

訓練事例と辞書用例を異なるモデルで表現した語義曖昧性解消

谷田部梨恵¹ 佐々木稔²

1 茨城大学大学院理工学研究科情報工学専攻

2 茨城大学工学部情報工学科

{19nm732r, minoru.sasaki.01}@vc.ibaraki.ac.jp

1 はじめに

語義曖昧性解消(WSD)は文中の多義語が辞書中のどの語義で使われるのかを識別するタスクである。現在、このWSDはディープラーニングを用いて行うことが主流で、そのシステムにおいて辞書やシソーラスといった外部知識を組み合わせることで有効性が示されている。現在までに、語義定義文と教師あり学習を組み合わせた GlossBERT [1]や BEM [2]、シソーラスの上位下位関係などの単語間関係を組み合わせた EWISE [3]や EWISER [4]などが存在する [5]。

しかし、辞書によっては同じ辞書の定義文を使用しても精度が低下することがある。上記の研究ではすべて WordNet の辞書から抽出されたラベルの意味区分や語義の定義文と用例文を使用している。WordNet を使うことで WSD への有効性は示されているが、他の辞書を使った場合でも有効かどうかは明らかになっていない。そこで、外部知識として日本語を対象とした SemEval2010 日本語 WSD タスクデータで使われる岩波国語辞典の語義定義文を EWISE と同じ手法で予備実験を行ったところ、低い分類精度となってしまい、有効性を確認することができなかった。

岩波国語辞典の語義定義文を使用しても有効性を確認できなかった要因のひとつとして、各語義に含まれる用例文は文の一部しか掲載されていないという点が考えられる。用例文の一部だけしか記述がないことから、他の語義と区別することが難しい用例文がいくつか存在する、例えば、動詞「上げる」に含まれる「よい状態に高める」という語義において、「腕を上げる(上達する)」や「上げたり下げたり(ほめたりくさしたり)」といった用例文が紹介されている。これらの用例文は前後の文脈によって「物を低い所から高い所にもってゆく」という別の語義として使用することも可能である。このため、

有効な語義の分散表現を求めることができなかったのではないかと考えられる。

そこで我々は、語義ラベルとして使用した辞書とは異なる辞書の情報を効果的に利用した WSD システムの構築を目指している。従来手法では辞書情報をモデル化する際に訓練データと同じ辞書が用いられるが、異なる辞書から得られるモデルを組み合わせることで有効な WSD を行うことが可能かどうか検討する。本稿では、完全な文として用例文が記述されている辞書として国立国語研究所の「基本動詞ハンドブック」に含まれる用例文を使用する [6]。この用例文集合を使用したニューラルネットワークのモデルを辞書のモデルとする。このモデルと SemEval2010 日本語 WSD タスクデータから学習したモデルを組み合わせることで WSD を行い、異なる辞書から構築されたモデルが有効かどうか検証する。

2 使用データと対象単語

本稿で使用するデータとして、日本語 WSD の評価データである SemEval2010 日本語 WSD タスクデータ [7]と利用可能な日本語辞書である「基本動詞ハンドブック」を使用する。

2.1 SemEval2010 日本語 WSD データ

SemEval2010 日本語 WSD タスクデータ(SE)は 50 個の対象単語に対して WSD の精度を評価するためのデータである [7]。各対象単語について、訓練データとテストデータとしてそれぞれ 50 個の用例文が用意されており、岩波国語辞典の意味区分に従ってラベル付けされている。

2.2 基本動詞ハンドブック

「基本動詞ハンドブック」(HB)は日常生活でよく使われる基本的な動詞単語の語義を日本語学習者や日本語教師に向けて解説したオンラインツールである。動詞単語に対して、語義の定義と 5 件程度の用例文、文法情報などが記載されている。用例文には

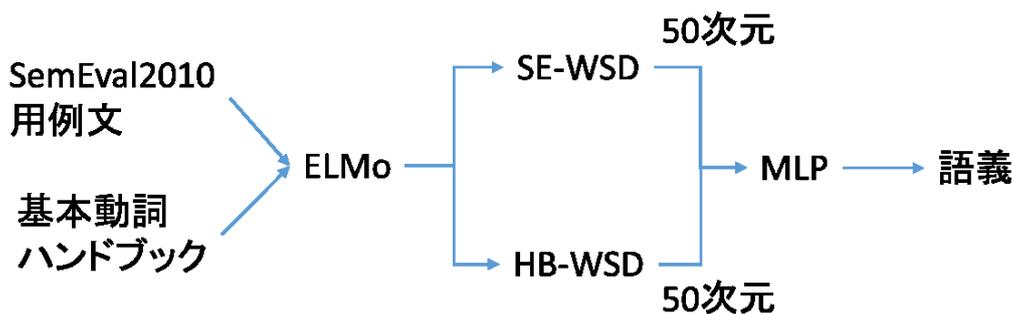


図1 並列 WSD モデル

手作業で岩波国語辞典の語義をラベル付けした。ハンドブックの語義定義文に対応する語義が岩波国語辞典に存在するとき、その語義のラベルを付与する。ハンドブックの語義が岩波国語辞典に存在しない場合やひとつの語義が複数の語義に対応する場合は、それらの用例文にはラベルを付与せず、使用しないこととした。

2.2 対象単語

対象単語は SemEval2010 日本語 WSD タスクデータと基本動詞ハンドブックの両方に登録された 10 個の動詞単語とする。具体的な対象単語は「あげる」、「上げる」、「揚げる」、「挙げる」、「いきる」、「生きる」、「活きる」、「出す」、「立つ」、「建つ」、「発つ」、「出る」、「のる」、「乗る」、「載る」、「見える」、「見る」、「持つ」、「やる」の 10 単語である。

3 WSD モデル

WSD モデルは訓練データである用例文を入力して、対象単語の語義推定を行う。異なる訓練データを用いて WSD モデルを構築し、その分類精度の比較を行う。

3.1 用例文の文脈ベクトル化

入力する用例文は対象単語前後の文脈を考慮したベクトルに変換される。文脈ベクトルへ変換するために、事前学習された ELMo モデルを用いて、用例文の分散表現を得る。このモデルは文脈情報を考慮した 2 層双方向 LSTM 構造で、本稿では ELMo モデルの最後の層の出力のみを使用する。

入力された用例文に対して対象単語及び前後 2 単語を抽出し、ELMo モデルに入力して 5 単語の分散

表現を得る。これを連結した 5120 次元ベクトルを WSD モデルに入力する文脈ベクトルとして使用する。

3.2 SE-WSD モデル

SemEval2010 日本語 WSD タスクデータを用いた WSD モデルを構築する。モデルには多層パーセプトロン(MLP)を使用する。今回の実験で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、学習の繰り返し数であるエポック数を 50 と設定して分類モデルの学習を行う。

3.3 HB-WSD モデル

「基本動詞ハンドブック」の用例文を用いた WSD モデルを構築する。モデルは SE-WSD と同じ多層パーセプトロン(MLP)を使用する。実験でも SE-WSD と同じパラメータ設定で分類モデルの学習を行う。

3.4 TR-WSD モデル

SemEval2010 日本語 WSD タスクデータの訓練データと「基本動詞ハンドブック」の用例文をひとつの訓練データにまとめて WSD モデルを構築する。SE-WSD と同じ MLP とパラメータ設定を使用して学習を行う。

3.5 並列 WSD モデル

並列 WSD モデルは学習済みの SE-WSD モデルと HB-WSD モデルを並列化したモデルで、訓練データのモデルと外部知識のモデルを異なるモデルとして分離したモデルである。図 1 に並列 WSD モデルの概要を示す。SemEval の訓練データとハンドブック

表 1 実験結果 (4 モデルのうち最も高い精度を太字で示す)

対象単語	SE-WSD	HB-WSD	TR-WSD	並列 WSD
あげる	0.487	0.453	0.567	0.520
いきる	0.940	0.967	0.953	0.973
だす	0.480	0.540	0.587	0.580
たつ	0.513	0.513	0.607	0.573
でる	0.713	0.653	0.660	0.607
のる	0.773	0.540	0.720	0.700
みえる	0.667	0.533	0.667	0.613
みる	0.793	0.700	0.780	0.773
もつ	0.827	0.820	0.867	0.833
やる	0.940	0.940	0.940	0.940
平均精度	0.713	0.666	0.735	0.711

の用例文に対して、用例文の文脈ベクトルを学習済みの SE-WSD モデルと HB-WSD モデルに入力し、その中間層の 50 次元ベクトルを生成する。2 つの中間層のベクトルを連結した 100 次元ベクトルを MLP に入力して、再度この MLP の訓練を行う。この MLP についても、中間層のサイズや学習パラメータ設定は SE-WSD と HB-WSD と同じものを使用する。

4 実験

本節では、3 節に示した 5 種類の WSD システムの分類精度を分析するために、10 個の動詞単語について評価データによる精度比較を行う。精度評価は Semeval2010 日本語 WSD タスクデータのテストデータを使用する。

4.1 実験結果

表 1 に 4 つの WSD モデルを用いた場合の語義識別精度を示す。平均精度は SE-WSD が 71.3%、HB-WSD が 66.6%、TR-WSD が 73.5%、並列 WSD は 71.1%であった。これらのモデルの中では TR-WSD の精度が最も高いことから、辞書中の用例文も訓練データに追加して学習をする方が効果的であると考えられる。

SE-WSD と HB-WSD を比較すると、HB-WSD は多くの動詞単語で精度が SE-WSD を上回ることができなかった。HB-WSD は「いきる」と「だす」の 2 単語で精度が高かった。どちらも少数語義の用例文を正しく識別できたことが要因だと考えられる。しかし、他の多くの単語で SE-WSD の方が高い精度で

あった。辞書の意味区分が異なるため語義の出現分布が偏ったことが要因として考えられる。また、「のる」のように辞書から語義の用例文を均等に抽出したことが要因で識別誤りが増えた単語も存在する。

一般的に、TR-WSD のように訓練データに辞書の用例文を追加すれば多くの単語で精度の上昇が期待できる。このとき、追加する用例文は岩波国語辞典のように用例文の一部を追加するのではなく、文として追加することで精度が上がると考えられる。今回の実験では訓練データと辞書用例文のモデルを並列化しても有効とは言えず、外部知識を使う場合でも用例文はひとつのモデルに集約する方が効果的であると考えられる。

5 おわりに

本稿では、語義ラベルで使った辞書とは異なる辞書の情報を異なるモデルで表現する WSD の効果について分析を行った。今回の実験では異なる辞書の用例文を訓練データに追加した方が高い精度であった。異なる辞書の効果的な活用方法と外部知識の効果的なモデル化を行うことが今後の課題である。

参考文献

1. HuangLuyao, ほか. GlossBERT: BERT for Word Sense Disambiguation with Gloss Knowledge. : the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on

Natural Language Processing (EMNLP-IJCNLP), 2019. ページ: pp. 3509-3514. .

2. **BlevinsTerra , ZettlemoyerLuke.** Moving Down the Long Tail of Word Sense Disambiguation with Gloss Informed Bi-encoders. : the 58th Association for Computational Linguistics (ACL2020), 2020. ページ: pp. 1006-1017. .

3. **KumarSawan, ほか.** Zero-shot Word Sense Disambiguation using Sense Definition Embeddings. : the 57th Annual Meeting of the Association for Computational Linguistics (ACL2019), 2019. ページ: pp. 5670-5681. .

4. **BevilacquaMichele , NavigliRoberto.** Breaking Through the (80%) Glass Ceiling: Raising the State of the Art in Word Sense Disambiguation by Incorporating Knowledge Graph Information. : the 58th Annual

Meeting of the Association for Computational Linguistics (ACL2020), 2020. ページ: pp. 2854-2864. .

5. **LoureiroDaniel , JorgeAlipio.** Language Modelling Makes Sense: Propagating Representations through WordNet for Full-Coverage Word Sense Disambiguation. : the 57th Annual Meeting of the Association for Computational Linguistics (ACL2019), 2019. ページ: pp. 5682-5691. .

6. **国立国語研究所.** 『基本動詞ハンドブック』.(オンライン) <https://verbhandbook.ninjal.ac.jp>.

7. **OkumuraManabu, ほか.** Semeval-2010 task: Japanese WSD. : the fifth International Workshop on Semantic Evaluation (SemEval-2010), 2010. ページ: pp. 69-74. .