BERT の学習済みモデルを用いた用例文ペアの同義判定

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（Word Sense Disambiguation, WSD）は文中に出現するある多義語が辞書中のどの語義として使われているかを推定するタスクである．このタスクはコンピュータが意味を理解するための基礎的な問題のひとつで，機械学習や文書要約など様々な自然言語処理タスクにおいて重要な役割を担うと考えられている．WSD は今日に至るまで様々なアプローチで研究が行われており，高い識別性能を持つのはコーパスを用いた教師あり学習に基づく手法であることが知られている．しかし，用例文に出現する単語に対して専門家が適切に語義を割り当てる必要があるため，語義タグ付きコーパスを大量に作成することはコストがかかるという問題がある．そのため，少量の語義タグ付きコーパスと大量の語義なしコーパスを用いた半教師あり学習に基づく WSD 手法も研究が行われている．

この半教師あり学習に基づくアプローチとして，Graph Convolutional Network (GCN)を用いた WSD 手法が存在する[1]．これは用例文間の類似性を表現したグラフ構造と従来の素性ベクトルを組み合わせて語義を識別する手法である．グラフ構造はコーパス中の各用例文をノードとし，類似する用例文ペアを辺で結んで作成され，用例文の類似性を表現している．このとき，類似性については Jaccard 係数や Cosine 類似度を使って判定が行われ，WSD の評価実験において高い識別精度が得られている．

しかし，この手法は対象単語前後の単語数に加えて，類似判定を行うための閾値を設定する必要があり，閾値を適切に設定することが難しいという問題がある．既存手法はすべての単語に対して一定の閾値で類似判定を行っているが，各単語によって適切な閾値は異なると考えられる．閾値を使わない相互k-近傍グラフでは，2 つのノードのk-近傍に互いのノードが存在するときに，これらのノードを辺で結ぶ[2][3]．この相互k-近傍グラフもパラメータkを設定する必要がある．そのため，前後の単語数以外に，閾値などのパラメータを設定しなくても，WSD に有効なグラフ構造を作成することが課題となっている．

そこで本稿では，対象単語前後の単語数を設定するだけで作成可能な，用例文間の関係を表現するグラフ構造の作成手法を提案する．提案手法は対象単語を含む用例文ペアに対して，汎用的な言語モデルである BERT [4]の学習済みモデルを用いて，同じ語義で使われているかどうかを判定するものである．これにより，用例文間の関連性をグラフ構造として適切に表現することが可能となると考えられる．提案手法の有効性を確認するために，Semeval2010 日本語 WSD タスクデータから作成したデータセットを用いて比較実験を行い，提案手法の有効性を示す．

Copyright(C) 2020 The Association for Natural Language Processing.
All Rights Reserved.

2. 用例文ペアの同義判定手法

2.1 Jaccard 係数，Cosine 類似度を用いる方法

対象単語を含むすべての用例文に対して，MeCab と UniDic を用いて形態素解析を行い前後二単語, 品詞, 品詞大分類, 係り受け, シソーラス情報を素性として抽出する．各用例文の素性集合について，類似度が最大となる用例文, さらに類似度が 0.9 以上（同じ語義であるという確信度が高い)である用例文を同義であると判定する．

ここで使用する類似度には，Jaccard 係数と Cosine 類似度の 2 種類を用いる．Jaccard 係数は二つの集合間で共通する素性の比率である．各用例文に含まれる素性集合 A と B が与えられた場合，Jaccard 係数 Jは以下のように表される．

$J(A,B) = |A \cap B|/|A \cup B|, (0 \leq J(A,B) \leq 1)$

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.

2.2 相互k-近傍グラフを用いる方法

相互k-近傍グラフは任意の二点間のノードがk-近傍でお互いに含まれている場合にのみエッジを張るグラフとなっている[2][3]．これを用いた判定方法として，まず用例文集合を 2.1 節と同様に素性を抽出し，この素性を用いて用例文間のユークリッド距離を計算する．各用例文に対して$k = 3$としてk-近傍を求め，エッジを張った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張る.

2.3 BERT を用いる方法

BERT は汎用的な言語モデルの事前学習を行う手法のひとつである[4]．大量のテキストデータを事前学習したモデルに少量の教師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．本論文では，BERT の学習済みモデルには国立国語研究所で作成された NWJC-BERT 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．この学習済みモデルの出力を 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.

まず, 各対象単語ついて, 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 つの用例文に対してそれぞれ MeCab と UniDic を用いて形態素解析を行って見出し語の単語列を求める. この単語列から抽出した対象単語の前後3単語を判定モデルに入力することで，同義と非同義の確率が出力される．訓練データにおいて，モデルからの出力と正解の同義ラベルとの誤差を最小化するように繰り返し学習が行われる.

同義かどうか分からない用例文ペアの同義判定を行う場合も学習時と同様に形態素解析を行い, 対象単語の前後3単語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も同義と非同義の確率となり，確率の大きい同義ラベルを判定結果として出力する.

Copyright(C) 2020 The Association for Natural Language Processing.
All Rights Reserved.

3. 実験

本研究において同義判定を行う対象単語は，Semeval2010 日本語 WSD タスクデータである対象単語の 50 個を利用する[5]. このデータには，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. ここから用例文ペアを作成し，岩波国語辞典第五版タグの中分類まで一致した場合，同じ語義として「1」，異なる場合は「0」としてラベル付けを行い，評価データを作成した.

同義判定の訓練データにはSemevalの訓練データ 50 文同士で用例文ペアを作成した 1,225 件のデータを使用する．また，同義判定のテストデータにはSemevalの訓練データとテストデータの用例文ペアである 2,500 件を使用する. このテストデータに対して同義判定を行い, 自動識別でラベル（「0」か「1」）を出力する. 出力したラベルと正解のラベルを比較して正解率を求める.

4. 実験結果

4.1 各方法の精度比較

BERT (前後 3 単語), Jaccard 係数，Cosine 類似度，および相互k-近傍グラフを用いる手法の実験結果を表 1 に示す．表 1 の結果を見ると，BERT の精度が最も高い結果となった．このことから同じ語義であるか判定するにはBERTを使用すると精度が高くなることがわかる. また, これを利用してグラフ構造を作成し，語義判別に使用することができれば，更なる語義曖昧性解消精度の向上が見込まれるのではないかと考えられる.

表 1　各方法における同義判定精度

BERT	Jaccard 係数	Cosine 類似度	相互 k-近傍
71.02	39.41	39.42	39.48

表 2　BERT に入力する前後の単語数を変えたときの同義判定精度

単語数	3 単語	5 単語	7 単語
精度	71.02	71.00	70.20

4.2 BERT に入力する前後の単語数の比較

BERT に入力する前後の単語数を 3 単語, 5 単語, 7 単語で行った実験結果を表 2 に示す．表 2 の結果から最も精度が高かったのは前後 3 単語となった．このことから，BERT で同じ語義であるか判定する際は前後の単語数が少ない方が効果的であると考えられる. また, 同義かどうか判定する際に必要となる単語が前後 3 単語中にあることが考えられ，それ以上の単語数となると不要な単語が増えて精度が下がると考えられる.

4.3 Jaccard 係数，Cosine 類似度の性能分析

Jaccard 係数と Cosine 類似度で極端に精度の悪い単語（例：14411「経済」）と精度の良い単語（例：545「上げる」）があったため，その精度の詳細を表 3 に示す．精度の悪くなった単語の 14411「経済」の分析を行うと, 語義が二種類のみであり, 片方の語義の用例文が極端に少ない場合, Jaccard 係数と Cosine 類似度は精度が悪くなるのではないかと考えられる．14411「経済」は「14411-0-0-1」「14411-0-0-2」の二つの語義しかないのと,「14411-0-0-2」は訓練データに二つ，テストデータに一つのみと偏っ

表 3　Jaccard 係数と Cosine 類似度での

Copyright(C) 2020 The Association for Natural Language Processing.
All Rights Reserved.

「上げる」と「経済」の同義判定精度

対象単語	Jaccard 係数	Cosine 類似度
545 上げる	80.52	80.36
14411 経済	7.76	7.68

ている．そのことから，この単語では，同じ語義を示す「1」の数は 2354，異なる語義を示す「0」の数は 146 となる．Jaccard 係数と Cosine 類似度は少なくとも 50 個に同じ語義である「1」をつないでいる．そのため，つないだ最低 50 個の「1」と 146 個の「0」を足して 196 前後の正解数となるので表 3 の実験結果のように精度が低くなったと考えられる．

また，Jaccard 係数と Cosine 類似度において最も精度の高かった「545」の語義は複数あり，同じ語義を示す「1」の数は 477，異なる語義を示す「0」の数は 2023 となり，異なる語義である「0」の数が多く，同じ語義である「1」が少ない．そのため，つないだ最低 50 個の「1」と 2023 個の「0」を足して 2073 前後の正解数となるので，Jaccard 係数と Cosine 類似度の精度が高くなったのではないかと考える．

5. おわりに

本稿では，対象単語を含む用例文ペアに対して，汎用的な言語モデルである BERT の学習済みモデルを用いて，同じ語義で使われているかどうかを判定する手法に有効性があるのかを分析した．BERT と Jaccard 係数と Cosine 類似度と相互k-近傍グラフを用いる手法の比較実験の結果，BERT による手法が最も高い精度となった．このことから，BERT による手法で用例文間の関連性をグラフ構造として適切に表現することが可能となると考えられる．そのため，対象単語前後の単語数を設定するだけで作成可能な，用例文間の関係を表現するグラフ構造の作成手法による Graph Convolutional Network (GCN)を用いた WSD 手法で高い効果を得られるのではないかと考えられる．今後は BERT を用いた同義判定によるグラフ構造を入力とした語義判別を行うことを課題とする．

参考文献

[1] 谷田部梨恵，佐々木稔，“グラフニューラルネットワークを用いた半教師あり語義曖昧性解消”，情報処理学会第 241 回自然言語処理研究会 (2019).

[2] 江里口瑛子，小林一郎，“グラフに基づく半教師あり学習のための潜在情報を考慮したグラフ構成”，人工知能学会全国大会論文集 3D4-2in (2013).

[3] 小嵜耕平，新保仁，小町守，松本裕治，“相互 k-近傍グラフを用いた半教師あり分類”，人工知能学会論文誌 Vol.28 No.4，pp.400-408 (2013).

[4] J. Devlin, M. Chang, K. Lee and K. Toutanova. “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.” NAACL-HLT, pp. 4171-4186 (2019).

[5] Okumura, M., Shirai, K., Komiya, K., Yokono, H. “Semeval-2010 task: Japanese WSD.，In: Proceedings of the SemEval-2010” ACL 2010, pp. 69-74 (2010).

Copyright(C) 2020 The Association for Natural Language Processing.
All Rights Reserved.