みの上限値は 3.0 とした。

#### 4.2 提案モデル

次に本研究で提案したモデルについて説明する。本研究では次の表 1 に記載されている複数のモデルを用いて正答率の比較を行った。

多チャンネルモデルは、ベースラインと殆ど同じ形であるが、入力層で文書行列を複数用意する点で異なる (図 3)。複数の行列を用意する方法として、単語ベクトルを均等に分割することで複数の行列を作成する。本研究では 300 次元のベクトルを 3 分割して、3 つの文書行列を作成した。

畳込み層 2 層モデルは、畳込み層を 1 層追加したモデルである。本研究では畳込み層を追加する方法を 2 つ用意し、それぞれ畳込み層 2 層モデル (1)、畳込み層 2 層モデル (2) とする (図 4)。畳込み層 2 層モデル (1) では、1 回目の畳込み層から抽出された特徴マップを行列の 1 行と考え、幾つか並べて新しい行列を作り、再度畳込みを行う。畳込み層 2 層モデル (2) では、最初にフィルタサイズが、縦横 1 のフィルタを用いて畳込みを行い、抽出された特徴マップを、ベースラインと同じ形の畳込み層を用いて再度畳込みを行う。その後のプーリング層、

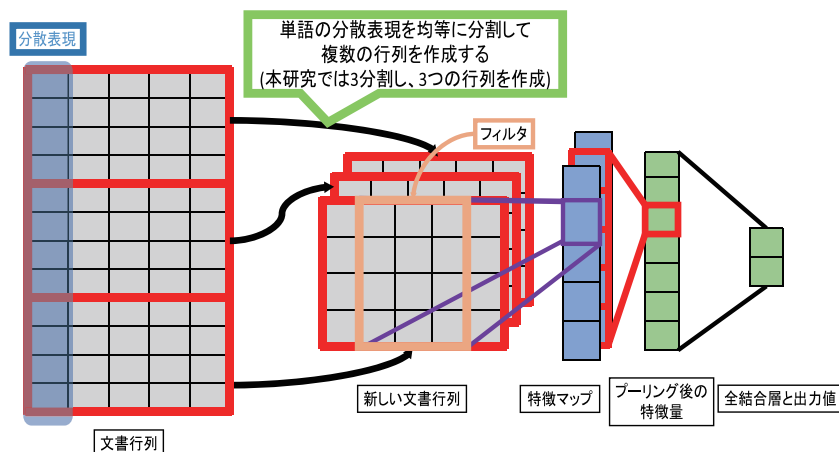


図 3: 多チャネルモデルの略図

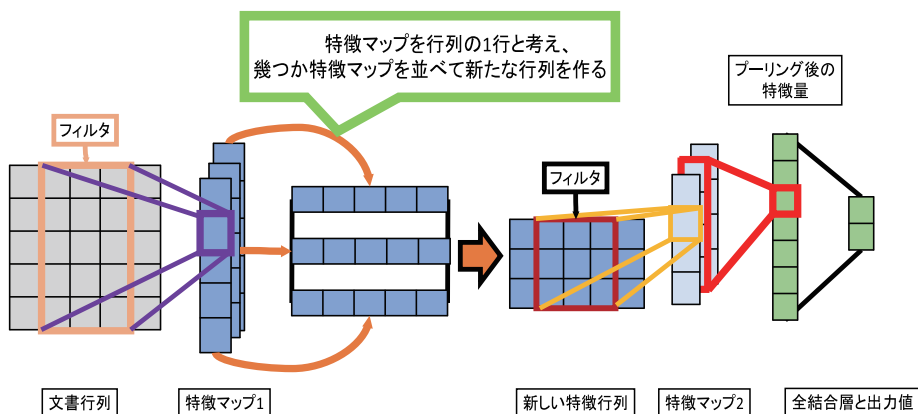


図 4: 畳込み層 2 層モデル (1) の略図

全結合層は共にベースラインと同じ形である。

位置情報追加モデルは、単語の分散表現を元に単語の位置情報を追加して分類を行う。本研究では、位置情報を求め、追加する方法を2つ用意し、それぞれ、位置情報追加モデル(1)、位置情報追加モデル(2)とした(図5)。位置情報追加モデル(1)では、単語間の分散表現のユークリッド距離を値とした位置行列を作り、文書行列と共に畳込み、プーリングを行い、抽出されたそれぞれの特徴量を組み合わせ、全結合層の入力層の値とする。文書行列における畳込み層、プーリング層はベースラインと同じ形であり、位置行列における畳込みフィルタのサイズは、文書行列で用いる畳込みフィルタサイズと等しい。位置行列におけるプーリング層は、ベースラインと同じく、特徴マップと同じサイズのフィルタを用いて最大プーリングを行う。位置情報追加モデル(2)では、単語の分散表現と0とのユークリッド距離を値とした位置ベクトルを作り、畳込み、プーリングから抽出された文章行列の特徴量と組み合わせ、新しい特徴量を生成し、分類を行う。

## 5. 評価

### 5.1 データセット

本研究で用いたデータセットは、Twitter データと映画レビューデータの2つである(表2)。Twitter データは、"Stanford

表 2: データセット

| データ名        | 文書の平均単語数 | 文書の最大単語数 | 文書数    | データ中で用いられる単語の種類数 |
|-------------|----------|----------|--------|------------------|
| Twitter データ | 17       | 49       | 15,000 | 23,964           |
| 映画レビュー      | 23       | 57       | 10,662 | 18,751           |

Twitter Corpus", 映画レビューは, "Movie Review Data" を用いた。各データセットには, negative, positive のラベル付けがされている。また, 共に前処理としてコンマやピリオドなどを取り除き, ゼロパディングを行うためにデータセットの各文書の前後に空白を追加した。本研究では, 訓練データを9割, テストデータを1割として学習に用いた。

### 5.2 評価手順

各データセットの単語を分散表現に変換し, 入力データとする。単語の分散表現を獲得する手法としては, word2vec [3] を用いた。本研究では, Google News データセットから事前学習を行ったモデル\*1を用いて300次元の分散表現を獲得す

\*1 <https://code.google.com/archive/p/word2vec/>

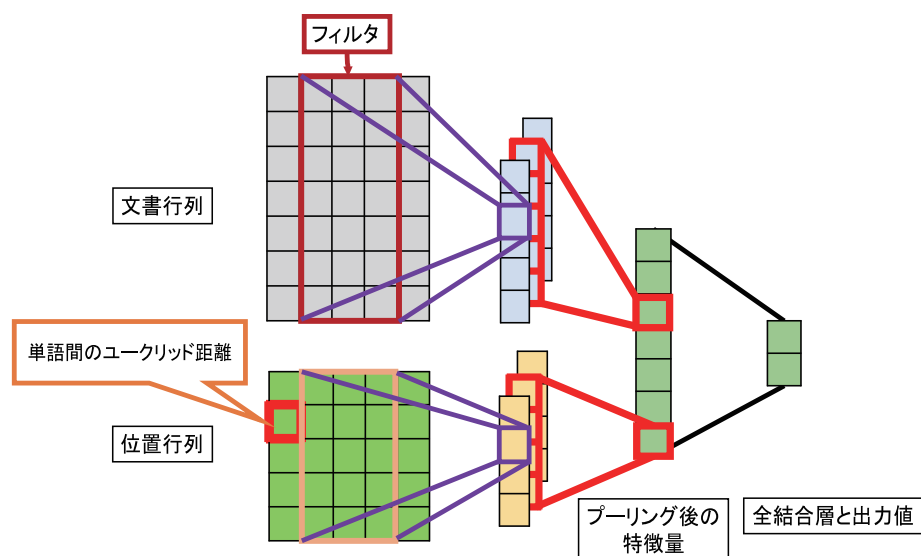


図 5: 位置情報追加モデル (1) の略図

表 3: 各モデルの正答率

| モデル名            | 全結合層の層数 | twitter における正答率 | 映画レビュー における正答率 |
|-----------------|---------|-----------------|----------------|
| ベースライン          | 1       | 81.5            | 82.0           |
|                 | 2       | 80.7            | 80.4           |
| 多チャンネルモデル       | 1       | <b>81.9</b>     | <b>82.3</b>    |
|                 | 2       | 80.6            | 80.4           |
| 畳込み層 2 層モデル (1) | 1       | 80.6            | 80.6           |
|                 | 2       | 80.9            | 79.8           |
| 畳込み層 2 層モデル (2) | 1       | 81.5            | 81.3           |
|                 | 2       | 80.4            | 80.0           |
| 位置情報追加モデル (1)   | 1       | 81.4            | 82.1           |
|                 | 2       | 81.0            | 80.2           |
| 位置情報追加モデル (2)   | 1       | 81.7            | 82.2           |
|                 | 2       | 80.8            | 80.6           |

る。評価手順としては、先ず初めに訓練データを用いてモデルを学習する。その後テストデータを用いて、正答率から各モデルの評価を行う。

### 5.3 評価結果

各モデルを用いて 2 つのデータセットに対して、150 エポック試行して正答率を求め、10 回の平均値を結果として表 3 に示す。

本研究で提案したモデルの中で最も正答率が良かったモデルは、多チャンネルモデルの全結合層が 1 層のモデルであった。また、ほとんどのモデルで全結合層が 2 層の場合より、1 層の場合の方が正答率が良かったという結果となった。多チャンネルモデル、位置情報追加モデルはベースラインより正答率が良く、一方で畳込み層 2 層モデルはベースラインより正答率が悪いという結果となった。

## 6. おわりに

本論文では、Kim [2] のモデルを元に新たなモデルを提案し、文書の positive, negative の正負感情の分類を行い、提案モデ

ルの正答率を比較し検証した。

今後の課題として、今回は正負感情の 2 値分類の問題にしか提案モデルを用いていないので、多種類の感情分類を行えるデータセットを用いて多分類を行いたい。また、多チャンネルモデルと位置情報追加モデルを組み合わせた、より複雑なモデルを考案することが出来たのならば、更に正答率が良いモデルとなるのではないかと考えた。

## 参考文献

- [1] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov. Improving neural networks by preventingco-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*, 2012.
- [2] Y. Kim. Convolutional neural networks for sentence classification. In *Proc. EMNLP*, pp. 1746–1751, Doha, Qatar, October 2014. Association for Computational Linguistics.
- [3] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. In *Proc. ICLR*, 2013.
- [4] T. Mikolov, M. Karafiat, L. Burget, J. "Honza" Cernocky, and S. Khudanpur. Recurrent neural network based language model. In *Proc. INTERSPEECH*, pp. 1045–1048, 2010.
- [5] R. Socher, B. Huval, C. D. Manning, and A. Y. Ng. Semantic compositionality through recursive matrix-vector spaces. In *Proc. EMNLP*, pp. 1201–1211, Jeju Island, Korea, July 2012. Association for Computational Linguistics.
- [6] I. Sutskever, O. Vinyals, and Q. V. Le. Sequence to sequence learning with neural networks. In *Proc. 27th NIPS*, pp. 3104–3112. Curran Associates, Inc., 2014.
- [7] S. Wang and C. Manning. Fast dropout training. In Sanjoy Dasgupta and David McAllester, editors, *Proc. 30th ICML*, Vol. 28 of *Proceedings of Machine Learning Research*, pp. 118–126, Atlanta, Georgia, USA, 17–19 Jun 2013. PMLR.