

オートエンコーダにおける単語ベクトルの学習

島岡 聖世 山本 風人 乾 健太郎
東北大学

simaokasonse@yahoo.co.jp, {inui, kazeto}@ecei.tohoku.ac.jp

1 はじめに

単語や文の意味的な特性を実数ベクトルでうまく表現することができれば、様々な自然言語処理タスクにおいて有用であることが知られている。文や節の意味的特性をベクトルで表現するアプローチのひとつとして、文や節を構成する各単語の意味的特性を表すベクトル(単語ベクトル)に数値演算を施すことによって文や節のベクトルを構成的に算出するものがあるが、中でも近年注目されている手法の一つに Unfolding Recursive Autoencoder[4](URA)がある。URAは文を構成する単語を構文的構造に基づいて再帰的に組み合わせ、文や節のベクトル表現を生成する枠組みである。この手法の優れた点は、単語対を組み合わせる際に用いる重み行列を教師無しで学習できる点である。しかし、その際の単語ベクトルの学習については既存の研究では詳しく言及されていない。そこで本稿では、URAの基礎技術であるオートエンコーダについて、重みパラメータの学習の際に単語のベクトルがどのように学習されるかに関して調査を行った結果を報告する。またその結果を元にオートエンコーダと言語モデルとの結合モデルを提案し、そのモデルがオートエンコーダの学習と単語ベクトルの学習を同時に行っていることを報告する。そして、この結合モデルがオートエンコーダだけでなくURAに対しても適用可能であることを示す。

以下の構成を次に示す。まず2節ではオートエンコーダの概要と、オートエンコーダの自然言語処理への適用例としてURAを紹介する。3節ではオートエンコーダの学習における単語ベクトルの振る舞いについて行った実験に関して報告する。4節ではオートエンコーダと言語モデルの結合モデルを提案し、学習実験の結果を報告する。5節では4節で提案したモデルがURAに対しても適用可能であることを示し、それを実証するいくつかの実験結果を報告する。最後に6節で本研究の全体の総括を行う。

2 先行研究

本節では先行研究としてオートエンコーダとその言語処理への応用について述べる。

2.1 オートエンコーダ

オートエンコーダは出力が入力と近い値を持つように学習するニューラルネットワークの枠組みである。オートエンコーダの一例を図1に示す。入力層を x 、中間層を p 、出力層を y とし、入力層から中間層への重み行列を W_e 、中間層から出力層への重み行列を W_d とする。このとき x と y の次元数は等しいものとする。このときの入出力の関係は、通常のニューラルネットワークと

同様に次式のように表される：

$$p = f(W_e \cdot x) \quad (1)$$

$$y = f(W_d \cdot p) \quad (2)$$

ただし、関数 f はニューラルネットワークの活性化関数であり、一般には \tanh 関数などのシグモイド関数を用いられる。オートエンコーダの学習は、再現誤差(Reconstruction Error) E^{rec} を最小化することによって行われる：

$$E^{rec}(x) = \|x - y\|^2 \quad (3)$$

このとき、出力と入力似た値であるほど E^{rec} は低い値をもつ。

なお、一般にオートエンコーダは次元圧縮の手法のひとつとして用いられる。特に活性化関数 f が線形である場合、中間層の情報表現は、広く知られた次元圧縮の手法である主成分分析と似たものになる。

2.2 Unfolding Recursive AutoEncoder

近年、オートエンコーダの自然言語処理への応用が大きな注目を集めている。その一つとして、SocherらのUnfolding Recursive Autoencoder [4]がある。これは、構文構造に基づいて単語ベクトルを再帰的にオートエンコーダで次元圧縮していき、文や節の意味的特徴を表すベクトルを生成するというものである。このモデルの優れた点としては、文や節の意味的な特性を単語ベクトルから構成的に生成できること、モデルパラメータを教師無しで学習できること、ベクトルの値とそれが表す意味的特性との対応が、文・節・単語など全ての単位で共通であること、などが考えられる。Socherらはこのモデルを言い換え認識のタスクに適用し、オートエンコーダによって得られた文や節のベクトルがそれを構成する単語の意味的特徴だけでなく構文情報もうまく捉えていること、またそのようなベクトルが解析に有用であることを経験的に示した。

3 オートエンコーダの学習における単語ベクトルの学習について

オートエンコーダのモデルパラメータの学習は、式3で定義された再現誤差を最小化することによって実現される。この学習においては、単語ベクトルを学習対象とみなすかどうか、学習対象とした場合に何が起るか、などの議論が考えられるが、我々が知る限りではこれまでの研究でそれらに関する議論は未だなされていない。

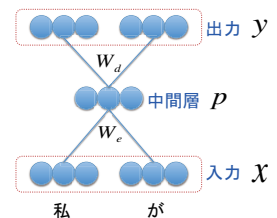


図1: オートエンコーダの例

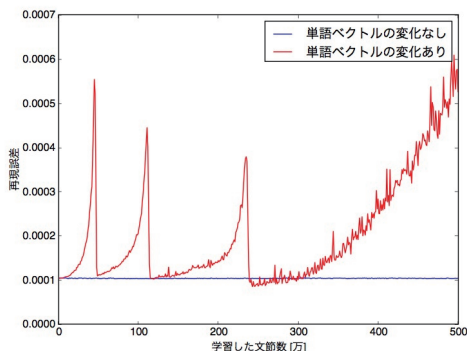


図 2: 乱数を初期値として学習した場合の結果

そこで本節では、オートエンコーダの重み学習における単語ベクトルの学習についていくつかの実験を行った結果を報告する。

3.1 行った実験

我々はオートエンコーダの重み学習における単語ベクトルの振る舞いについて、以下の調査を行った：

- オートエンコーダの学習において、単語ベクトルを同時に学習した場合と固定した場合、初期値を乱数で与えた場合とあらかじめ言語モデルで学習した単語ベクトル(このモデルの詳細は5節で扱う)を与えた場合で、それぞれ学習後の単語ベクトルがどのように変化したかを比較する。
- オートエンコーダの学習において、あらかじめ言語モデルで学習した単語ベクトルの一部を乱数で初期化した上で学習を行う。この時に乱数で初期化した単語ベクトルがどのように変化するかを調べ、オートエンコーダの学習における未知語の学習に対する知見を得る。

実験の評価は、オートエンコーダの再現誤差がどのように推移するかと、学習後の単語ベクトルが単語間の意味的な類似度をどれだけ捉えられているかの、2つの観点に基づいて行った。後者については、同じ品詞を持つ単語のペアを正例、異なる品詞を持つ単語のペアを負例として、それぞれ1000個ずつ用意し、各ペアの単語間のコサイン類似度の平均値を算出することで評価した。このとき正例のコサイン類似度は高い方が望ましく、負例のコサイン類似度は低い方が望ましい。

学習のコーパスにはTSUBAKI[5]を用いた。これはWeb一億ページに対して構文情報を付与したコーパスになっている。本実験ではこの中から2つの形態素のみからなる文節を500万文節抽出し、これをオートエンコーダの学習データとして用いた。単語ベクトルの評価用データセットも同様にTSUBAKIから抽出した。TSUBAKIに含まれる単語のうち、同じ品詞を持つ単語ペアおよび異なる品詞を持つ単語ペアをそれぞれ1000ペア抽出し、評価用データとした。

単語ベクトルを乱数で初期化するには、平均0.0標準偏差0.01のガウス分布に従う乱数を用い、重み行列の初期化にも同様の乱数を用いた。また学習率は 10^{-3} の一定値とし、正則化は係数 10^{-6} のL2正則化を用いた。単語ベクトルの次元は25次元とした。

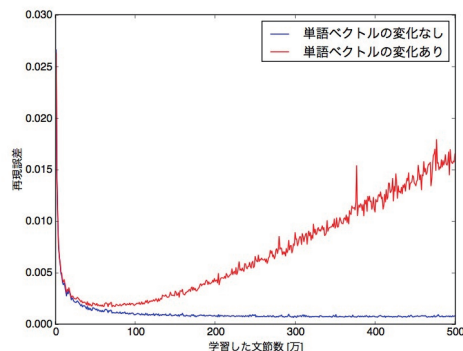


図 3: 言語モデルによって学習した単語ベクトルを初期値として学習した場合の結果

設定	正例	負例
ランダム・初期値	0.0022	-0.0045
ランダム・学習後	-0.0023	-0.0069
言語モデル・初期値	0.5042	0.3237
言語モデル・学習後	0.5087	0.3260
未知語・初期値	0.0076	0.0134
未知語・学習後	0.0085	0.0099

表 1: 単語ペアに対する各単語ベクトルにおける類似度

3.2 実験結果と考察

初期値を乱数で与えた場合の結果を図2に、あらかじめ言語モデルによって学習した単語ベクトルを初期値として与えた場合の結果を図3に、各実験によって学習された単語ベクトルについて同じ品詞の単語および異なる品詞の単語のペアに対するコサイン類似度を計算した結果を表1に示す。

学習後の単語ベクトルについて、いずれの実験設定においても表1ではほぼ変化がみられないことから、オートエンコーダの学習において単語ベクトルを学習することは困難であると考えられる。また、いずれの初期値設定においても単語ベクトルの学習を行った場合に再現誤差が増加していることから、出力層で入力層を再現するオートエンコーダの枠組みにおいて再現対象である入力層を再現誤差によって変化させてしまうことが推測される。そして図2における学習曲線の不自然な推移については、種々の追加調査から、入力ベクトルの大きさに対して学習率が相対的に高すぎる場合にこのような現象が起こることが推測されている。以上のような結果から、単純なオートエンコーダの学習においては、単語ベクトルの学習は行うべきでないといえる。

4 オートエンコーダによる言語モデル学習

本節では、単語ベクトルを変化させることが有益となるのは明示的な教師信号が存在する場合のみであることを検証するため、オートエンコーダに新たにニューロンを追加し、オートエンコーダとNeural Language Model(NLM)の結合学習を行うようなモデルを提案する。

4.1 オートエンコーダとNeural Language Modelの結合モデル

Neural Language Model[1]とは、単語列を入力としてそれが言語として自然かどうかの二値分類をニューラルネットワークを用いて行うモデルである。言語として

自然であるとはそれが実際の文章で使われ得るということを示す。例えば文節「私 が」は自然であり、文節「私 ます」は不自然である。我々は NLM とオートエンコーダを結合的に学習することによって、オートエンコーダの重みの学習と単語ベクトルの学習を効果的に行うようなモデルを提案する。

オートエンコーダと NLM の結合モデルの例を図 4 に示す。ここで、NLM の重み行列を W_p とする。NLM の出力層はロジスティックシグモイド関数 $g(a)$ を用いて次のように表される：

$$t = g(W_p \cdot p) \quad (4)$$

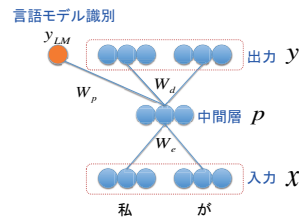


図 4: オートエンコーダと Neural Language Model の結合モデル

NLM は入力としてオートエンコーダの中間層の値を受け取り、NLM の出力はオートエンコーダに入力された単語列が言語として自然かどうかの二値分類に対する解析結果を表す。出力は、入力が自然である場合は 1 に近い値を、不自然である場合は 0 に近い値を示すことを目指す。

x をオートエンコーダの入力、 t を教師信号、 y_{LM} を NLM の出力とし、NLM の誤差関数 E^{pre} を交差エントロピー誤差関数を用いて次のように定義する：

$$E_x^{pre} = -t \log y_{LM} - (1 - t) \log(1 - y_{LM}) \quad (5)$$

右下の添字はこの誤差関数がオートエンコーダの入力 x に依存していることを表す。この誤差を以下**予測誤差**と表現することにする。さらに、予測誤差と再現誤差との重み係数を α 、オートエンコーダの入力 x から Pseudo-Negative Sampling[3] を用いて生成した負例を x^- としたとき、結合モデル全体の誤差関数として次のような関数 E を定義する：

$$E(x) = \alpha(E_x^{pre} + E_{x^-}^{pre}) + (1 - \alpha)E^{rec}(x) \quad (6)$$

学習は誤差関数 E を確率的勾配降下法を用いて最小化することで実行され、この時の誤差関数に対する各パラメータの勾配はニューラルネットワークにおける誤差逆伝播のテクニックを用いて計算される。

4.2 行った実験

提案したモデルによって単語ベクトルが学習されることを確認するために実験を行った。単語ベクトルの初期値は 3 章で扱った実験と同様の乱数を用いた。3 章の実験と同様に Tsubaki から抽出した 2 形態素からなる文節 500 万個を対象とし、結合モデルの学習における再現誤差と識別誤差の推移を観察した。このとき重み係数は $\alpha = 0.1$ とし、その他のパラメータは 3 章の実験と同様の設定とした。また学習後の単語のベクトルは 3 章と同じ方法で評価を行った。

4.3 結果と考察

実験における言語モデルの予測誤差の推移を図 5 に、オートエンコーダの再現誤差の推移を図 6 に、3 章の実験結果と今回の実験結果との入出力のコサイン類似度の比較を図 7 に、学習された単語ベクトルにおける関

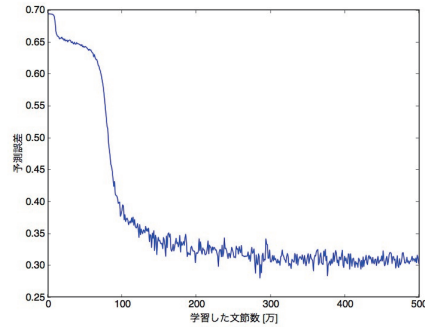


図 5: 結合モデルの学習における予測誤差の推移

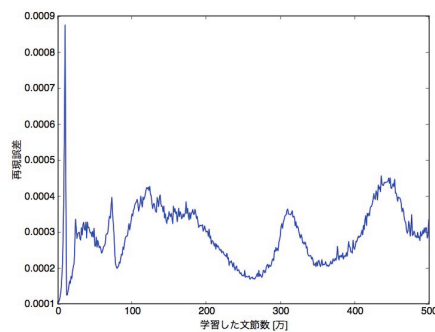


図 6: 結合モデルの学習における再現誤差の推移

連語・非関連語に対するペアに対するコサイン類似度を表 2 に示す。

学習が進むにつれ予測誤差が低下していることや、表 2 の結果から、今回のモデルによって単語ベクトルが学習できていることが確認できる。また、図 7 の結果より、オートエンコーダの学習という観点においても結合モデルが通常モデルと比肩する結果が得られていることが分かる。

5 URA に対する言語モデル結合学習の適用

5.1 URA に対する Neural Language Model の結合学習

本節では 4 節で扱ったオートエンコーダによる言語モデル学習の枠組みを拡張し、URA による言語モデル学習を提案する。提案モデルの一例を図 8 に示す。図中の重み行列の定義は 4 節と同様とする。

URA では文を構成する単語ベクトルが入力となる。ネットワークは中心の層に関して対称となるように構築され、その対称構造の入力側の各中間層はいくつかの単語ベクトルの情報をまとめたものと見なせ、また中心の層は文全体の情報をまとめたものと考えられる。そこで、それらの中間層のそれぞれに 4 章と同様のニューロン層を追加し、URA と言語モデルとの結合学習を実現する。このとき、言語モデルの学習では、その中間層が表す単語列が言語として自然かどうかを識別する。モデル全体の学習は URA の再現誤差と各

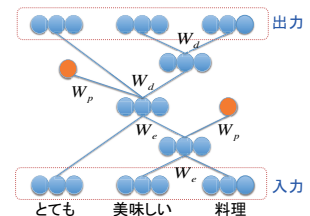


図 8: URA への結合モデルの適用

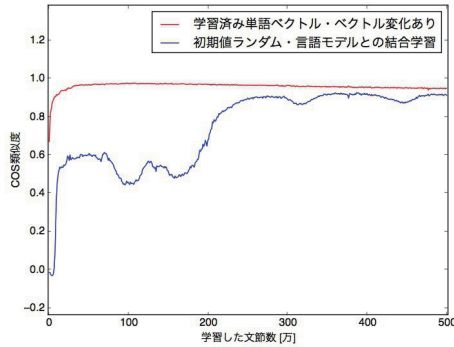


図 7: 入出力のコサイン類似度の比較

設定	正例	負例
結合モデル	0.5367	0.2823

表 2: 結合モデル学習後の単語ベクトルにおける単語ペアに対する類似度

言語モデルの予測誤差を重み付け平均した誤差関数を最小化することで行う。

5.2 実験

文の構文木を作成する際に、機能語などの文の意味に大きく影響を与えない語を文から除外した上で構文木を作成した。解析器には Cabocha[6] を使用した。機能語などを除外することが今回の学習の上で有用であることは、事前実験によって確かめられている。言語モデル学習の正例および負例は 4 節の実験と同様の方法で生成した。学習コーパスには TUBAKI[5] から約 2000 万文を抽出して使用した。その際、学習を容易にするために入力ベクトルの数が 8 個以上になる文は学習の対象から除外している。

また、得られた単語ベクトルがどの程度単語の意味的特性を捉えているかを評価するため、日本語 WordNet[2] から同一品詞である類義語のペアと類義語でないペアを 30000 組作成し、各ペアに対してコサイン類似度を計測した。ただし、学習コーパスの規模が不十分であることが予想されたため、各ペアを構成する単語にはコーパス中で最頻出の 5000 語のみを対象とした。

5.3 結果と考察

結果を表 3 に示す。類似語ペアのコサイン類似度の方が大きいことから、提案モデルにより学習された単語ベクトルが単語の意味的特性をある程度捉えられていることが確かめられた。

また、得られた単語ベクトルの意味的特性を視覚的に観察するために、極性をもった単語をいくつか選び、それらの単語ベクトルを主成分分析を用いて可視化した。その結果の一部を図 9 に示す。この図から、極性が負の単語 (赤) と極性が正の単語 (青) がある程度分離されていることが確認できた。

6 結論

本稿では、オートエンコーダの学習における単語ベクトルの学習に注目し、調査を行った。その結果、オートエンコーダの学習においては単語ベクトルは意味のある情報を獲得できず、単語ベクトルを学習することがオートエンコーダ自体の学習に悪影響を及ぼすことが確

設定	類義語	非類義語
URA による言語モデル学習	0.7554	0.6619

表 3: 類義語・非類義語のペアに対する類似度

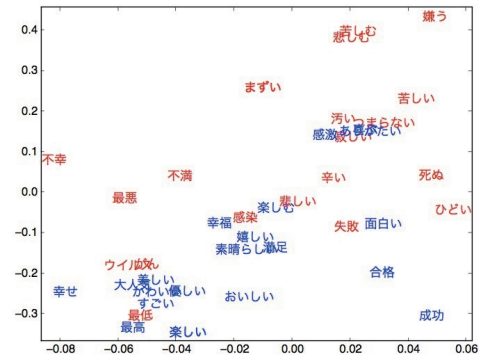


図 9: 単語ベクトルの可視化

認できた。また、オートエンコーダと Neural Language Model の結合学習モデルを提案し、言語モデルの教師信号を与えることでオートエンコーダの学習と同時に単語ベクトルも学習できることを確かめた。これらは、オートエンコーダを自然言語処理に応用する上で基礎的かつ重要な知見であると考えられる。

そして、提案モデルが Unfolding Recursive Encoder に対しても適用可能であることを経験的に示した。今後は URA に対する言語モデルの結合学習に対する詳しい検証や、これらのモデルのタスク適用などを検討していきたい。

謝辞

本研究は、東北大学工学部 情報知能システム総合学科「Step-QI スクール」の支援を受けた。

参考文献

- [1] R. Collobert and J. Weston. A unified architecture for natural language processing: Deep neural networks with multi-task learning. In *ICML*, 2008.
- [2] K. Uchimoto M. Utiyama H. Isahara, F. Bond and K. Kan-zaki. Development of the japanese wordnet. In *LREC'08*, 2008.
- [3] D. Okanohara and J. Tsujii. A discriminative language model with pseudo-negative samples. In *ACL*, pp. 73–80, 2007.
- [4] R. Socher and E. H. Huang and J. Pennington and A. Y. Ng and C. D. Manning. Dynamic Pooling and Unfolding Recursive Autoencoders for Paraphrase Detection. In *NIPS*. 2011.
- [5] K. Shinzato, T. Shibata, D. Kawahara, and S. Kurohashi. Tsubaki: An open search engine infrastructure for developing information access methodology. *JIP*, Vol. 20, No. 1, pp. 216–227, 2012.
- [6] 工藤拓, 松本裕治. チャンキングの段階適用による日本語係り受け解析. Vol. 43, No. 6, pp. 1834–1842, 2002.