

# 用例文間の意味的な類似関係を用いた 半教師あり語義曖昧性解消

谷田部 梨恵<sup>1</sup> 佐々木 稔<sup>2,a)</sup>

受付日 2021年1月27日, 採録日 2021年7月7日

**概要:** 単語に対して前後の文などを手がかりとして適切な語義を判定する語義曖昧性解消は今日に至るまで様々なアプローチで研究が行われている。先行研究では、教師あり学習を用いた分類手法に対して誤り分析が行われ、主要な要因として語義ラベルを付与した用例文集合である訓練データの不足をあげている。そのため、語義曖昧性解消の精度を向上させるためにはさらに多くの用例文の追加やその他の有効な情報の追加が求められている。しかし、学習用のデータを新たに追加することは、語義識別に精通した専門家による正解ラベル付与が必要となるためコストがかかるという問題がある。そこで、本研究では用例文間の同義関係を用いた半教師あり語義曖昧性解消手法を提案する。この手法では、訓練データと語義ラベルなしの用例文集合から同義関係を表したグラフを作成し、このグラフと訓練データに対してグラフニューラルネットワークで学習して語義識別モデルを構築する。Semeval-2010 日本語タスクのデータセットと SENSEVAL2 English Lexical Sample タスクデータを用いて評価実験を行った結果、提案手法はグラフ構造を使わない手法や既存手法と比較して精度が向上し、グラフ構造の利用が有効であることを示した。また、提案手法は学習済み言語モデルの BERT を用いて同義かどうかを判定するパラメータ設定のないグラフ構築手法とパラメータ設定が必要な Cosine 類似度や相互  $k$  近傍グラフによるグラフ構築手法のどちらを用いても高い精度で語義を識別できることを示した。

**キーワード:** 語義曖昧性解消, グラフ構造, 半教師あり学習, グラフニューラルネットワーク

## Semi-supervised Word Sense Disambiguation Using Semantic Similarities between Examples

RIE YATABE<sup>1</sup> MINORU SASAKI<sup>2,a)</sup>

Received: January 27, 2021, Accepted: July 7, 2021

**Abstract:** Word Sense Disambiguation (WSD) is a well-known problem in the natural language processing. In recent years, there has been increasing interest in applying neural networks and machine learning techniques to solve WSD problems. However, these previous approaches often suffer from the lack of manually sense-tagged examples of specific words due to costly annotation work. Therefore, we consider that more sense-tagged examples and some other useful information need to be added to improve the performance of the WSD model. In this study, we propose a semi-supervised WSD method using semantic similarities between example sentences. In this method, to represent a semantic similarity relation obtained from sense labeled examples and unlabeled examples, we also propose a graph construction method that does not require any parameters using BERT pre-trained model, as well as graph construction methods using cosine similarity and mutual  $k$ -neighborhood graphs. We construct a WSD classifier for the target word by using a graph convolutional network for training data and graph structure. The effectiveness of the proposed method was evaluated on the Semeval-2010 Japanese WSD task dataset and on the SENSEVAL-2 English lexical sample task. Experimental results show that the proposed method improves the performance of WSD compared to the method without graph structure.

**Keywords:** word sense disambiguation, graph structure, semi-supervised learning, graph convolutional neural network

## 1. はじめに

日常で使用する単語には様々な意味を持つものがある。たとえば「意味」という単語では、「その言葉の表す内容、意義」と「表現や行為の意図・動機」などという語義を複数持つように、文章中で使われる単語には様々な語義が存在する [1]。このように様々な語義を持つ単語に対して前後の文などを手がかりとして適切な語義を判定するタスクとして語義曖昧性解消というものがある。

計算機による単語の語義曖昧性解消は、今日に至るまで様々なアプローチで研究が行われている [2]。語義曖昧性解消システムの誤り原因の分類が行われ、語義ラベルを付与した用例文集合である訓練データが不足して誤る事例の多いことが主要な要因として指摘されている [3]。最近では、対象単語の前後に出現する単語を分散表現と呼ばれるベクトルで表現し、そのベクトルを入力してニューラルネットワークを用いて分類する手法が主流である [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15]。しかし、新たに学習データを追加するには、用例文における単語の正解語義の割当てに精通した専門家によるラベル付与が必要となるためコストがかかるという問題がある。そのため、訓練データに用例文を追加する以外にも、語義曖昧性解消に有効な何らかの情報を追加することが求められている。

そこで、本研究では用例文間の類似関係を用いた半教師あり語義曖昧性解消手法を提案する。この手法では、訓練データと語義ラベルなしの用例文（ラベルなし用例文）集合から代表的な言語モデルである BERT を用いてノードを用いて、エッジを用例文間の同義関係を表すグラフを作成し、このグラフと訓練データに対してグラフニューラルネットワークで学習して語義識別モデルを構築する。得られたモデルに、語義を調べたい単語を含む用例文を入力することで対象単語の語義を推定する。既存の半教師あり手法では、ラベルなし用例文を少数の訓練データで学習した語義識別モデルによって自動でラベル推定し、その結果を訓練データに追加することが一般的である [6], [15], [16], [17]。これらの手法は用例文の特徴と語義ラベルの関係を用いて識別モデルの学習を行うが、用例文間の同義関係を考慮していないという課題が存在する。用例文間の同義関係をとらえてグラフを構築する簡単な方法としては、Cosine 類似度や相互  $k$  近傍グラフを用いた判定方法がある。この手法では、距離や類似度だけで同義判定を行うため、訓練データにある語義ラベルの情報を利用できないという課題や判

定するためのパラメータ設定が必要であるという課題が存在する。これらの課題に対して、提案手法では語義曖昧性解消において訓練データとグラフから得られる分散表現とともに用いることで、従来手法と比較して効果的な特徴をとらえることができると考えられる。さらに、提案手法では BERT の事前学習済みモデルを利用してグラフを構築する方法も提案し、パラメータ設定をすることなく用例文の語順を考慮した同義文判定を行う。

提案手法の語義識別性能を評価するために、日本語の評価データである SemEval-2010 日本語 WSD タスクのデータセット [18] と英語の評価データである SENSEVAL2 English Lexical Sample タスクデータ [19] を用いて評価実験を行う。ただし、本研究では辞書未定義語義の識別に関しては研究対象とせず、文脈情報を幅広くとらえて効果的に語義曖昧性解消を行うことを研究目的とする。対象単語に対する語義の識別精度を評価することで、提案手法が語義曖昧性解消に有効であることを示す。

本研究の貢献は以下の 2 点にまとめられる。

- 半教師あり学習に基づく語義曖昧性解消において、同義関係を表したグラフ構造から得られる分散表現が語義曖昧性解消に有効であることを示した。日英の評価データによる比較を行った結果、各データにおける最も精度の高い既存の半教師あり手法と比較して精度が向上し、提案手法が有効であることを示した。
- 同義関係を表すグラフにおいて、言語モデルの BERT を用いて同義かどうかを判定するパラメータ設定のないグラフ構築手法を示した。また、このグラフ構築手法とパラメータ設定が必要な Cosine 類似度や相互  $k$  近傍グラフによるグラフ構築手法はどちらを用いても高い精度で語義を識別できることを示した。

本論文は次のように構成されている。2 章では、機械学習に基づく語義曖昧性解消における既存研究と本研究の位置付けについて述べる。3 章では、本研究において使用するグラフニューラルネットワークを用いた半教師あり学習手法である Planetoid について概説する。4 章では、提案手法である Planetoid を用いた半教師あり語義曖昧性解消システムについて詳しく述べる。5 章では、評価実験と比較手法について説明し、6 章で実験結果を示す。7 章では実験結果に対して考察を行い、8 章で結論を述べる。

## 2. 関連研究

語義曖昧性解消を自動的に行う場合、対象となる多義語に対して複数ある語義候補から適切な語義を選択する分類問題として定式化される。分類問題として語義曖昧性解消を行う際、対象となる多義語の前後に出現する単語などの様々な手がかりを用いて語義の分類が行われる [2]。この共起単語のほかにも、対象単語や共起単語の品詞情報、読み仮名、係り受け情報、シソーラスから得られた概念情報

<sup>1</sup> 株式会社日立産業制御ソリューションズ  
Hitachi Industry & Control Solutions, Ltd., Hitachi, Ibaraki  
316-1221, Japan

<sup>2</sup> 茨城大学工学部情報工学科  
Ibaraki University, Hitachi, Ibaraki 316-8511, Japan

a) minoru.sasaki.01@vc.ibaraki.ac.jp

といった語彙素性と呼ばれる特徴が語義曖昧性解消に有効であることが示されている [18], [20]. 語義曖昧性解消のアプローチは大きく分けて, 知識に基づく手法, 教師あり学習, 教師なし学習 [21], および, 半教師あり学習を用いた手法が存在する [2]. 本研究は半教師あり学習に基づくアプローチを採用している.

教師あり学習は各単語にラベルとなる語義が付与された用例文集合から構築された分類器を用いて, 入力された単語の用例文に対して語義を予測するアプローチである. 人間が適切に語義ラベルを付与した用例文を入力するため分類精度が高いという利点がある. 近年ではニューラルネットワークに基づく手法が主流で, 共起単語の分散表現を素性として利用して教師あり学習を用いて語義曖昧性解消を行う手法 [4], [15] や Long Short-Term Memory (LSTM) を用いて対象単語周辺の単語の語順を考慮して語義を推定する手法がある [6]. また, All-words 型語義曖昧性解消に対して単語系列を語義ラベルの系列に変換する系列ラベリング問題と見なして, 双方向 LSTM や Encoder-Decoder モデルなどの様々なモデルを使って有効性を評価した研究も存在する [7], [8]. しかし, 高い分類精度を得るためには, 語義ラベルを付与した大量の訓練データが必要となり, 訓練データを人手で作るコストが大きいという欠点がある.

この課題を解決するアプローチとして, 半教師あり学習を用いた手法が存在する. 半教師あり手法は一般的に, 少数の訓練データで語義識別モデルを学習し, ラベルなし用例文をこのモデルで語義を推定して, その結果を訓練データに追加する. 訓練データから作成した分類器の予測結果に基づく手法として, Co-training や Self-training を用いた手法がある [22]. Co-training や Self-training による分類手法は訓練データから得られる分類器を使用し, ラベルなしデータに確信度付きのラベルを付与して, それを利用することで分類器を改善し, そのうえで学習と識別を行う. 同じような訓練データの拡張手法として, 国語辞典の例文から得られる素性に一致する文は同じ語義で使用されると仮定して語義ラベル付きデータを抽出する藤田らの手法も存在する [16]. データをある空間へマッピングする分類手法として, データ間の距離学習手法を応用した手法 [23] や生成モデル [24], Stacked Denoising Autoencoder を使用した手法 [25] が存在する. これらの手法は, まずラベルなしデータを分離し, 空間にマップする. 次にラベル付きデータもその空間にマップし, その空間上で分類器の学習と識別を行う. これらの手法は特徴として語彙素性を用いており, 近年有効性が示されている単語の分散表現を特徴として用いていない.

グラフに基づいて語義ラベルを予測するアプローチでは, 外部知識を用いて語義をグラフのノード, 類義関係や上位下位関係をエッジで表す手法が存在する [26]. 本研究では, 外部知識は使わず, 用例文をノード, 用例文間の同義関係

をエッジでグラフを構築する. そのため, 前述のラベル付きデータから作成した分類器の予測結果に基づく半教師あり学習手法の中に分類できる. 代表的な手法として, ラベル伝搬法 (Label Propagation) を用いて訓練データを拡張する手法がある [6], [27]. これは用例文から抽出した素性データのグラフを作成し, ラベルの自動推定を行う手法である. 類似度の最も高い文どうしは同一語義を持つと仮定し, ラベルを伝搬させることでラベルなしデータにラベルを付与する. また, 最近の研究では Gaussian Random Field (GRF) を用いてラベルなしデータのラベルを推定する手法も存在する [15]. 本研究もこれらの手法と同じくグラフに基づいてラベル予測を行っているが, 提案手法は訓練データとグラフの分散表現をともに特徴として用いることが従来手法と異なる点である.

半教師あり学習手法とは異なるが, 用例文における単語の分散表現に辞書やシソーラスなどの外部知識を組み合わせた語義曖昧性解消についての研究も数多く行われている. このアプローチによる研究として, 語義定義文と教師あり学習を組み合わせた GlossBERT [9] や BEM [13], 上位下位関係などの単語間関係を組み合わせた EWISE [10] や EWISER [14], シソーラスの類義語関係や上位下位関係を用いて語義や概念の分散表現を構築して語義曖昧性解消を行う手法が存在する [11], [12]. これらの手法では単語間関係をグラフ構造として表現していることから, 単語をノード, 単語間関係をエッジで表現するグラフとなっている. 本研究は用例文をノード, 用例文間の意味的な類似関係をエッジとしたグラフ構造を用いており, これらの手法とは異なる構造のグラフである.

### 3. Planetoid グラフニューラルネットワークを用いた半教師あり学習

本章では, 提案手法において利用する半教師あり学習モデルである Planetoid の概要を説明する.

#### 3.1 Planetoid の概要

Planetoid は半教師あり深層学習手法として 2016 年に提案されたものである [28]. Planetoid には Transductive モデルと Inductive モデルという 2 つのモデルがあり, 本研究ではテストデータのように訓練データに存在しないデータについてラベル予測を行う Inductive モデルを使用する.

Inductive モデルのネットワーク構造を図 1 に示す. 入力データとして, 訓練データとラベルなしデータ, およびデータ間の接続情報を表現したグラフを用いる. 訓練データから得られる特徴ベクトルをフィードフォワードニューラルネットワーク (FF-NN) に入力し (図 1 の FF-NN(1)), 正解ラベルが出力されるようにモデルを学習する. 入力データから得られるグラフを別の FF-NN に入力し (図 1 の FF-NN(3)), データ間の接続情報に基づいて入力データ

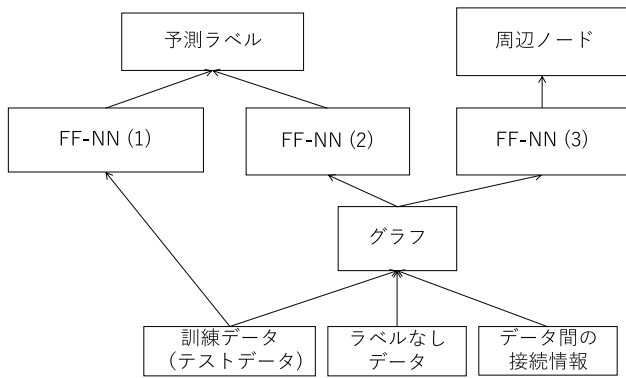


図 1 Planetoid-Inductive モデルのネットワーク構造

Fig. 1 WSD Model Using Graph Embeddings Based Learning.

のノードと隣接する周辺ノードを出力するようにモデルを学習する。この FF-NN(3) のモデルに訓練データを入力し、グラフの周辺に存在するラベルなしデータのサンプリングを行って正例と負例を抽出し、訓練データのラベルを正例に割り当てる。この結果を用いて、すべてのデータに対する FF-NN(3) の最終層のベクトルを FF-NN(2) に入力し、割り当てられたラベルを出力するようにモデルを学習する。最後に、訓練データを図 1 の FF-NN(1) と FF-NN(2) に入力して、得られる隠れ層の重みベクトルを連結したベクトルに対して正解ラベルが出力されるように分類モデルを学習する。学習を行う際、訓練データを学習したときの損失とラベルなしデータでグラフ構造を予測したときの損失の 2 つの合計を損失関数として最小化させる。この学習を何度も繰り返し実行することによって、損失が小さくなるように隠れ層の重みを更新する。

訓練データによる学習を行った分類モデルに対して、未知のデータに対するラベル推定を行う。未知のデータを分類モデルに入力すると、分類モデルはそのデータに対応して推定されたラベルを出力する。

### 3.2 入力 (データとグラフ)

Inductive モデルへの入力は以下のとおりである。

- 訓練データ：ラベルありデータの特徴ベクトル  $x_1, \dots, x_l$
- 訓練データのラベル：語義ラベルの数  $s$  を次元数とする One-Hot ベクトル  $y_1, \dots, y_l$  (正解語義ラベルに対応する要素を 1, 残りの要素を 0 とする  $s$  次元ベクトル)
- 訓練データとラベルなしデータ：ラベルありとラベルなしの両方の特徴ベクトル  $allx = \{x_1, \dots, x_l, x_{l+1}, \dots, x_N\}$  ( $N$  は  $allx$  に含まれるデータ数)
- グラフ  $G$  の隣接行列  $A$  ( $A \in R^{N \times N}$ )：ノード  $x_i$  と  $x_j$  が接続している場合、エッジには  $a_{ij}$  の重みを持つ。接続していない場合はエッジの重みは 0 である。

### 3.3 グラフニューラルネットワークの学習

Inductive モデルは訓練データを学習したときの損失  $L_s$  とグラフ構造を予測したときの損失  $L_u$  の 2 つの合計を損失関数として最小化させるように学習を行う。

この Inductive モデルにおいて、訓練データやグラフのサンプリングにより得られたラベルなしデータを変換して特徴ベクトル  $x$  を作成し、この特徴ベクトル  $x$  をフィードフォワードニューラルネットワーク (FF-NN(2)) に入力する。このネットワークにおける第  $k$  番目の隠れ層  $h^k(x)$  は以下の式で表現される。

$$h^k(x) = \text{ReLU}(W^k h^{k-1}(x) + b^k)$$

ここで  $W^k$  と  $b^k$  は第  $k$  番目の隠れ層のパラメータであり、 $\text{ReLU}(x)$  は隠れ層 (50 次元) の活性化関数である Rectified Linear Unit (ランプ関数) を表している。訓練データの特徴ベクトルを入力したときの中間層のベクトル  $h^k(x)$  とグラフから得られた特徴ベクトルを入力したときの中間層のベクトル  $h^l(x)$  を連結し、 $h(x) = [h^k(x), h^l(x)]$  を求める ( $[h^k(x), h^l(x)]$  は 2 つのベクトル  $h^k(x)$  と  $h^l(x)$  を水平方向に連結したベクトルを表す)。連結したベクトル  $h(x)$  を重みのパラメータ  $w_y$  を持つ softmax 層に入力して、ラベル数の要素を持つベクトルに変換して、各ラベルを選択する確率値  $p(y|x)$  を計算する。確率  $p(y|x)$  は以下の式で表される。

$$p(y|x) = \frac{\exp(h(x)w_y)}{\sum_{y'} \exp(h(x)w_{y'})}$$

Inductive モデルは最終的にこの確率値が最大になるラベルを予測結果として出力する。

このモデルに訓練データを入力した場合のラベル予測における損失  $L_s$  を求める。訓練データを学習するモデルとして、3 層の多層パーセプトロンを用いて隠れ層の学習を行う。  $l$  個の訓練データに対する交差エントロピーによる損失  $L_s$  は以下の式で表される。

$$L_s = -\frac{1}{l} \sum_{i=1}^l \log p(y_i|x_i)$$

次に、グラフからサンプリングを行ったときの損失  $L_u$  を求める。Planetoid ではグラフ構造に基づくサンプリングとラベル情報に基づくサンプリングの 2 種類が存在し、これらのサンプリングは 2 つのパラメータによって切り替えている。パラメータ  $r_1$  ( $0 < r_1 < 1$ ) は正例か負例かのラベルを決めるための閾値、パラメータ  $r_2$  ( $0 < r_2 < 1$ ) はグラフに基づくサンプリングとランダムウォークに基づくサンプリングを切り替えるための閾値である。

1 回のサンプリングで、2 つの乱数  $rand_1$  ( $0 < rand_1 < 1$ ) と  $rand_2$  ( $0 < rand_2 < 1$ ) を取得し、サンプリング方法を切り替える。  $rand_1 < r_1$  のとき、正例のラベルを表す

$\gamma = +1$ , それ以外のときは負例のラベルを表す  $\gamma = -1$  をあらかじめ設定する.  $rand_2 < r_2$  の場合, ランダムウォークに基づくサンプリングを行う. 訓練データとラベルなしデータに対して, 初期ノード  $S_1 = x_i$  から確率  $a_{ij} / \sum_{j'=1}^N a_{ij'}$  に従ってノード  $x_j$  に遷移するランダムウォークを行い, 長さ  $q$  のノード列  $S$  を求める.  $\gamma = +1$  の場合, ノード列  $S$  からエッジの数を表すノード間距離のパラメータ  $d$  よりも少ない距離のノードペアの集合を求め, その中からノード  $i$  と周辺ノード  $c$  のペア  $(i, c)$  を正例としてサンプリングする.  $\gamma = -1$  の場合, 訓練データとラベルなしデータからランダムに選んだノード  $c$  のペア  $(i, c)$  を負例としてサンプリングする.  $rand_2 \geq r_2$  の場合はラベルに基づくサンプリングを行う. 訓練データの中から対応するノード  $i$  と  $c$  をサンプリングし,  $\gamma = +1$  であればペア  $(i, c)$  を正例,  $\gamma = -1$  であればペア  $(i, c)$  を負例とする.

すべてのデータに対する交差エントロピーによる損失の期待値  $L_u$  は以下の式で表される.

$$\begin{aligned} L_u &= -E_{(i,c,\gamma)} [\log(\gamma w_c^T h^l(x))] \\ &= -E_{(i,c,\gamma)} [I(\gamma = 1) \log \sigma(w_c^T h^l(x)) \\ &\quad + I(\gamma = -1) \log \sigma(-w_c^T h^l(x))] \end{aligned}$$

ここで,  $w_c$  と  $h^l(x)$  はグラフから得られた特徴ベクトルを入力するニューラルネットワークの重みと中間層のベクトルを表す. また,  $I(\cdot)$  は引数が真であれば1, 偽であれば0を返す指標関数,  $\sigma(x)$  はシグモイド関数  $\sigma(x) = 1/(1+e^{-x})$  として定義される,

これら2つの損失  $L_s$  と  $L_u$  を足し合わせた全体的な損失  $L$  は以下の式で表される.

$$L = L_s + L_u$$

この損失  $L$  を最小化するように学習を行う.

損失を最小化するための学習手法には確率的勾配降下法 (SGD) を用いたミニバッチ学習を行う [29]. この学習法は学習用データ (訓練データとグラフ) の中から, いくつかデータを取出して損失関数を計算し, 最適化することでモデルパラメータ  $w$  を更新する. 損失関数  $L(w)$  に対して学習率であるパラメータ  $\varepsilon$  を用いて, 以下の式に示す勾配ステップをとることで最適なモデルパラメータ  $w$  を求める.

$$w = w - \varepsilon \frac{\partial L(w)}{\partial w}$$

$L_s$  の最適化を求めるときの学習率は  $\varepsilon_s (= 0.1)$  とし,  $L_u$  の最適化を求めるときの学習率は  $\varepsilon_u (= 0.001)$  とする.

損失関数を最適化したことで得られた識別モデルは3.4節に示す未知データの識別で使用される.

### 3.4 学習されたモデルを用いた未知データの識別

未知データをベクトル化し, このベクトルを学習したモデルに入力すると, あらかじめ設定されたラベルごとに予測確率を計算する. この予測確率に対して最も確率の大きいラベルを判定結果とし, 識別結果として出力する.

## 4. 半教師あり学習手法を用いた語義曖昧性解消手法

本章では, 提案手法である半教師あり学習手法を用いた語義曖昧性解消手法のシステムについて述べる.

### 4.1 語義曖昧性解消システムの概要

本システムは語義識別モデルの学習と語義を知りたい用例文の語義推定を行う. 語義識別モデルは訓練データと用例文間の関係を示すグラフを入力して学習を行うことで得られる. 入力となるグラフを作成するために, 訓練データの用例文ペアを用いて同義判定モデルの学習を行い, このモデルを用いて未知の用例文ペアの同義判定を行う. 訓練データとラベルなしデータ, および用例文間の接続情報を3章の Planetoid に入力して識別モデルを学習する. そこで得られた語義識別モデルに用例文を入力することで対象単語の語義を推定することが可能となる.

### 4.2 素性 (双方向言語モデルを用いた文脈ベクトル)

データ入力部分では入力文の文脈を考慮して, 事前学習された双方向言語モデル (biLMs) から得られた文脈に応じた単語表現を特徴量として適用する. これは事前学習された ELMo モデルを用いて, 対象単語を含む入力文の分散表現を得る. この事前学習された ELMo モデルは, テキストから文脈情報を学習するための2層双方向 LSTM 構造である. 本システムでは事前学習モデルの最後の層の出力のみを使用する.

訓練データとラベルなし用例文, テストデータの用例文から得られる単語列をそれぞれ ELMo モデルに入力し, 対象単語および前後2単語の5つの単語分散表現を得る. そして, これら5つの単語分散表現を連結した5,120次元ベクトルを生成する. 単語分散表現は, 入力文の各単語に対して文脈に応じて生成されるため, 単語の意味ごとに異なる分散表現が生成される傾向がある. この5つの単語分散表現により, 対象となる単語の文脈を限定することができる. この分散表現を利用し, 「力士の住む寮の場所を教えて」という文の対象単語を「場所」としてその後2単語を分散表現として表す流れを図2に示す.

### 4.3 Planetoid に入力するグラフ構造

Planetoid で用いるグラフ構造は用例文に対応するノードと用例文間の関連を示すエッジからなる. グラフ中のノードは訓練データとラベルなしデータに含まれる用例文

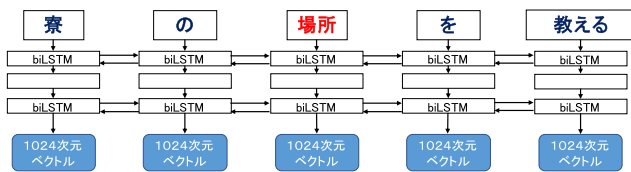


図 2 単語分散表現の流れ

Fig. 2 Context Representation Derived from Pre-trained Bi-directional Language Model (ELMo).

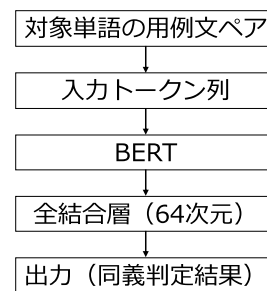


図 4 BERT による同義判定モデル

Fig. 4 Sentence Pair Classification by Fine-tuning BERT.

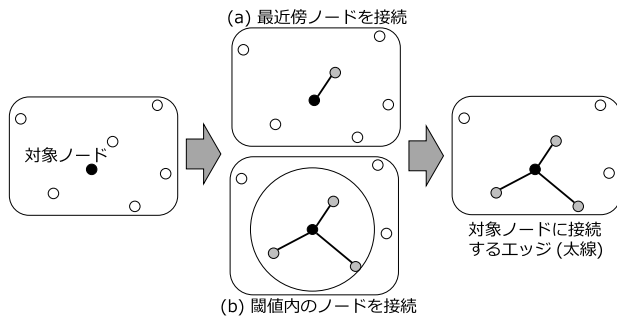


図 3 Cosine 類似度を用いたグラフ構造の作成方法

Fig. 3 Edge Connection Between Examples Using Cosine Similarity.

を表す。エッジはノード間で関連があると判断された場合に結ばれることとする。本研究では、判断基準となるパラメータ設定が必要な Cosine 類似度を用いた方法と相互  $k$  近傍グラフを用いた方法、およびパラメータ設定が不要な BERT による同義判定を用いた方法の 3 種類を用いて効果を比較する。

#### 4.3.1 Cosine 類似度を用いたグラフ構造

グラフ構造は訓練データとラベルなしデータに含まれる各用例文をノードとする。ノード間のエッジの有無は Cosine 類似度を用いて判定する。Cosine 類似度は 2 つの素性ベクトルがなす角の余弦で、2 つの用例文が類似すれば 1 に近い値を持つ。

図 3 に Cosine 類似度を用いてエッジをつなげる様子を示す。各ノードにおいて、他のノードのとの類似度を計算し、図 3 の (a) に示す最も類似度の高いノードと図 3 の (b) に示すしきい値以上の類似度を持つノードをすべて接続する。これにより得られたグラフを図 1 に示す Planetoid のデータ間の関係として入力する。

Cosine 類似度のしきい値は 0.9 と設定する。ここで、同じ語義を持つ用例文は周辺に類似した単語や品詞が出やすく、異なる語義を持つ用例文は周辺に異なった単語や品詞などが出現しやすいと仮定している。最高類似度だけでなく、確信度の高い用例文も組み込むことが重要であると考えたため、このような設定とした。

#### 4.3.2 相互 $k$ 近傍グラフを用いたグラフ構造

データ集合からグラフ構造を求めるための半教師あり手法の 1 つとして、相互  $k$  近傍グラフが存在する。相互  $k$  近傍グラフは任意の 2 点間のノードが  $k$  近傍でお互いに含

まれている場合にのみエッジを結ぶグラフである。この相互  $k$  近傍グラフを用いて、用例文集合のグラフ構造を構築する。訓練データとラベルなしデータに対して、4.2 節で得られた素性を用いて KD-tree を作成し、各ノード  $n_i$  ( $1 \leq i \leq N$ ) についてユークリッド距離の最も近い  $k$  個のノード  $kNN(n_i)$  を求める。グラフにおける任意のノードペア  $n_i$  と  $n_j$  に対して、 $n_i \in kNN(n_j)$  かつ  $n_j \in kNN(n_i)$  であればエッジを結ぶ。各用例文に対して  $k = 3$  として  $k$  近傍を求め、エッジを結んだ用例文ペアを同義であると判定する。このとき、ノードの次数はすべて  $k$  以下となり、極端に高い頂点次数はできない。そのため、最近傍の用例文も同義としてエッジを結ぶ。このようにして得られたグラフ構造を図 1 におけるデータ間の接続情報として Planetoid に入力する。

#### 4.3.3 BERT を用いたグラフ構造

BERT を用いたグラフ構造は図 4 に示す BERT による同義判定モデルを用いて作成する。BERT は汎用的な言語モデルの事前学習を行う手法の 1 つである [30]。大量のテキストデータを事前学習したモデルに少量の訓練データを追加学習させることで、テキスト分類などの様々なタスクで先行研究を超える分類性能を達成している。本研究では、BERT の学習済みモデルを用いて Sequence Classification タスクとして用例文ペアが同義か非同義かの判定を行う。用例文ペアをそれぞれ単語列に分割し、指定した数の単語列を用いて BERT へ入力するトークン列を作成して BERT の学習済みモデルに入力する。このモデルから出力された分散表現を 3 層の全結合層に入力し、用例文間の同義判定タスクに転移学習することで、対象単語が同じ語義で使われているかどうか判定するモデルを構築する。

各対象単語について、訓練データにあるすべての用例文ペアを用いて同義判定モデルの学習を行う。2 つの用例文  $s_1$  と  $s_2$  をそれぞれ単語列に分割し、指定した数の単語列を用いて BERT へ入力するトークン列を作成する。本研究では、使用する単語列として、対象単語とその前後 3 単語 (後述する表 2, 表 3 では前後 3 単語と表記) と用例文のすべての単語列 (表 2, 表 3 では文単位と表記) の 2 種類を用いて効果を比較する。対象単語と前後 3 単語を用いる場合、各

用例文の単語列  $(w_{s1}^1, w_{s1}^2, \dots, w_{s1}^7)$  と  $(w_{s2}^1, w_{s2}^2, \dots, w_{s2}^7)$  に対して、文頭を表す '[CLS]' と文の区切りを表す '[SEP]' という特別なトークンを用いて、

$$\{ '[CLS]', w_{s1}^1, w_{s1}^2, \dots, w_{s1}^7, '[SEP]', w_{s2}^1, w_{s2}^2, \dots, w_{s2}^7, '[SEP]' \}$$

というトークン列を作成して BERT の学習済みモデルに入力する。この入力で得られる BERT の隠れ層のうち、'[CLS]' トークンに対応する最終の隠れ層を用例文ペアの分散表現とする。用例文のすべての単語列を用いる場合は各用例文のすべての単語列を用いてトークン列を作成し、このトークン列を BERT の学習済みモデルに入力する。

得られた用例文ペアの分散表現は同義ラベルとあわせて 4 層の多層パーセプトロン (MLP) で学習し、同義判定モデルを構築する。このとき、2 つの隠れ層の次元はともに 64、バッチサイズは 32、MLP の学習回数は 20 回とする。

グラフ構造を作成するには、既知の訓練データ間同義関係に加えて、訓練データとラベルなしデータの用例文ペア、および、ラベルなしデータ間の用例文ペアについて同義判定を行う必要がある。未知の用例文ペアに対して同義判定を行う場合も学習時と同様に指定した数の単語列から得られるトークン列を判定モデルに入力する。この入力に対するモデルからの出力は同義と非同義の確率となり、同義の確率が非同義よりも大きいときに同義であると判定し、同義である用例文のノード間にエッジを結ぶことでグラフを作成する。

#### 4.4 Planetoid を用いた語義識別モデルの学習

訓練データ、ラベルなしデータに対して、4.3 節に示したグラフ作成方法を用いて得られたグラフと訓練データを 3 章に示した Planetoid に入力して、語義識別モデルの学習を行う。Planetoid における訓練データの特徴ベクトルは 4.2 節で得られた素性ベクトルを用いる。

#### 4.5 学習されたモデルを用いた語義の識別

4.4 節で得られた語義識別モデルに対して、語義を推定したいテストデータを入力することで、語義識別結果が出力される。テストデータの特徴ベクトルも訓練データと同様に 4.2 節で得られた素性ベクトルを用いる。

### 5. 実験

本章では、グラフニューラルネットワークを用いた半教師あり学習による語義曖昧性解消を用いて、精度比較実験を行う。実験は同じ実験を 3 回行い、その平均精度を実験結果として使用する。

#### 5.1 日本語データによる日本語 WSD 実験

##### 5.1.1 日本語語義曖昧性解消評価データセット

本研究における日本語語義曖昧性解消の評価データとし

て、SemEval-2010 日本語 WSD タスクデータ [18] を使用する。SemEval-2010 日本語 WSD タスクはコーパス中に出現する対象単語に対して、岩波国語辞典第 5 版で定義された語義の中から適切な語義を推定することを課題とし、タスクが対象とするコーパスが多岐の分野にわたって構成されている点と使用するデータ中に辞書に定義されていない語義が出現するという特徴がある。

このタスクの対象単語は 50 単語あり、名詞単語が 22 単語、動詞単語が 23 単語、形容詞単語が 5 単語となっている。各単語には訓練データとテストデータとして、その単語を使用した用例文がそれぞれ 50 個用意されている。訓練データは国立国語研究所が開発した現代日本語書き言葉均衡コーパス (BCCWJ) に収録された白書、新聞、書籍の分野から選ばれ、テストデータはこれらの分野に加えて Yahoo!知恵袋も対象として用例文が選ばれている。

配布されたデータには形態素情報と分野コード、および語義 ID が付与されている。形態素情報は単語境界、品詞、基本形、および読みから成り立っている。選ばれたデータに対して ChaSen と UniDic を用いて形態素解析を行い、形態素解析の誤りを人手で修正した形態素情報を各単語に割り当てている。分野コードは文書が白書、新聞、書籍、知恵袋のどの分野から抽出したのかを示し、用例文が存在する分野が分かるようになっている。語義 ID は岩波国語辞典第 5 版の語義に従って定義された語義のコードで、用例文中の対象単語に対して手作業で語義 ID が付与されている。ただし、対象単語が辞書に定義されていない語義で使われている場合は未定義語義を表すコードが語義 ID として割り当てられている。

##### 5.1.2 日本語ラベルなし用例文データ

実験用のラベルなし用例文データには、国立国語研究所が開発した現代日本語書き言葉均衡コーパス (BCCWJ) から抽出した用例文集合を利用する。BCCWJ は日本語の様々なジャンルの文書を収録した、書き言葉の全体像を把握するために構築されたコーパスである。

##### 5.1.3 日本語語義曖昧性解消実験の設定

日本語語義曖昧性解消実験では 5.1.1 項で説明した SemEval-2010 日本語 WSD タスクデータを使用する。これまでの研究において、このデータには付与された語義ラベルに誤りが存在することが指摘されている [3]。しかし、誤りが修正されたデータは公開されていないため、実験の再現性を考慮して公開された訓練データとテストデータをそのまま使用する。

実験で使用されるラベルなしデータは各対象単語について BCCWJ からランダムに選んだ 200 文の用例文を使用する。BCCWJ はコーパスの規模が大きいため、出現する用例文をすべてラベルなしデータとして追加すると、グラフが大きすぎて実験をすることができない。利用可能な環境で語義識別モデルの学習を完了するために、ラベルなし

データの数を限定して実験を行った。

4章に示した提案手法において、訓練データとラベルなしデータの用例文に対して双方向言語モデルに入力する単語列を求める際に、MeCabを用いて形態素解析を行う。このとき、MeCabの辞書にはUniDicのバージョン2.3.0を使用する。SemEval-2010日本語WSDタスクデータには形態素解析によって各単語に品詞、読み、基本形の情報が与えられているが、今回の実験では使用していない。同じ入力文でもUniDicのバージョン違いによって異なる形態素解析結果とならないように、すべての用例文に対して同じUniDicのバージョン2.3.0を使用する。

4.3.3項におけるBERTの同義判定でグラフ構造を作成する際、判定に必要なBERTの日本語学習済みモデルとしてNWJC-BERT[31]を用いる。NWJC-BERTは「国語研日本語ウェブコーパス」に対して、形態素解析器MeCabとUniDicを用いて形態素解析を行って見出し語に変換した単語列をBERTで学習した言語モデルである。このNWJC-BERTのパラメータは固定して使用する。また、実験において同義判定モデルを構築するための訓練データには、SemEval-2010日本語WSDタスクデータの各対象単語について、訓練データ50文どうしで用例文ペアを作成した1,225件のペアを訓練データとして使用する。用例文ペアの各用例文をNWJC-BERTと同じ形態素解析を行い、見出し語の単語列に変換してBERTの入力トークン列作成のために使用する。

4.2節と4.3.1項、4.3.2項において、素性である文脈ベクトルを求めるために日本語で事前学習されたELMoモデルが必要となる。本論文での実験では、文献[32]の研究で構築、公開されているELMoの日本語学習済みモデルを使用する\*1。これは楽天市場に投稿された商品レビューとYahoo!映画に投稿された映画レビューのデータを用いて作られた事前学習済みモデルである。

3章のPlanetoidでは、FF-NN(1)の学習を10,000回、FF-NN(2)の学習とFF-NN(3)の学習を各1,000回行った。これらのモデルに対して、訓練データを用いて素性とグラフの分散表現を連結し、それを入力として語義ラベルを出力するための学習を1,000回繰り返した。

## 5.2 英語データによる英語WSD実験

### 5.2.1 英語語義曖昧性解消評価データセット

英語を対象とした評価実験として、SENSEVAL-2 English Lexical Sampleタスクデータ[6]も使用する。このデータには73語の対象単語があり、29個の名詞単語、29個の動詞単語、および15個の形容詞単語で構成されている。全対象単語において、訓練データとして使用する用例文が8,611個、テストデータとして使用する用例文が4,328

個用意され、対象単語にはWordNet 1.7プレリリース版の語義が付与されている。

### 5.2.2 英語ラベルなし用例文データ

英語のラベルなし用例文データにはOMSTI(One Million Sense-Tagged Instances)を使用する。OMSTIは813,798文からなるコーパスで、WordNet 3.0の語義が付与されている。今回の実験では、SENSEVAL-2のデータとは異なるバージョンの語義が使われているため、ラベルなし用例文データとして使用する。このデータから各対象単語について最大1,000件の用例文を抽出して、グラフ構造の作成に利用する。

### 5.2.3 英語語義曖昧性解消実験の設定

実験で使用するラベルなしデータは各対象単語についてOMSTIから最大1,000文の用例文をランダムに選んで使用する。日本語での実験とは異なり、ラベルなしデータが少数しか存在しない対象単語が存在する。そのため、ラベルなしデータの数が多数存在する場合のみ、使用する用例文の数を限定して実験を行う。

4.3.3項における同義判定で使用するBERTのトークナイザー学習済みモデルには'bert-base-cased'を利用する。'bert-base-cased'は書籍と英語版Wikipediaに対して、WordPieceで単語分割を行い、30,000語の語彙を用いて得られた単語列をBERTで学習した言語モデルである。この英語学習済みモデルを用いて、用例文ペアの同義判定を行う。'bert-base-cased'のパラメータは固定して使用する。また、SENSEVAL2 English Lexical Sampleタスクの各対象単語について、訓練データにあるすべての用例文ペアを同義判定モデル構築のための訓練データとして使用する。

4.2節と4.3.1項、4.3.2項において必要な英語の学習済みELMoモデルには、AllenNLPが公開する英語の'Original'モデルを用いる\*1。このモデルはニュース記事を集めた'1 Billion Word Benchmark'コーパスで学習されたものである。各データに含まれる用例文はスペースを区切り文字として分割して単語列に変換する。用例文に対して見出し語化は行わずに出現形の単語列を求め、この単語列を学習済みELMoモデルに入力して文脈ベクトルを生成する。

3章のPlanetoidでは、FF-NN(1)の学習を10,000回、FF-NN(2)の学習とFF-NN(3)の学習を各1,000回行った。これらのモデルに対して、訓練データを用いて素性とグラフの分散表現を連結し、それを入力として語義ラベルを出力するための学習を1,000回繰り返した。

## 5.3 評価方法

今回の実験では、語義曖昧性解消手法の評価指標として精度のマイクロ平均を用いて比較評価を行う。

語義曖昧性解消は分類問題として扱われることから、その評価は一般的に精度、再現率、F値の指標を用いて行わ

\*1 <https://allennlp.org/elmo>



表 1 テストデータに対する予測と正解を示した混同行列

Table 1 Confusion Matrix for a WSD Classifier.

		正解語義	正解以外の語義
モデル予測	予測語義	TP	FP
	上記以外の語義	FN	TN

れる。これらの指標を定式化するため、表 1 に示すようなテストデータに対する 4 種類の予測と正解の組合せを用いて説明する。テストデータとして  $N$  個の用例文を用意し、各用例文をモデルに入力したときの予測語義を求め、それらの予測語義が正解語義と一致する用例文の数が TP、正解語義と一致しない用例文の数が FP と定義する。このとき、 $TP + FP = N$  となる。これらの値から、精度 (Precision) はモデルが語義を予測した用例文数に対して正解語義と一致した用例文数の割合として定義され、以下の式となる。

$$\text{Precision} = \frac{TP}{TP + FP}$$

また、テストデータである用例文  $N$  個の正解語義に対して、各正解語義がモデルの予測語義と一致する用例文の数を TP、予測語義と一致しない用例文の数を FN と定義する。このとき、 $TP + FN = N$  となる。これらの値から、再現率 (Recall) はテストデータに対して用例文の正解語義がモデルの予測語義と一致した用例文数の割合として定義され、以下の式となる。

$$\text{Recall} = \frac{TP}{TP + FN}$$

ただし、表 1 にある TN の値については、精度と再現率の計算には必要がないため考慮しない。

F 値は精度と再現率の調和平均として定義され、以下の式で表される。

$$F = \frac{2\text{Recall} \cdot \text{Precision}}{\text{Recall} + \text{Precision}}$$

今回の実験において精度と再現率を計算する場合、上述のように  $TP + FP = TP + FN = N$  となるため、精度と再現率が同じ値になる。さらに、調和平均である F 値も精度と同じ値となる。したがって、本論文では精度のみを用いて語義曖昧性解消手法の比較評価を行う。

精度にはマイクロ平均とマクロ平均があるが、今回の実験ではマイクロ平均を使用して評価する。マイクロ平均は、各セットの正解値とデータ数をそれぞれ合計してから評価値を計算する。たとえば、 $N$  個のテストデータのセットを用意し、 $i$  個目のテストデータのセットは  $n_i$  個データが存在する場合を考える。このとき、 $n_i$  個中  $x_i$  個正解したとすると、精度のマイクロ平均は以下の式で表される。

$$\text{Precision}_{\text{micro}} = \frac{\sum_{i=1}^N x_i}{\sum_{i=1}^N n_i}$$

## 5.4 比較手法

- word2vec の単語分散表現と MLP, Planetoid を用いた手法

nwjc2vec [33] を用いて word2vec の単語分散表現を取得し、5 つの単語に対して 200 次元の単語分散表現を連結した 1,000 次元のベクトルを素性とした語義曖昧性解消モデルを用いて実験を行う。MLP は中間層のノード数を 50、最適化手法として確率的勾配降下法を使用、学習の繰返し数であるエポック数を 50、バッチサイズを 200 として、モデルの学習を行う。

また、word2vec の単語分散表現と 4.3.3 項で説明した BERT のグラフ構造を 3 章の Planetoid で学習した語義曖昧性解消モデルでも実験を行う。BERT に入力する単語列には対象単語と前後 3 単語を使用する。

- ELMo による単語分散表現を用いた手法

4.2 節で説明した ELMo による 5,120 次元の単語分散表現を素性とした語義曖昧性解消モデルを用いて実験を行う。この手法の MLP は word2vec の場合と同じ設定でモデルの学習を行う。

- 藤田らの手法

半教師あり学習を用いた日本語語義曖昧性解消の既存手法として、藤田らの手法 [16] との精度比較を行う。文献 [16] では書籍 (PB) の用例文を追加した場合に精度が最も高かったが、一般的に追加すべき用例文の分野は分からないため、BCCWJ 全部を対象として追加した結果と比較を行う。また、別の手法では [17]、岩波国語辞典の用例文を訓練データに追加したうえで半教師あり学習を行うことで、高い識別精度が得られている。しかし、予備実験において辞書の用例文を追加しても効果が得られなかったため、本実験では文献 [16] を比較対象として採用する。

- Taghipour and Ng (2015) の手法

素性ベクトル word2vec の分散表現を連結した素性を用いた教師あり学習語義曖昧性解消手法 [5] の精度を参考として比較する。

- Kågebäck and Salomonsson (2016) の手法

素性として双方向 LSTM で得られる分散表現を用いた教師あり語義曖昧性解消手法 [7] の精度を参考として比較する。

- Sousa et al. (2020) の手法

半教師あり学習を用いた英語語義曖昧性解消の既存手法として、Sousa et al. (2020) の手法 [15] との精度比較を行う。これは Gaussian Random Field (GRF) を用いてラベルなしデータのラベルを推定する手法で、現時点において既存研究の中で最も精度が高い半教師あり学習手法である。

## 6. 実験結果

### 6.1 SemEval-2010 日本語 WSD タスクの結果

提案手法と比較手法に対して、SemEval-2010 日本語

表 2 SemEval-2010 日本語 WSD タスクの結果

Table 2 Experimental Results on SemEval-2010 Japanese WSD Task.

	精度(%)
藤田らの手法[16]	79.20
word2vec, MLP(グラフなし)	74.79
ELMo, MLP(グラフなし)	79.01
word2vec, Planetoid(BERT)	79.29
提案手法(Cosine 類似度)	80.64†
提案手法(相互 $k$ 近傍)	80.89††
提案手法(BERT 文単位)	80.92††
提案手法(BERT 前後 3 単語)	<b>80.93††</b>

表 3 SENSEVAL2 English Lexical Sample タスクの結果

Table 3 Experimental Results on SENSEVAL-2 Lexical Sample Task.

	精度(%)
(Taghipour and Ng, 2015)	66.2
(Kågebäck and Salomonsson, 2016)	66.90
ELMo, MLP(グラフなし)	65.86
(Sousa et al., 2020)	68.7
提案手法(Cosine 類似度)	71.50†
提案手法(相互 $k$ 近傍)	71.43††
提案手法(BERT 文単位)	71.69††
提案手法(BERT 前後 3 単語)	<b>71.70††</b>

WSD タスクデータを利用した実験における識別精度を表 2 に示す. ELMo を用いたグラフなし手法と提案手法との実験結果の間で  $t$  検定を行い, 両側検定の有意水準 5% を †, 有意水準 1% を †† で満たしていることを表す. 表 2 の結果を見ると, 提案手法はグラフを用いない場合と比較し, 提案手法の識別精度が word2vec, ELMo を素性とした場合よりも高くなった. また, 提案手法において BERT と Cosine 類似度, 相互  $k$  近傍グラフを用いた場合で比較を行うと, 前後 3 単語を入力とした BERT を利用した場合の精度が最も高く, Cosine 類似度を利用した場合より少し高かった. 相互  $k$  近傍グラフと前後 3 単語の BERT, 文単位で入力した BERT については精度差はほとんどなかった. 藤田らの手法 [16] の結果と比較すると, 提案手法は SemEval-2010 日本語 WSD タスクデータを用いた半教師付き学習法において, 既存手法を上回る結果を得た.

## 6.2 SENSEVAL2 English Lexical Sample タスクの結果

SENSEVAL2 English Lexical Sample タスクを利用した半教師あり NN と既存手法の識別精度の比較を表 3 に示す. ELMo を用いたグラフなし手法と半教師あり学習手法の実験結果の間で日本語データと同様に  $t$  検定を行い, 両側

検定の有意水準 5% を †, 有意水準 1% を †† で満たしていることを表す. 表 3 の結果を見ると, 提案手法は, Taghipour and Ng (2015) [5] や Kågebäck and Salomonsson (2016) [7] の教師あり手法, Sousa et al. (2020) [15] の半教師あり語義曖昧性解消手法よりも高い精度を達成している. 提案手法と ELMo を素性とした MLP の結果を比較すると, すべてのグラフ作成手法で提案手法の方が高い識別精度となっている. 提案手法のグラフ作成方法を比較すると, 識別精度に差はほとんどなかった.

## 7. 考察

### 7.1 グラフありとグラフなしの比較

SemEval-2010 日本語 WSD タスクデータを利用した表 2 に示すように, 提案手法と ELMo の分散表現を素性とする MLP において 5% の有意差検定を行ったところ, すべてのグラフ作成方法について有意な差がみられた. SENSEVAL2 English Lexical Sample タスクデータを利用した表 3 においても同様に, 提案手法と ELMo の分散表現を素性とする MLP において 5% の有意差検定を行ったところ, すべてのグラフ作成方法について有意な差がみられた.

以上のことから SemEval-2010 日本語 WSD タスクデータと SENSEVAL2 English Lexical Sample タスクデータにおいて, グラフ構造を使用した半教師あり語義曖昧性解消手法が語義曖昧性解消に効果があるということが分かった.

### 7.2 BERT と Cosine 類似度, 相互 $k$ 近傍グラフの比較

表 2 におけるグラフを作成する際の類似度計算方法を変更して比較すると, BERT を利用したグラフで最も高い識別精度となった. この実験結果と相互  $k$  近傍グラフでの実験結果に対して 5% の有意差検定を行ったところ, 有意な差はみられなかった. これは, Planetoid のグラフ学習による最適化で同じような精度になったと考えられる. この結果から, BERT によるグラフ, Cosine 類似度を用いたグラフ, 相互  $k$  近傍グラフのどのグラフを使用しても高い精度で語義を識別できることが分かった.

### 7.3 同義文判定における入力文の長さの比較

SemEval-2010 日本語 WSD タスクを利用した結果の表 2 では, 同義文判定における入力文の長さにおける WSD 精度差はほとんどみられなかった. SENSEVAL2 English Lexical Sample タスクを利用した表 3 においても, 同義文判定における入力文の長さにおける WSD 精度差はほとんどみられなかった. このことから同義文判定における入力文の長さは提案手法においてはほとんど精度に影響がでないことが考えられる.

### 7.4 半教師あり学習の既存手法との比較

SemEval-2010 日本語 WSD タスクデータの文脈ベクト

ルを用いた実験結果である表 2 の提案手法 (BERT 前後 3 単語) の精度と藤田らの手法の精度を比較すると, 提案手法の精度の方が 1.73 ポイント高かった. そのため, 半教師あり学習手法としては提案手法の方が語義曖昧性解消の効果が高いと考えられる. また, 提案手法は藤田らの手法におけるデータ追加に比べると, 文どうしのペアを作成または類似度を計算すれば比較的簡単にラベルなし用例文を追加することが可能である. 辞書に記載されていない新しい語義も用例文を追加することで, 語義のラベルを容易に加えることもできる. 藤田らの手法のように辞書の用例文を用いて質の高い語義ラベル付き用例文を追加することで, 提案手法においても語義曖昧性解消の精度がさらに向上することが期待できる.

SENSEVAL2 English Lexical Sample タスクデータの実験結果である表 3 の提案手法の精度と既存手法である Taghipour and Ng (2015) [5], Kågebäck and Salomonsson (2016) [7], Sousa et al. (2020) [15] の精度を比較すると, 提案手法の精度が高くなった. そのため, 英語の語義曖昧性解消においても, 半教師あり学習手法としては提案手法の方が語義曖昧性解消の効果が高いと考えられる.

### 7.5 日本語単語の難易度別精度比較

SemEval-2010 日本語 WSD タスクデータの対象単語を難易度に分けて, 各手法の精度を難易度ごとに比較する. ここでは既存研究との比較を容易にするため, 対象単語の難易度を文献 [18] において設定された対象単語の難易度を用いる. この難易度はテストデータにおける語義の頻度分布のエントロピーを用いて分類可能であるが, 訓練データや形態素解析器によって単語の出現頻度が変化するため難易度は変化してしまう [16], [18]. 今後の研究における難易度別評価を容易にするため, 以前に設定された対象単語の難易度を用いることとする. 対象単語の難易度は低難易度 ( $D_{easy}$ ), 中難易度 ( $D_{mid}$ ), 高難易度 ( $D_{diff}$ ) の 3 段階に分かれ, 各難易度に含まれる対象単語の数は  $D_{easy}$  が 11 単語,  $D_{mid}$  が 19 単語,  $D_{diff}$  が 20 単語である. 各難易度における対象単語は以下のようにになっている.

- $D_{easy}$ : 「生きる・活きる」「大きい」「考える」「経済」「強い」「場所」「一つ」「開く」「文化」「外・他」「遣る」
- $D_{mid}$ : 「相手」「教える」「可能」「関係」「技術」「現場」「子供」「時間」「社会」「情報」「高い」「電話」「場合」「初め・始め」「始める」「早い・速い」「求める」「もの」「良い・善い・好い」
- $D_{diff}$ : 「合う・会う・遭う・遇う・逢う」「上げる・揚げる・挙げる」「与える」「意味」「入れる・容れる」「市場」「進める・勧める・薦める・奨める」「為る」「出す」「立つ・建つ・発つ」「手」「出る」「取る・採る・執る・捕る」「乗る・載る」「入る・這入る」「前」「見える」「認める」「見る・視る・観る」「持つ」

表 4 難易度ごとの語義識別精度 (%)

Table 4 Precision for Each Difficulty Class.

手法	$D_{easy}$	$D_{mid}$	$D_{diff}$
文献[18]の最頻出語義	95.09	68.21	55.30
文献[18]の Baseline	94.18	74.11	66.00
文献[18]の RALI-2 [34]	94.00	75.58	67.40
word2vec, MLP(グラフなし)	94.06	74.91	64.07
ELMo, MLP(グラフなし)	95.15	79.23	69.93
word2vec, Planetoid(BERT)	<b>95.58</b>	77.79	71.77
提案手法(Cosine 類似度)	95.39	79.65	73.47
提案手法(相互 $k$ 近傍)	95.52	79.40	74.27
提案手法(BERT 文単位)	95.21	79.33	<b>74.57</b>
提案手法(BERT 前後 3 語)	95.52	<b>79.82</b>	73.97

表 4 に難易度ごとに分けた各手法の語義識別精度を示す. 表 4 には参考として SemEval-2010 日本語 WSD タスクデータでのベースラインシステムの精度と最も結果の良かった手法 [34] の精度も掲載した [18]. 表 4 を見ると, 提案手法は高難易度において他の手法よりも高い精度で識別している. これはグラフニューラルネットワークを用いた半教師あり学習の効果が高いと考えられる. word2vec を用いた素性の場合には 64.07% から 71.77% と 7.7 ポイントの精度向上, ELMo を用いた素性の場合には 69.93% から提案手法で最も高い 74.57% と 4.64 ポイントの精度向上がみられた. どのような素性であっても, 同義関係を表すグラフを用いた半教師あり学習を行うことが高難易度の単語に対する語義曖昧性解消において有効であることが分かる. また, 提案手法は Cosine 類似度によるグラフ (73.47%), 相互  $k$  近傍グラフ (74.27%), BERT によるグラフ (文単位で 74.57%, 前後 3 単語で 73.97%) のどれを利用してほとんど同じ識別精度であることから, どちらのグラフを利用して Planetoid において学習を繰り返すことで語義曖昧性解消に有効なモデルが得られるのではないかと考えられる.

中難易度では, 提案手法と素性に ELMo を用いた MLP において高い精度で識別できていることから, ELMo で得られる文脈ベクトルの効果が現れていると考えられる. 素性として ELMo と word2vec を比較した場合, 提案手法で最も高い精度が 79.82%, Planetoid (素性が word2vec, グラフが BERT) の識別精度が 77.79% となり, 提案手法の方が 2.03 ポイント高い識別精度である. また, グラフを使わない MLP を識別モデルとした場合も word2vec で 74.91%, ELMo で 79.23% と, ELMo を素性とする方が 4.32 ポイント高い精度であった. しかし, グラフの有無による精度の差がみられなかった. このことから, 中難易度では素性として使用した文脈ベクトルが語義曖昧性解消に大きく影響していると考えられる.

低難易度では, 提案手法は最頻出語義の精度を超える高い精度で識別できていることが分かる. 従来研究でも 94% 以上の高い精度で識別できているが, 最頻出語義の精

度を超えることができなかつたため、低難易度では多数を占める最頻出語義を選ぶ手法が効果的だといわれてきた。たとえば、対象単語「強い」は語義1が50文のうち46文存在するため最頻出語義を選ぶと92%の精度で識別するが、提案手法では平均92.67%の精度で識別できている。このことから、提案手法は最頻出語義だけでなく、少数しか出現しない語義も含めて語義曖昧性解消に有効であると考えられる。

## 8. 結論

本論文では、機械学習に基づく語義曖昧性解消における訓練データ不足の問題に対処するため、用例文間の類似関係を用いた半教師あり語義曖昧性解消手法を提案し、その有効性について分析を行った。提案手法の有効性を分析するために、Semeval-2010日本語語義曖昧性解消タスクのデータセットとSENSEVAL2 English Lexical Sampleタスクデータを用いて評価実験を行った。実験の結果、どちらのデータセットにおいても、提案手法は既存の最も精度の高い半教師あり手法と比較して語義識別精度が向上したことから、半教師あり学習に基づく語義曖昧性解消において同義関係を表したグラフ構造から得られる分散表現が語義曖昧性解消に有効であることが分かった。また、同義関係を表すグラフにおいて、言語モデルのBERTを用いて同義かどうかを判定するパラメータ設定のないグラフ構築手法の提案も行った。このグラフ構築手法とパラメータ設定が必要なCosine類似度や相互 $k$ 近傍グラフによるグラフ構築手法はどちらを用いても高い精度で語義を分類できることを示した。

今後の課題として、グラフ構造を作成する際に言い換え技術の利用や係り受け情報の利用を行うこと、計算時間を短縮するためにラベルなし用例文をフィルタリングすること、岩波国語辞典の用例文を効果的に追加することがあげられる。

謝辞 本研究はJSPS科研費18K11422の助成を受けたものです。

## 参考文献

[1] 西尾 実, 岩淵悦太郎, 水谷静夫: 岩波国語辞典, 岩波書店 (1994).

[2] Navigli, R.: Word sense disambiguation: A survey, *ACM Comput. Surv.*, Vol.41, No.2, article 10, pp.1–69 (2009).

[3] 新納浩幸, 村田真樹, 白井清昭, 福本文代, 藤田早苗, 佐々木稔, 古宮嘉那子, 乾 孝司: クラスタリングを利用した語義曖昧性解消の誤り原因のタイプ分け, *自然言語処理*, Vol.22, No.5, pp.319–362 (2015).

[4] Sugawara, H., Takamura, H., Sasano, R. and Okumura, M.: Context Representation with Word Embeddings for WSD, *Proc. PACLING 2015*, pp.108–119 (2015).

[5] Taghipour, K. and Ng, H.T.: Semi-Supervised Word Sense Disambiguation Using Word Embeddings in General and Specific Domains, *Proc. NAACL-HLT2015*,

pp.314–323 (2015).

[6] Yuan, D., Richardson, J., Doherty, R., Evans, C. and Altendorf, E.: Semi-supervised word sense disambiguation with neural models, *COLING 2016*, pp.1374–1385 (2016).

[7] Kågebäck, M. and Salomonsson, H.: Word Sense Disambiguation using a Bidirectional LSTM, *Proc. 5th Workshop on Cognitive Aspects of the Lexicon (CogALex-V)*, pp.51–56 (2016).

[8] Raganato, A., Delli Bovi, C. and Navigli, R.: Neural Sequence Learning Models for Word Sense Disambiguation, *Proc. EMNLP*, pp.1156–1167 (2017).

[9] Huang, L., Sun, C., Qiu, X. and Huang, X.: GlossBERT: BERT for Word Sense Disambiguation with Gloss Knowledge, *Proc. 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp.3509–3514 (2019).

[10] Kumar, S., Jat, S., Saxena, K. and Talukdar, P.: Zero-shot Word Sense Disambiguation using Sense Definition Embeddings, *Proc. 57th Annual Meeting of the Association for Computational Linguistics (ACL2019)*, pp.5670–5681 (2019).

[11] Loureiro, D. and Jorge, A.: Language Modelling Makes Sense: Propagating Representations through WordNet for Full-Coverage Word Sense Disambiguation, *Proc. 57th Annual Meeting of the Association for Computational Linguistics (ACL2019)*, pp.5682–5691 (2019).

[12] Vial, L., Lecouteux, B. and Schwab, D.: Sense Vocabulary Compression through the Semantic Knowledge of WordNet for Neural Word Sense Disambiguation, *Proc. 10th Global WordNet Conference (GWC2019)* (2019).

[13] Blevins, T. and Zettlemoyer, L.: Moving Down the Long Tail of Word Sense Disambiguation with Gloss Informed Bi-encoders, *Proc. 58th Association for Computational Linguistics (ACL2020)*, pp.1006–1017 (2020).

[14] Bevilacqua, M. and Navigli, R.: Breaking Through the (80%) Glass Ceiling: Raising the State of the Art in Word Sense Disambiguation by Incorporating Knowledge Graph Information, *Proc. 58th Annual Meeting of the Association for Computational Linguistics (ACL2020)*, pp.2854–2864 (2020).

[15] Sousa, S., Milios, E. and Berton, L.: Word sense disambiguation: an evaluation study of semi-supervised approaches with word embeddings, *Proc. 2020 International Joint Conference on Neural Networks (IJCNN)*, pp.1–8 (2020).

[16] 藤田早苗, Kevin Duh, 藤野昭典, 平 博順, 進藤裕之: 日本語語義曖昧性解消のための訓練データの自動拡張, *自然言語処理*, Vol.18, No.3, pp.273–291 (2011).

[17] Fujita, S. and Fujino, A.: Word Sense Disambiguation by Combining Labeled Data Expansion and Semi-Supervised Learning Method, *ACM Trans. Asian Language Information Processing*, Vol.12, No.2, Article No.7 (2013).

[18] Okumura, M., Shirai, K., Komiya, K. and Yokono, H.: On SemEval-2010 Japanese WSD Task, *自然言語処理*, Vol.18, No.3, pp.293–307 (2011).

[19] Kilgariff, A.: English lexical sample task description, *Proc. 2nd International Workshop on Evaluating Word Sense Disambiguation Systems (SENSEVAL '01)*, pp.17–20 (2001).

[20] Zhong, Z. and Ng, H.T.: It Makes Sense: A Wide-Coverage Word Sense Disambiguation System for Free Text, *Proc. ACL 2010 System Demonstrations*, pp.78–

- 83 (2010).
- [21] Yarowsky, D.: Unsupervised Word Sense Disambiguation Rivaling Supervised Methods, *Proc. 33rd Annual Meeting of the Association for Computational Linguistics*, pp.189–196 (1995).
- [22] Mihalcea, R.: Co-training and Self-training for Word Sense Disambiguation, *Proc. 8th Conference on Computational Natural Language Learning (CoNLL-2004)*, pp.33–40 (2004).
- [23] Sasaki, M. and Shinnou, H.: Word Sense Disambiguation Based on Distance Metric Learning from Training Documents, *Proc. 6th International Conference on Advances in Semantic Processing (SEMAPRO2012)*, pp.54–58 (2012).
- [24] Hu, Z., Luo, F., Tan, Y., Zeng, W. and Sui, Z.: WSD-GAN: Word Sense Disambiguation Using Generative Adversarial Networks, *Proc. AAAI Conference on Artificial Intelligence*, Vol.33, No.01, pp.9943–9944 (2019).
- [25] Kouno, K., Shinnou, H., Sasaki, M. and Komiya, K.: Unsupervised Domain Adaptation for Word Sense Disambiguation using Stacked Denoising Autoencoder, *Proc. 29th Pacific Asia Conference on Language, Information and Computation (PACLIC-29)*, pp.224–231 (2015).
- [26] Agirre, E., López de Lacalle, O. and Soroa, A.: Random walks for knowledge-based word sense disambiguation, *Computational Linguistics*, Vol.40, No.1, pp.57–84 (2014).
- [27] Niu, Z., Ji, D. and Tan, C.: Word sense disambiguation using label propagation based semi-supervised learning, *Proc. 43rd Annual Meeting on Association for Computational Linguistics (ACL05)*, pp.395–402 (2005).
- [28] Yang, Z., Cohen, W.W. and Salakhutdinov, R.: Revisiting Semi-Supervised Learning with Graph Embeddings, *Proc. 33rd International Conference on International Conference on Machine Learning (ICML2016)*, Vol.48, pp.40–48 (2016).
- [29] Bottou, L.: Large-scale machine learning with stochastic gradient descent, *COMPSTAT*, pp.177–186 (2010).
- [30] Devlin, J., Chang, M., Lee, K. and Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, *NAACL-HLT*, pp.4171–4186 (2019).
- [31] 浅原正幸, 西内沙恵, 加藤 祥: NWJC-BERT: 多義語に対するヒトと文脈化単語埋め込みの類似性判断の対照分析, 言語処理学会第26回年次大会, pp.961–964 (2020).
- [32] Bataa, E. and Wu, J.: An Investigation of Transfer Learning-Based Sentiment Analysis in Japanese, *Proc. 57th Annual Meeting of the Association for Computational Linguistics (ACL2019)*, pp.4652–4657 (2019).
- [33] Asahara, M.: NWJC2Vec: Word embedding dataset from ‘NINJAL Web Japanese Corpus’, *Terminology*, Vol.24, pp.7–22 (2018).
- [34] Brosseau-Villeneuve, B., Kando, N. and Nie, J.: RALI: Automatic Weighting of Text Window Distances, *Proc. 5th International Workshop on Semantic Evaluation (SemEval-2010)*, pp.375–378 (2010).



谷田部 梨恵

2019年茨城大学工学部情報工学科卒業。2021年同大学大学院理工学研究科情報工学専攻博士前期課程修了，修士（工学）。同年（株）日立産業制御ソリューションズ入社。



佐々木 稔（正会員）

2001年徳島大学大学院博士後期課程修了，博士（工学）。同年茨城大学工学部情報工学科助手，2005年同講師を経て，2021年より同准教授。機械学習に基づく自然言語処理・テキストマイニングに関する研究に従事。言語

処理学会，計量国語学会，人工知能学会，電子情報通信学会各会員。