修　士　学　位　論　文

教師あり学習により構築した
語義の分散表現による語義曖昧性解消

平成29年度

茨城大学大学院理工学研究科

情報工学専攻
山木　翔馬

平成 29 年度茨城大学大学院理工学研究科情報工学専攻　修士学位論文

教師あり学習により構築した
語義の分散表現による語義曖昧性解消

氏名：16NM724R 山木 翔馬
指導教員：新納 浩幸 教授

論文要旨

本論文では単語の分散表現と教師データを用いた教師あり学習による語義の分散表現を構築する手法を提案し, 語義曖昧性解消 (Word Sense Disambiguation; WSD) の評価実験によって従来の教師なし学習により構築した語義の分散表現との比較を行い, その考察を述べる.

深層学習の手法を利用して単語の意味を低次元の密なベクトルで表現した分散表現 (Word Embeddings) は, 今や自然言語処理の様々な分野で有効な結果を残している .近年ではさらに, 語義ごとの分散表現 (Multi-sense Embeddings) を構築する研究がなされており , いくつかのタスクでは単語の分散表現よりも語義の分散表現を用いる方が良い結果を出すことを示す研究もある.

現在研究されている語義の分散表現の構築手法の多くは, 単語の分散表現を構築する従来の教師なしの言語モデルを拡張したものがほとんどである. そのため, コーパス中に現れる語義の出現頻度の情報を得ることができないという問題点がある.

本論文では教師データを用いて, 語義の出現頻度を考慮した語義の分散表現の構築手法を提案する. 具体的には, あらかじめ学習させた単語の分散表現を用いて教師データ中の各語義に対する用例の文脈ベクトルを求め, 文脈ベクトルと語義の出現頻度に基づいて単語の分散表現を語義の分散表現に分解する.

実験では単語の分散表現として Wikipedia の日本語記事を学習させたものと国立国語研究所が作成した分散表現 nwjc2vec を用いた. 教師データとして SemEval-2 の日本語辞書タスクのデータを用い, 語義の分散表現を構築し ,WSD による評価実験を行った. 実験の結果, 作成した語義の分散表現による WSD の精度向上は確認できなかったものの, 語義の分散表現に類似する分散表現を持つ単語を分析したところ, 語義の分散表現が正しく作られていることが分かった.

また教師なし学習による語義の分散表現の構築手法として Multi Sense Skip-gram (MSSG) モデルと Non-Parametric Multi Sense Skip-gram (NP-MSSG) モデルを実装し, 語義の分散表現を構築した. これらの教師なし学習により構築した語義の分散表現と我々の提案手法により構築した語義の分散表現を用いて SemEval-2 の日本語辞書タスクデータのうち名詞である 20 単語を対象に WSD の実験を行ったところ，MSSG モデルが他の 2 つの教師なしの手法に比べて高い正解率を出すことが確認できた. さらに対象単語ごとの正解率を見ると, 語義の出現頻度の差が大きい単語に対しては提案手法が高い正解率となり , 語義の出現頻度の差が小さい単語に対しては MSSG モデルが高い正解率となることが分かった. また NP-MSSG モデルは他の手法に比べ正解率は低かったものの, 教師データ中に出現しない語義の分散表現を構築できる可能性があることが分かった.

Master's Thesis in Scholastic 2017, Major in Computer and Information Sciences,
Granduate School of Science and Engineering, Ibaraki University

Multi Sense Embeddings by Supervised Learning for Word Sense Disambiguation

Author : Shoma Yamaki (16NM724R)
Adviser : Prof. Hiroyuki Shinnou

ABSTRACT

In this paper, we propose the method to construct distributed representation for each sense of ambiguous word from distributed representation for word and training data, and describe a consideration comparising multi-sense embeddings by superviesed learning and unsupervised learning by experiments of Word Sense Disambiguation (WSD).

Word embeddings which expresses the meanings of a word in a low-level dense vector leaned by deep learning has effective results in various fields of natural language processing. In recent years, further studies have been made to construct a distributed represetation for each word sense, and several studies have shown that using multi-sense embeddings is better than using word embeddings in some tasks.

In this paper, we propose the mehod to construct multi-sense embeddings considering the frequency of the word meanings. Specifically, using word embeddings learned in advance, context vectors of sentences for each word sense in training data are calculated, and word embeddings are decomposed into multi-sense embeddings based on the context vector and the frequency of word senses.

In the experiments, we use embeddings learned from Wikipedia's Japanese article and nwjc2vec by National Institute for Japanese Language and Linguistics. We used SemEval-2 Japanese task data as training data to built multi-sense embeddings, and performed experiments with WSD. As a result of the experiments, we could not improve the accuracy of WSD using multi-sense embeddings, however, when we analyzed the words that have embeddings similar to multi-sense embeddings , we found that the multi-sense embeddings were correctly constructed.

Furthermore, we implemented Multi Sense Skip-gram (MSSG) Model and Non-Parametric Multi Sense Skip-gram (NP-MSSG) model as a method as unsupervised methods. We conducted WSD experiments using these models and our proposed method, and confirmed that MSSG model gives higher accuracy than the other two methods.

Moreover, according to the accuracy for each word, we found that the proposed method gives a high accuracy for words with a large difference in the frequency, and the MSSG model gives a high accuracy for words with a small difference in the frequency. In addition, although the NP-MSSG model is inferior in the accuracy to other methods, we found it's possible to construct embeddings of unknown sense in training data.

目次

表目次

図目次

第 1 章

序論

1.1 概要

本論文では単語の分散表現と教師データを用いて語義の分散表現を構築する手法を提案する．

深層学習 (Deep Learning) の手法を利用して単語の意味を低次元の密なベクトルで表現した分散表現 (Word Embeddings) は，今や自然言語処理の様々な分野で有効な結果を残しているが，近年では分散表現の学習手法を拡張して語義ごとの分散表現を構築する研究がなされており，いくつかのタスクでは単語の分散表現よりも語義の分散表現を用いた方が良い結果が得られることを示す報告がある．

現在研究されている語義の分散表現の構築手法の多くは，単語の分散表現を構築する従来の教師なしの言語モデルを拡張したものがほとんどである．そのためコーパス中に現れる語義数や語義の出現頻度の情報を得ることができないという問題点がある．

本論文では教師データを用いて，語義数と語義の出現頻度を考慮した語義の分散表現を構築する手法を提案する．具体的には，あらかじめ得た単語の分散表現を用いて教師データ中の各語義に対する用例の文脈ベクトルを求め，文脈ベクトルと語義の出現頻度に基づいて単語の分散表現を語義の分散表現に分解する．

実験では単語の分散表現として Wikipedia の日本語記事を word2vec[*1]で学習させたものと国立国語研究所が作成した分散表現 nwjc2vec[8] を用いた．教師データは SemEval-2 の日本語辞書タスクのデータを用いて語義の分散表現を構築し，語義曖昧性解消 (Word Sense Disambiguation) による評価実験と，語義の分散表現の類似単語による評価を行った．実験の結果，作成した語義の分散表現を WSD に用いる実験では精度の向上は確認できなかったが，構築した語義の分散表現と類似する分散表現を持つ単語を分析したところ，語義の分散表現が

[*1] https://code.google.com/archive/p/word2vec

正しく作られていることが分かった．

また，教師なし学習による語義の分散表現を構築する一般的な手法として知られる Multi Sense Skip-gram (MSSG) モデル [1], MSSG モデルを拡張した Non-Parametric Multi Sense Skip-gram (NP-MSSG) モデル [4] を用いて語義の分散表現を構築し，我々の提案手法による語義の分散表現と比較したところ，MSSG モデルによる語義の分散表現を用いた WSD の正解率の方が良い結果を出すことが分かった．

さらに対象単語別に正解率を見ると，語義の出現頻度の差が大きい単語は提案手法が良い正解率となり，語義の出現頻度の差が小さい単語は MSSG モデルが良い正解率を出した．

1.2　本論文の構成

本論文では，はじめに理論とその手法について紹介する．第 2 章で語義曖昧性解消 (Word Sense Disambiguation; WSD) について，第 3 章では深層学習を利用した分散表現の構築モデルについて述べる．　そして教師データあり学習による語義の分散表現の構築について述べ，その実験と結果の考察を行う．第 4 章では教師データを使い単語の分散表現から語義の分散表現を構築する手法を提案する．第 5 章では提案手法により構築した語義の分散表現による WSD の実験の内容と結果を示し，第 6 章で実験結果について考察する．最後に第 7 章で結論を述べる．

第 2 章

語義曖昧性解消

2.1 概要

文章中に出現した単語が複数の意味（ 語義）を持っている場合，その単語が表す語義を一意に識別するタスクのことを語義曖昧性解消 (Word Sense Disambiguation; WSD) という．

例えば「 相手」という単語には表 2.1 のように 4 つの語義が存在する．

表 2.1 「 相手」の語義

1	一緒にする相手．相棒．仲間．
2	対抗すること．また，対抗する人．
3	付き合うこと．世話をすること．また，その人．
4	はたらきかける対象．

ここで以下のように「 相手」に関する例文が与えられたとする．

1. 次の試合の相手は強豪校のエースだ．
2. 私が結婚する相手は同じ職場の同僚です．

例文 1 での「 相手」は表 2.1 の 2 番目の語義「 対抗すること．また，対抗する人」に該当するが，例文 2 での「 相手」は 3 番目の語義「 付き合うこと．世話をすること．また，その人．」となる．このように，与えられた文章中に出現する単語の語義を一意に識別することを語義曖昧性解消と呼ぶ．

2.2　一般的な手法

WSD の手法は大きく分けて教師あり学習と教師なし学習がある．教師あり学習とは WSD の対象単語を含んだ用例とその単語の正解となる語義がラベルとして付与されたデータ (ラベル付きデータ，または教師データ) を用いて分類器を学習する手法である．一方，教師なし学習はラベルが付与されていない用例のみのデータ (ラベル無しデータ) から分類器を学習する手法である．教師あり学習による WSD ではラベルが付与されている単語に対しての分類器しか作れないのに対し，教師なし学習では用例中に出現する単語全てに対して分類器を作れるため，all-words WSD として区別されることもある．ここでは教師あり学習による WSD の手法について説明する．

教師あり学習による WSD の手法では，入力された用例から対象単語の周辺単語をベクトルで表現し，そのベクトルが入力されたときに正解となるラベルを出力するような分類器を学習させる．

周辺単語をベクトル化する方法として最も単純なものに Bag-of-Words(BoW) というものがある．これは文章中に単語が含まれているかどうかだけを考え，単語の並び順などは考慮しないモデルである．具体的にはコーパス中に現れる単語が V 種類あるとすれば，各単語に $1 \sim V$ までの id を付与する．そして周辺単語をベクトルにする際には，周辺単語の id に相当するベクトルの次元の値が 1 となり，その他の値は 0 になるようにする．例えば先にの示した例文「 次の試合の相手は強豪校のエースだ」において，単語「 相手」の周辺単語をベクトル化した場合，そのベクトルは V 次元で,「 次」「 の」「 試合」「 は」「 強豪校」「 エース」「 だ」のそれぞれの単語の id 番目の次元だけが 1 で他の次元は 0 となる．

WSD の手法はこのように対象単語の文脈（ 周辺単語）をベクトル化し，その文脈ベクトルを用いて Support Vector Machine (SVM) などにより分類器を学習させ，識別に用いる．

第 3 章

分散表現

3.1 概要

2.2 節で述べたように，WSD のアプローチとしては対象単語の周辺単語をベクトルで表現する手法が多い．2.2 節では単語をベクトル化する手法として BoW について説明したが，近年深層学習の手法を利用して単語の意味を低次元の密なベクトルで表現する分散表現が注目されている．分散表現により単語の意味をベクトル表現した場合，単語の間の距離が BoW を用いるよりもより正確に求められるようになる．そのため，様々な自然言語処理のタスクに分散表現が利用され，有効な結果を残している．

単語の分散表現の構築には Feedforward Neural Network Language Model や Reccurent Neural Network Language Model などのニューラルネットワークに基づく言語モデルを用いる方法が多く研究されているが，中でも Mikolv らが提案した skip-gram モデルと Conteniouse Bog-of-Words (CBoW) モデルは，言語モデルを単純化することでベクトル表現の学習の高速化に成功した [3]．これらのモデルをツール化した word2vec は分散表現を獲得する手段として広く使われている．

単語の分散表現の学習モデルを拡張することで語義ごとの分散表現を構築する研究も多くされている． Huang らの研究では 1 つの単語にあらかじめ指定した語義数のベクトルを与えるモデルとして skip-gram モデルを拡張した Multi Sense Skip-gram (MSSG) モデルを提案している． Neelakantan らは MSSG モデルをさらに拡張し，語義の数を自動で決める Non-Parametric Multi Sense Skip-gram (NP-MSSG) モデルを提案している． Li らの研究では語義の分散表現が実際の自然言語処理で有効であることを示している [2]．この研究では MSSG モデル，NP-MSSG モデルによって構築された語義の分散表現を自然言語処理の様々なタスクに利用するためのパイプラインアーキテクチャを提案し， part-of-speech tagging, semantic relation identification, semantic relatedness のタスクにおいて語義の分散表現が有効であることを示した．

本章では単語の分散表現を構築する言語モデルとして広く使われている skip-gram モデルと，それを拡張した MSSG モデルについて説明する．

3.2　Skip-gram モデル

3.2.1　目的関数

skip-gram モデルは，ある単語が与えられた時にその周辺の単語を予測する言語モデルである．以下のような文（単語列）があるとする．

$$w_{t-b}, w_{t-b+1}, \cdots, w_{t-1}, w_t, w_{t+1}, \cdots, w_{t+b-1}, w_{t+b}$$

ここで対象とする単語 w_t と w_t の文脈

$$c_t = \{w_{t-b}, w_{t-b+1}, \cdots, w_{t-1}, w_{t+1}, \cdots, w_{t+b-1}, w_{t+b}\}$$

を考え，単語 w_t から文脈 c_t が出現する確率 $p(c_t|w_t)$ をモデル化し，対数尤度

$$L = \sum_{t=1}^{T} \log p(c_t|w_t) \tag{3.1}$$

を最大化する．T はコーパスの単語数（語彙数ではなく，単語列の長さ）である．式 3.1 にはパラメータとして w_t の分散表現が含まれているため，これを最大化することで「良い」分散表現を得ることができる．

skip-gram モデルでは $p(c_t|w_t)$ を以下のようにモデル化する．

$$p(c_t|w_t) = \prod_{c \in c_t} p(c|w_t)$$

これを式 3.1 に代入すると，以下が得られる．

$$L = \sum_{t=1}^{T} \sum_{c \in c_t} \log p(c|w_t) \tag{3.2}$$

次に $p(c|w_t)$ を以下のように条件付き確率として softmax 関数を使ってモデル化する．

$$p(c|w_t) = \frac{\exp(\boldsymbol{v}_c \cdot \boldsymbol{v}_{w_t})}{\sum_{w' \in V} \exp(\boldsymbol{v}_c \cdot \boldsymbol{v}_{w'})} \tag{3.3}$$

ここで $\boldsymbol{v}_w$ が単語 w の分散表現であり，V はコーパス中に現れる全種類の単語となる．つまりこの式は単語 w_t とその単語 c の共起する確率を表しており，その確率は単語 w_t の分

散表現 $\boldsymbol{v}_{w_t}$ と単語 c の分散表現 $\boldsymbol{v}_c$ の内積で計算している．これを式 3.2 に代入すると以下の式が得られる．

$$L = \sum_{t=1}^{T} \sum_{c \in c_t} \left((\boldsymbol{v}_c \cdot \boldsymbol{v}_{w_t}) - \log \sum_{w' \in V} \exp(\boldsymbol{v}_c \cdot \boldsymbol{v}_w^{w'}) \right) \tag{3.4}$$

この分散表現をパラメータとした目的関数を最大化することで単語の分散表現を求めることができる．しかし，式 3.4 の

$$\sum_{w' \in V} \exp(\boldsymbol{v}_c \cdot \boldsymbol{v}_{w'})$$

を見ると，内積の計算を V 回繰り返していることが分かる．一般的なコーパスでは語彙数 V は $10^5 \sim 10^7$ ほどのオーダーであるため，この部分の計算コストが高すぎるという問題に直面する．次節ではこの部分の計算を高速化する Negative Sampling という手法について説明する．

3.2.2 Negative Sampling

Negative Sampling の目的は式 3.4 を計算コストの低い式で近似することである．そのためにまず，ある単語のペア (w_t, c) がコーパスから生成されたかの確率を考え，それを $p(D = 1|w_t, c)$ とし，逆にコーパスから生成されていない確率を $p(D = 0|w_t, c)$ とする．これらの確率を用いて，式 3.4 を以下のように近似する．

$$p(c|w_t) = \frac{\exp(\boldsymbol{v}_c, \boldsymbol{v}_{w_t})}{\sum_{w' \cdot \boldsymbol{v}_{w'})}} \approx p(D = 1|c, w_t) \prod_{c' \in Ng} p(D = 0|c', c) \tag{3.5}$$

これは softmax 関数の近似を行う Noise Contractive Estimation という手法である．また上式の Ng はコーパス中からノイズ分布と呼ばれるある分布にしたがってサンプリングした k 個の単語の集合である．

つぎに $p(D = 1|w_t, c)$ をシグモイド関数として以下のように定義する．

$$p(D = 1|c, w_t) = \sigma(\boldsymbol{v}_c, \boldsymbol{v}_{w_t})$$

すると $p(D = 0|c, w_t)$ は確率の定義から

$$p(D = 0|c', c) = 1 - p(D = 1|c', c) = 1 - \sigma(\boldsymbol{v}_{c'} \cdot \boldsymbol{v}_c) = \sigma(-\boldsymbol{v}_{c'} \cdot \boldsymbol{v}_c)$$

となる．これを式 3.4 に代入すると以下のようになる．

$$L = \sum_{t=1}^{T} \sum_{c \in c_t} \left(\log \sigma(\boldsymbol{v}_c \cdot \boldsymbol{v}_{w_t}) + \sum_{c' \in Ng} \log \sigma(-\boldsymbol{v}_{c'} \cdot \boldsymbol{v}_c) \right)$$

3.2.3　ニューラルネットワークによるモデル化

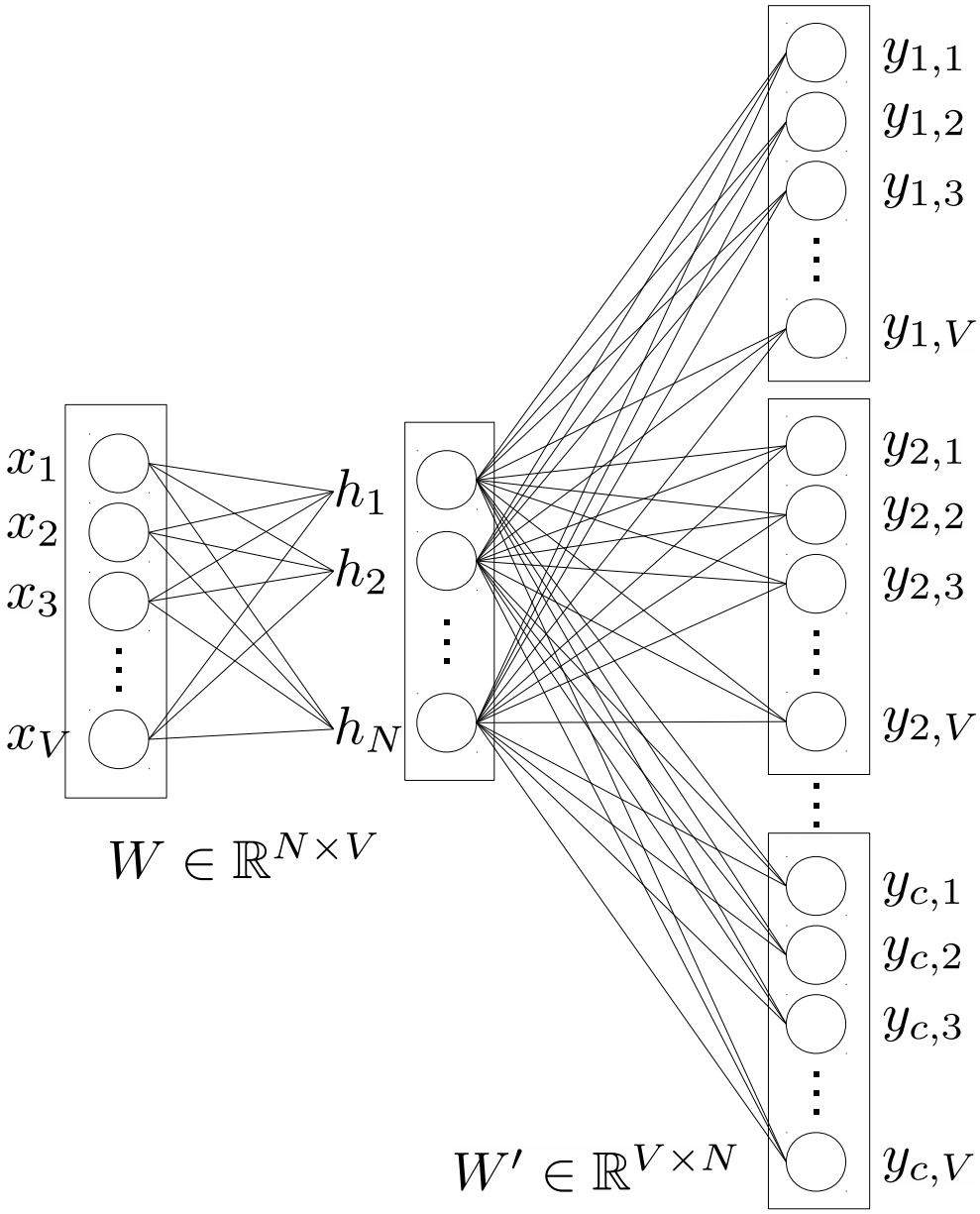

図 3.1　skip-gram モデルのネットワーク図

skip-gram モデルのネットワーク図を図 3.1 に示す．

入力層のベクトル $\boldsymbol{x}$ は V 次元となる．ベクトル x は入力される単語に対応する次元

だけ 1 となり，その他の次元は 0 となる one-hot-vector である．また隠れ層の次元数はは学習したい分散表現の次元数となっており，入力層から隠れ層への重み $\boldsymbol{W}$ が分散表現となる．そして出力層はコンテキストサイズ（対象単語の前後何単語までを周辺単語として取るか）の数だけのベクトルで，それぞれのベクトルが入力層と同じく，出力する単語に相当する one-hot-vector となる．

出力層の活性化関数は softmax 関数を使うため，最終的な出力値は式 3.3 と同じになることが分かる．

3.3　Multi Sense Skip-gram モデル

3.3.1　概要

次に Huang らにより提案された Multi Sense Skip-gram (MSSG) モデルについて説明する．前述したように skip-gram モデルは一つの単語に対して一つの分散表現ベクトルを学習するモデルであった．一方，MSSG モデルは skip-gram モデルを拡張することで語義ごとの分散表現を学習することができる．具体的には，周辺単語のコンテキストベクトルから，平均コンテキストを計算し，あらかじめ用意されている語義候補の中から最もコンテキストの類似した語義を選択する．そして選択された語義ベクトルに従って周辺コンテキストが決定していると仮定し，skip-gram モデルと同じように学習する．　MSSG モデルのネットワーク図を 3.2 に示す．

3.3.2　目的関数

MSSG モデルの目的関数は以下のようになる．

$$L = \sum_{(w_t,c)\in D^+} \sum_{c\in c_t} \log p(D=1|\boldsymbol{v_s}(w_t,s_t),\boldsymbol{v_g}(c)) + \sum_{(w_t,c'_t)\in D^-} \sum_{c'\in c'_t} \log p(D=0|\boldsymbol{v_s}(w_t,s_t),\boldsymbol{v_g}(c')$$

$\boldsymbol{v_s}(w_t,s_t)$ は単語 w_t の語義 s_t の分散表現を表す．また $\boldsymbol{vg}(c)$ は単語 c のグローバルな分散表現を表している．この式を見ると，語義の分散表現 $bmvs(w_t,s_t)$ を用いている以外は skip-gram モデルの目的関数と同じであることが分かる．　次に語義 s_t をどのように選択するのかについて説明する．

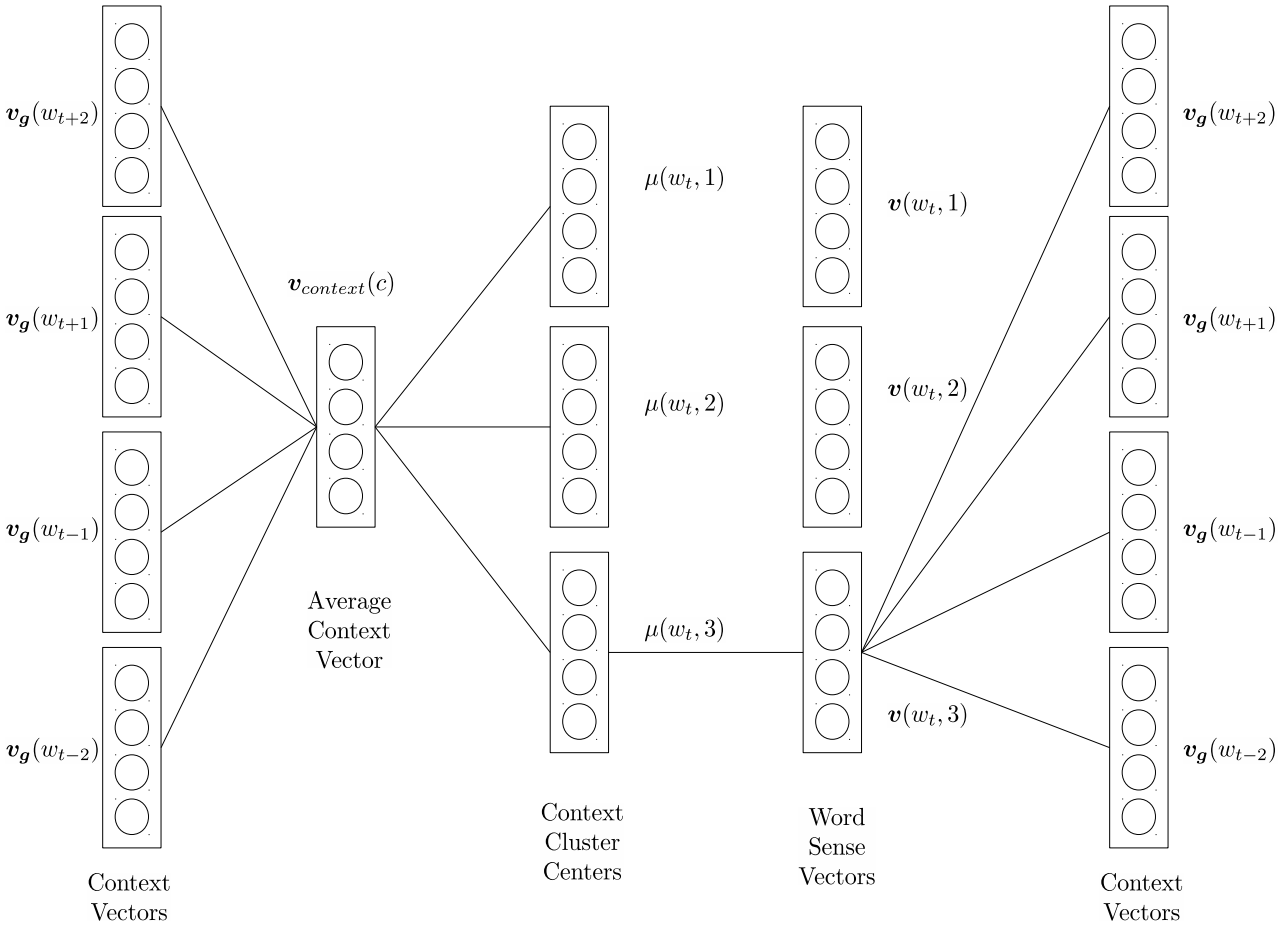

図 3.2　MSSG モデルのネットワーク図

3.3.3　語義の選択

単語 w_t とその文脈

$$c_t = \{w_{t-b}, w_{t-b+1}, \cdots, w_{t-1}, w_{t+1}, \cdots, w_{t+b-1}, w_{t+b}\}$$

が与えられたとき，周辺単語の文脈ベクトルを

$$\boldsymbol{v}_{context}(c_t) = \frac{1}{2b} \sum_{c \in c_t} \boldsymbol{vg}(c)$$

で表す．この時の単語 w_t の語義 s_t は

$$s_t = \underset{k=1,2,\ldots,K}{\arg\max} \{sim(\mu(w_t, k), \boldsymbol{v}_{context}(c_t))\}$$

となる．K はあらかじめパラメータとして与えられる語義の数である．

第 4 章

教師あり学習による語義の分散表現の構築

4.1 教師データの利用

語義の分散表現を構築する上で語義の出現頻度は有力な情報となるが，3.3 で述べたように MSSG モデルのような教師なし学習による手法では語義の出現頻度が得られないため，対象単語の文脈ベクトルとあらかじめ設定した語義の文脈ベクトルとの類似度によって語義を推定している．

教師データを用いる場合も文脈ベクトルは有効な情報である．対象単語 w_t に対する教師データ中の i 番目の用例

$$w_1, w_2, \cdots, w_t, \cdots, w_m$$

の文脈ベクトル $\boldsymbol{u_i}$ は, 単語 w の分散表現ベクトルを $\boldsymbol{v}(w)$ としたとき，

$$\boldsymbol{u_i} = \frac{\sum_{i=1}^{m} \boldsymbol{v}(w_i)}{m}$$

で表される．また教師データ中の語義 c_i に対する用例の集合を C_i としたとき，語義 c_i の文脈ベクトル $\boldsymbol{u_{c_i}}$ は用例の文脈ベクトルの平均

$$\boldsymbol{u_{c_i}} = \frac{1}{|C_i|} \sum_{i \in C_i} \boldsymbol{u_i}$$

となる．

ここで語義の文脈ベクトル $\boldsymbol{u_{c_i}}$ を語義の分散表現ベクトルとして利用する方法も考えられるが，この方法では語義の出現頻度が反映されていない．本論文では語義の文脈ベクトル $\boldsymbol{u_{c_i}}$ と語義の出現頻度を考慮した語義の分散表現の構築法を提案する．

4.2　提案手法

単語の分散表現は，その単語が持つ語義の分散表現の和になると仮定する．つまり，単語 w の分散表現を $\boldsymbol{v}(w)$，語義の分散表現を $\boldsymbol{e_1}, \boldsymbol{e_2}, \boldsymbol{e_3}$ としたとき，

$$\boldsymbol{v}(w) = \boldsymbol{e_1} + \boldsymbol{e_2} + \boldsymbol{e_3}$$

が成り立つとする．このとき，これらのベクトルの k 次元目の値においても

$$w^{(k)} = e_1^{(k)} + e_2^{(k)} + e_3^{(k)}$$

が成り立つ．ここで提案する手法は，単語の分散表現と語義の文脈ベクトル $\boldsymbol{u_{c_1}}, \boldsymbol{u_{c_2}}, \boldsymbol{u_{c_3}}$ の k 次元目の値に注目し，$u_1^{(k)}, u_2^{(k)}, u_3^{(k)}$ の値の差が小さければ

$$e_i^{(k)} = \frac{|C_i|}{|C_1| + |C_2| + |C_3|} w^{(k)}$$

と，$w^{(k)}$ を語義の出現頻度で分解し，逆に $u_1^{(k)}, u_2^{(k)}, u_3^{(k)}$ の差が大きければ

$$e_i^{(k)} = w^{(k)}$$
$$e_j^{(k)} = 0$$

と，一つの語義の分散表現の k 次元目の値に単語の分散表現の k 次元目の値を入れ，他の語義の分散表現の k 次元目の値は 0 とする．

次に値の差の大小を比較するための基準値を設定する．教師データの総数を N，教師データ中の再頻出語義 (Most Frequent Sense; MFS) の数を M としたとき，基準値 tr を

$$tr = \log \frac{M + 0.01}{N + 0.01}$$

とする．tr は決定リストのデフォルト規則適用の証拠の強さを表している．$u_1^{(k)}, u_2^{(k)}, u_3^{(k)}$ の差の大きさ m は

$$m = |\max(u_1^{(k)}, u_2^{(k)}, u_3^{(k)})|$$

とし，次のように正規化する．

$$n = \sum_i |u_i^{(k)} w^{(k)}|$$
$$r = \frac{N}{n}$$
$$d = \log \frac{mr + 0.01}{nr + 0.01}$$

tr と d を比較し，$tr > d$ であれば差が小さい，$tr \leq d$ であれば差が大きいと判定する．

この操作を分散表現のすべての次元に対して行い，得られたベクトル $\boldsymbol{e_1}, \boldsymbol{e_2}, \boldsymbol{e_3}$ を語義の分散表現とする．

第 5 章

実験

5.1 実験設定

5.1.1 提案手法

実験では教師データとして SemEval-2 の日本語辞書タスクのデータを用いる．このデータは 50 個の異なる多義語で構成されており，各単語ごとに訓練データ 50 用例，テストデータ 50 用例が用意されている．

単語の分散表現は分散表現の精度の違いによる比較を行うために，wikipedia の日本語記事を word2vec で学習させた 200 次元のベクトルと，国立国語研究所が作成した 200 次元のベクトル nwjc2vec の 2 つを用意した．

また WSD による評価実験ではベースラインとして文脈ベクトルを素性とした単純な SVM を実装した．

SVM の学習には scikit-learn[*1]の LinearSVC を用いた．

5.1.2 提案手法と教師なしモデルの比較

提案手法の学習で必要な学習済みの単語の分散表現は国立国語研究所が作成した nwjc2vec を使用し，教師なし学習で用いるコーパスは毎日新聞の 1993 年から 1999 年の新聞記事データを用いた．なお，単語の分散表現と語義の分散表現のベクトルは 200 次元とした．

教師なし学習のモデルである MSSG モデルと NP-MSSG モデルによる分散表現の構築では，対象となる単語がコーパス中にある程度の頻度で出現する必要がある．そこで本実験では SemEval-2 の対象単語の内，品詞が名詞である単語 20 単語に対して語義の分散表現を学習

[*1] http://scikit-learn.org/stable/index.html

し，WSD の実験を行った．また教師なし学習により構築した語義の分散表現を用いて WSD を行う場合，構築された語義の分散表現がテストデータのどの語義に対応するか判断できない．そのため本実験において語義の分散表現と語義の対応は，各単語ごとに平均正解率が最も高くなるような組み合わせとした．

なお教師なし学習のモデルは GitHub 上で公開されいるツール[*2]を用いた．

5.2　実験結果

5.2.1　提案手法

はじめに wikipedia の分散表現を用いた提案手法と SVM による WSD の分類結果を表 5.2.1 に示す．

表 5.1　wikipedia の分散表現を用いた提案手法と SVM による WSD の分類結果

手法	平均正解率
MFS	0.688
SVM	0.712
提案手法	0.692

次に nwjc2vec を用いた提案手法と SVM による WSD の分類結果を表 5.2.1 に示す．

表 5.2　nwjc2vec の分散表現を用いた提案手法と SVM による WSD の分類結果

手法	平均正解率
MFS	0.0.688
SVM	0.757
提案手法	0.736

実験の結果，いずれの手法も nwjc2vec を用いた方が wikipedia の分散表現を用いるよりも高い正解率となった．また手法別に見ると提案手法は MFS よりは高い正解率を出すものの，SVM よりはわずかに低い正解率となった．

次に，構築した語義の分散表現がどのような意味を表しているのかを調べるために，単語「 意味」の各語義の分散表現と最も類似度の高い分散表現を持つ単語をそれぞれ 10 単語ずつ

[*2] https://github.com/ExplorerFreda/Multi-Sense-Skip-Gram

求めた．結果を表 5.2.1 に示す．なお，岩波国語辞書における「 意味」の語義は以下の通りである．

2843-0-0-1 その言葉の表わす内容．意義．
2843-0-0-2 表現や行為の意図，動機．
2843-0-0-3 表現や行為のもつ価値．意義．

表 5.3　単語「 意味」の各語義の類似単語

語義 ID	類似単語
2843-0-0-1	趣旨，形式，主旨，自体，面白味，面白み，意図，2843-0-0-3，要素
2843-0-0-2	ニュアンス，比喩，文脈，暗喩，真意，婉曲，恣意，語弊，含意，隠喩
2843-0-0-3	本質，意識，美徳，含意，アイデンティティ，内実，2843-0-0-1，先人，大局，根底，文脈

5.2.2　提案手法と教師なしモデルの比較

提案手法，MSSG モデル，NP-MSSG モデルによる平均正解率を表 6.1 に示す．

実験の結果，いずれの手法も SVM よりも低い正解率となった．特に語義数を自動的に決める NP-MSSG モデルは，他の手法に比べて著しく低い正解率となった．

また，提案手法，MSSG モデル，NP-MSSG モデルによる WSD の各単語に対する正解率を表 5.5 にまとめた．

表 5.4　提案手法，MSSG モデル，NP-MSSG モデルにおける平均正解率

手法	平均正解率
SVM	0.774
提案手法	0.699
MSSG モデル	0.737
NP-MSSG モデル	0.752

表 5.5　提案手法，MSSG モデル，NP-MSSG モデルによる各単語に対する平均正解率

対象単語	SVM	提案手法	MSSG	NP-MSSG
相手	0.68	0.58	0.72	0.48
意味	0.42	0.34	0.38	0.30
可能	0.58	0.50	0.54	0.50
関係	0.96	0.96	0.96	0.68
技術	0.78	0.68	0.76	0.52
経済	0.98	0.98	0.90	0.76
現場	0.76	0.62	0.74	0.42
子供	0.58	0.52	0.60	0.50
時間	0.76	0.72	0.72	0.58
市場	0.58	0.40	0.48	0.34
社会	0.88	0.84	0.82	0.80
情報	0.78	0.64	0.72	0.58
手	0.64	0.46	0.48	0.40
電話	0.80	0.68	0.70	0.56
場合	0.74	0.66	0.78	0.62
場所	0.90	0.90	0.90	0.78
一	0.92	0.92	0.92	0.42
文化	1.00	1.00	1.00	0.88
他	1.00	1.00	1.00	0.80
前	0.66	0.58	0.62	0.52
平均	0.774	0.699	0.737	0.572

第 6 章

考察

6.1　提案手法の改善

WSD による実験では提案手法よりも SVM による分類結果の方が高い正解率となった．その原因として提案手法は教師データから得た語義の文脈ベクトルを用いているため，SVM のように未知データに対する汎用性が低いからだと考えられる．そこで我々は SVM を学習させた際の識別超平面を文脈ベクトルの代わりに用いる手法を実装し，精度の改善が見られるか実験を行った．

つまり，実験で学習させた SVM の識別超平面

$$g(\boldsymbol{x}) = \boldsymbol{v}^T \boldsymbol{x} + b$$

の重み v を用いて語義の分散表現を構築する．実験で述べたように SVM を学習させる際の素性は語義の文脈ベクトルとなっているため，SVM の識別超平面は教師データ中の用例を語義ごとに分離する一つのベクトルである．このベクトルを用いることで汎用性のある語義の分散表現を構築することができると考えられる．この手法を用いた WSD の分類結果を表 6.1 に示す．

表 6.1　SVM の識別超平面を用いた提案手法による WSD の分類結果

手法	平均正解率
SVM	0.757
提案手法	0.736
提案手法 +SVM	0.752

実験の結果，SVM の正解率には劣るものの，提案手法に比べると WSD の精度を改善することができた．

6.2　提案手法と教師なしモデルの比較

WSD による実験では SVM に比べ提案手法，MSSG モデル，NP-MSSG モデルのいずれの手法も低い結果となった．MSSG モデルと NP-MSSG モデルの正解率が低い原因としては，語義の分散表現の学習において単語の分散表現を学習する時とは別のコーパスを用いており，SemEval-2 の教師データを用いていないためだと思われる．一方教師データを用いている提案手法の正解率が低い原因は前述したが，他の原因として単語の分散表現を語義の分散表現に分解する際に，特徴的な一つの語義に対して重みを与えているため，各語義の間で特徴の少ない語義について正しい学習ができていないのではないかと考えられる．

例えば「 経済」「 関係」「 一」「 文化」「 他」のような特定の語義の出現頻度が際立って多い単語については SVM と同程度の高い正解率となっていることが確認できる．

また NP-MSSG モデルの特徴として，教師なし学習であるため，教師データ中に出現しない語義の分散表現も構築するというものがある．例えば「 前」という単語の覚悟着の頻度は表 6.2 のようになっている．教師データ中に出現しない語義の用例が 7 個あるが，NP-MSSG モデルでは単語「 前」に対して 3 つのベクトルを作っており，48488-X-X の語義の用例 7 個のうち 3 つの用例で正しい識別を行っていた．

このことから，NP-MSSG モデルは WSD の精度は低いものの，教師データ中に出現しない未知の語義に対する分散表現を構築できる可能性があると考えられる．

表 6.2　単語「 前」の語義の頻度

	教師データ	テストデータ
48488-0-0-1	19	12
48488-0-0-2	31	31
48488-X-X-X	0	7

第 7 章

結論

本論文では教師データを用いて単語の分散表現から語義の分散表現を構築する手法を提案した．具体的には教師データから得た語義の文脈ベクトルと語義の出現頻度の情報を用いて単語の分散表現を語義の分散表現に分解するというものである．従来の教師なし学習による言語モデルを用いた語義の分散表現の構築では，コーパス中に出現する語義の出現頻度の情報が得られないという問題がある．そこで我々は教師データを用いることでこの問題を解決できると考えた．　実験では単語の分散表現として Wkipedia の日本語記事を学習させたものと国立国語研究室が作成した分散表現 nwjc2vec を用い，教師データとして SemEval-2 の日本語辞書タスクのデータを用いた．提案手法により作成した語義の分散表現を用いて WSD の実験を行ったところ，文脈ベクトルを素性とした SVM による分類器と比べて精度の改善は見られなかった．しかし語義の分散表現と類似する分散表現を持つ単語を分析した結果，語義の分散表現が正しく作られていることが分かった．　さらに教師なし学習による語義の分散表現の構築手法として MSSG モデルと NP-MSSG モデルを実装し，構築された語義の分散表現を用いて WSD の実験を行った．実験の結果，提案手法と NP-MSSG モデルに比べて MSSG モデルが高い正解率を出すことが確認できた．対象単語別に正解率を見ると，語義の出現頻度の差が大きい単語に対しては提案手法が高い正解率となり，語義の出現頻度の差が小さい単語に対しては MSSG モデルが高い正解率となることが分かった．また NP-MSSG モデルは他の手法に比べて正解率は低かったものの，教師データ中に出現しない語義の分散表現を構築できる可能性があることが分かった．

謝辞

本研究を進めるにあたり，多くのご指導，ご協力を頂いた指導教員の新納浩幸教授に感謝致します．また，日常の議論を通じて多くの知識や示唆を頂いた佐々木稔講師，古宮嘉那子講師と新納研究室の皆様に感謝します．

参考文献

[1] Eric H Huang, Richard Socher, Christopher D Manning, and Andrew Y Ng. Improving word representations via global context and multiple word prototypes. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1*, pp. 873–882. Association for Computational Linguistics, 2012.

[2] Jiwei Li and Dan Jurafsky. Do multi-sense embeddings improve natural language understanding? *arXiv preprint arXiv:1506.01070*, 2015.

[3] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pp. 3111–3119, 2013.

[4] Arvind Neelakantan, Jeevan Shankar, Alexandre Passos, and Andrew McCallum. Efficient non-parametric estimation of multiple embeddings per word in vector space. *arXiv preprint arXiv:1504.06654*, 2015.

[5] 山木翔馬, 新納浩幸. 教師あり・教師なし学習により構築した語義の分散表現を用いた語義曖昧性解消に関する一考察. 言語処理学会第 24 回年次大会, to appear, 2018.

[6] 山木翔馬, 新納浩幸, 古宮嘉那子, 佐々木稔. 教師データを用いた語義の分散表現の構築. 言語処理学会第 23 回年次大会発表論文集, 2017.

[7] 新納浩幸. Chainer による実践深層学習. オーム社, 2016.

[8] 浅原正幸, 岡照晃. nwjc2vec:『国語研日本語ウェブコーパス』に基づく単語の分散表現データ. 言語処理学会第 23 回年次大会, to flapper, 2017.