

# word2vec と深層学習を用いた大規模評判分析

加藤 和平\*      大島 考範\*      二宮 崇\*\*

\* 愛媛大学 工学部 情報工学科

\*\* 愛媛大学 大学院理工学研究科 電子情報工学専攻

{kato@ai.,ohsh@ai.,ninomiya@}cs.ehime-u.ac.jp

## 1 はじめに

近年、情報インフラの発達によって多様な商品、サービスが提供されている。これらは有形無形問わずに膨大な量があるが、そのような商品あるいはサービス群の中から商品の善し悪しを判断することができれば非常に有用である。楽天データは、各商品に対しレビュー文書と評価値が付与された大規模データであり、約1億5600万商品、約6400万レビューから成る。楽天データから評判分析の学習を行うことで実用的かつ精度の高い評判分析器が実現されることが期待される。

本論文は word2vec と深層学習を用いた大規模評判分析について報告する。近年、機械学習の分野、特に画像認識において深層学習が非常に高い成果を上げており、大きな注目を浴びている。本研究では中間層が0層、1層または2層から成るニューラルネットワークを用いて、大規模評判分析データである楽天データを対象に評判分析を行う。本研究では 1-of- $K$  表現の単語ベクトルを直接用いるのではなく、word2vec により得られる分布表現の単語ベクトル (単語概念ベクトル) を用いてニューラルネットワークの学習を行う。

## 2 関連研究

word2vec[2] はニューラルネットワークを用いた単語の分布表現 (単語の概念を表す低次元の密なベクトルによる表現) を獲得する手法である。テキスト中の各単語を周辺の単語から予測する擬似的な単語予測のタスクを設定し、このタスクを大量のテキストからニューラルネットワークで学習し、中間層における各単語の重みを抽出することによって、単語に対する概念ベクトルを獲得する。周辺の単語の重みベクトルの和を中間層の値とするモデルは CBOW モデルと呼ばれ、周辺の単語のうちの一つに対する重みベクトルを中間層の値とするモデルは Skip-gram モデルと呼ばれ

る。CBOW モデルでも Skip-gram モデルでも、入力層と中間層をつなぐ重み行列 (各単語に対する重みベクトルの集合) が、word2vec が最終的に生成する単語分布表現 (各単語に対する概念ベクトル) となる。単語を分布表現に変換することにより単語を意味空間上の1点に対応させることができ、意味的に関連の強い単語グループに分類したり、単語に対する意味的な計算を可能とする利点がある。単語に対する重みベクトルの次元数は中間層のノード数と等しく、word2vec において数百程度がよく用いられる。つまり、入力層における語彙数 (数十万) の次元を数百程度に次元削減していることになる。

深層学習は、効率的な多層ニューラルネットワークの学習法であり、機械学習の分野、特に画像認識において非常に高い成果を上げている。本研究では画像認識のための深層学習のツールキットとして有名な Caffe[3] を用いて多層ニューラルネットワークの学習を行う。Caffe は高速なフレームワークであり、モデルファイルによりネットワークの構造を与えることができるため、多層化、損失関数の設定、活性化関数の設定を容易にすることが可能となっている [1]。ただし、現在の Caffe の実装では 1-of- $K$  単語ベクトルのような疎なベクトルを効率的に処理するにはなっておらず、本研究では word2vec を併用し次元削減を行う。

## 3 word2vec を用いた多層ニューラルネットワークによる大規模評判分析

本論文は word2vec による単語分布表現 (単語の概念を表す低次元の密なベクトルによる表現) を用いた、多層ニューラルネットワークによる大規模評判分析について報告する。多層ニューラルネットワークの中間

表 1: データセット

	データサイズ	レビュー件数	内訳
訓練データ	4.655GB	7,606,045	レビューデータ 01-06
開発データ	1.02GB	1,655,042	レビューデータ 11
テストデータ	1.12GB	1,818,107	レビューデータ 12

層は0層, 1層または2層について試した. 出力層にはソフトマックス関数を適用し, 入力層には word2vec による単語分布表現を用いた文書ベクトルを入力として与える. 図 1 の (A), (B), (C) はそれぞれ中間層が 0 層の場合, 1 層の場合, 2 層の場合のネットワーク構造を表している. 0 層のモデルはロジスティック回帰モデルと等価なモデルとなる.

あるレビュー文書  $r$  を単語列  $w_1 w_2 \dots w_n$  としたとき,  $r$  に対する文書ベクトル  $\mathbf{d}$  は, 次式により与えられる.

$$\mathbf{d} = \sum_{i=1}^n \text{word2vec}(w_i)$$

ただし,  $\text{word2vec}(w_i)$  は word2vec により与えられる  $w_i$  の単語分布表現である.

## 4 実験

本節は従来手法と提案手法の評価実験について述べる.

### 4.1 データセット

本研究の実験ではデータセットとして楽天から公開されている楽天データの中から, 楽天で取り扱う商品に対するレビューデータ群及びその商品に対する評価を用いる. 詳細は表 1 に記す.

### 4.2 実験設定

単語分布表現の獲得には word2vec を用いる. 本研究では word2vec の出力として 200 次元, 400 次元, 600 次元, 800 次元, 1000 次元という 5 パターンの単語分布表現を得た. その際 word2vec は次のオプションで実行した. モデルとしては CBOW モデルを用いる (-cbow 1), ウィンドウサイズは最大 8 単語 (-windows 8), ネガティブサンプリングのネガティブサンプルの個数は 25 個 (-negative 25) とし, 階層的ソフトマックスは使用しなかった (-hs 0). また, 学習の繰り返し

表 2: Caffe 用データセット

次元数	訓練データ件数	開発データ	テストデータ
200 次元	7,686,045	1,655,042	1,818,107
400 次元	5,368,709	1,655,042	1,818,107
600 次元	3,579,139	1,655,042	1,818,107
800 次元	2,684,354	1,655,042	1,818,107
1000 次元	2,147,483	1,655,042	1,818,107

表 3: Caffe の設定

パラメータ	数値
中間層数	0-2
バッチサイズ	100
ノード数	400
iteration 間隔	1,000
最大 iteration 回数	100,000

回数は 15 回 (-iter 15) とし, 頻出語をランダムに削除した (-sample 1e-3).

深層学習には Caffe を用いた. 楽天データが大規模データであることと Caffe の実装の制限のため, word2vec を用いて次元削減をしても 200 次元より大きい次元になる場合には全データを用いて学習することができない. 従って, 200 次元より大きな次元で学習する場合は訓練データのサイズを変更して学習することにした. 表 2 は, 今回の実験で用いたデータのサイズと次元数との関係を表している. また Caffe のパラメータについては表 3 に記載する. ニューラルネットワークの構造についてはそれぞれ図 1 に示される通りである.

### 4.3 実験結果

比較する従来法として, Liblinear[4] を用いた L2 正則化ロジスティック回帰と比較する. LibLinear のハイパーパラメータについては開発データを用いて調整した. 提案手法に対する実験では, 3 種類のモデル (中間層 0 層, 中間層 1 層, 中間層 2 層) と 5 種類の次元数 (200 次元, 400 次元, 600 次元, 800 次元, 1000 次元) を用いた性能評価を開発データに対して行った.

表 4 はテストデータに対する実験結果を示し, 表

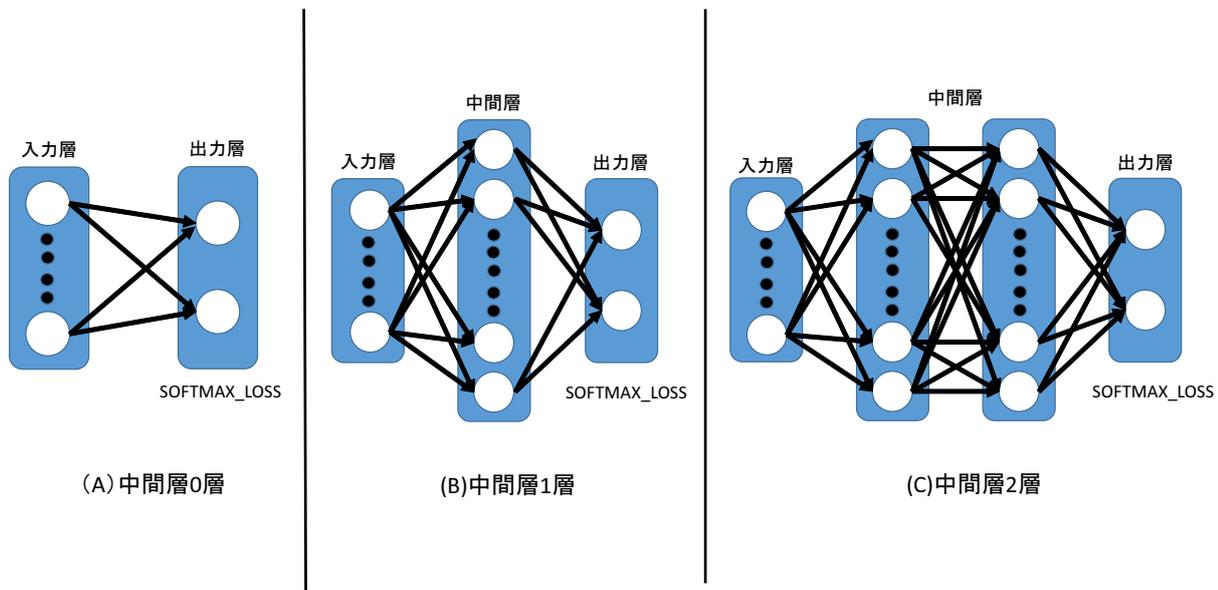


図 1: 深層学習におけるニューラルネットワークの構造

表 4: テストデータに対する精度

学習モデル	精度 (%)
LR(1-of- $K$ )	91.0151
NN-L0(w2v)	90.188
NN-L1(w2v)	90.555
NN-L2(w2v)	90.675

表 6: 訓練データサイズ固定で学習した場合のテストデータに対する精度

学習モデル	精度 (%)
NN-L0(w2v)	90.156
NN-L1(w2v)	90.566
NN-L2(w2v)	90.637

表 5: 開発データに対する精度 (%)

次元数	NN-L0(w2v)	NN-L1(w2v)	NN-L2(w2v)
200	90.008	90.577	90.653
400	90.025	90.642	90.703
600	90.203	90.830	90.920
800	90.121	90.679	90.774
1000	89.725	90.331	90.594

5 は 10 万イテレーションの学習を行った場合の各モデルと次元数に対する開発データでの精度を表している。表中における“LR(1-of- $K$ )”は 1-of- $K$  単語ベクトルを用いたロジスティック回帰の結果を示し、“NN- $L_i$ (w2v)”は word2vec により得られる単語分布表現を入力とした中間層数  $i$  の中間層から成るニューラルネットワークの結果を示す。開発データに対する性能評価の結果、全てのモデルにおいて 600 次元のときが最も良い性能であったため、テストデータに対する実験では 600 次元の入力を用いて評価を行った。また、最適なイテレーション回数を開発データを用いて設定した。

上記の実験では、Caffe の制限により訓練データのサイズと次元数を変更させながら学習を行ったが、訓

練データサイズを固定した場合の性能についても評価した。訓練データのレビューデータ数を 1000 次元の場合の 2,147,483 件に統一して実験を行った結果、開発データに対する性能評価は表 7 のようになり、テストデータに対する性能評価は表 6 のようになった。

実験結果より、中間層数を増やした学習の方がより高い精度を出していることがわかる。次元数に関しては訓練データサイズ固定の場合の実験結果より、必ずしも高次元のデータセットから高精度が得られるわけではないことが分かる。また訓練データサイズ変動の場合と訓練データサイズ固定の場合でも精度が同程度であることから、少量のデータセットにおいて提案手法が特に有効であると言える。提案手法は 1-of- $K$  表現の単語ベクトルを用いたロジスティック回帰の性能には及ばなかったが、ほぼ同程度の性能であることがわかる。

## 5 おわりに

本研究は、大規模評判分析データである楽天データに対し、word2vec と深層学習を用いた評判分析の学

表 7: 訓練データサイズ固定で学習した場合の開発データに対する精度 (%)

次元数	NN-L0(w2v)	NN-L1(w2v)	NN-L2(w2v)
200	90.008	90.572	90.665
400	90.038	90.625	90.691
600	90.203	90.817	90.894
800	90.122	90.672	90.779
1000	89.725	90.331	90.594

習を行い、その性能評価について報告した。本研究では、ニューラルネットワークの入力として 1-of- $K$  表現の単語ベクトルを直接用いるのではなく、word2vecにより得られる分布表現の単語ベクトルを入力として用い、中間層が 0 層、1 層または 2 層から成るニューラルネットワークを用いて、楽天データに対する評判分析を行った。多層ニューラルネットワークのツールとしては Caffe を用いた。

Caffe は疎なベクトルに対して効率的に学習を行う仕様となっていないため、大規模データである楽天データに対し直接 1-of- $K$  表現の単語ベクトルから学習することは難しい。本研究は word2vec を用いて次元削減を行うことでデータサイズを削減し Caffe で学習することを実現した。データサイズと次元数との間にトレードオフができるため、データ数と次元数を同時に増減させてその性能を評価した。また、データサイズを固定し次元数を増減させた場合の性能についても評価した。実験結果より、ニューラルネットワークの中間層を増やすほど性能が向上することが確認できた。1-of- $K$  表現の単語ベクトルを用いたロジスティック回帰の性能には及ばなかったが、ほぼ同程度の性能であることがわかった。

## 参考文献

- [1] Jia, Yangqing and Shelhamer, Evan and Donahue, Jeff and Karayev, Sergey and Long, Jonathan and Girshick, Ross and Guadarrama, Sergio and Darrell, Trevor (2014). "Caffe: Convolutional Architecture for Fast Feature Embedding." arXiv preprint arXiv:1408.5093.
- [2] word2vec, <https://code.google.com/p/word2vec/>
- [3] Caffe, <http://caffe.berkeleyvision.org/>
- [4] Liblinear, <http://www.csie.ntu.edu.tw/~cjlin/liblinear/>