概要

単語が複数の意味 (語義) を持つことを単語の多義性という．機械が人間の言語を理解する上で単語の多義性を適切に取り扱うことは重要な問題である．単語の多義性に関する代表的な研究に語義曖昧性解消 (Word Sense Disambiguation; WSD) がある．WSD とは，ある文脈における単語の意味として適切なものを，あらかじめ定義された語義の集合の中から 1 つ選ぶタスクである．WSD は，自然言語処理分野において長年にわたり研究されてきた．近年の WSD の研究は，正解の語義が付与された用例集合を訓練データとして，ニューラルネットワークにより語義の分類モデルを教師あり学習することが主流である．ただし，従来の WSD の研究のほとんどは，訓練データに出現する語義の中からテストデータにおける単語の語義を選択することを想定している．しかし，語義は日々変化し，新しい語義も生まれている．そのため，単にあらかじめ定義された語義集合の中から該当する語義を選ぶのではなく，対象単語が訓練データに出現しない未知の語義 (新語義) であるのかを判定することが求められる．さらに，新語義に対し，人間の理解を助けるような何らかの手がかりを提示することも重要な課題である．例えば，新語義の上位概念を提示することで，人間がその新語義の大まかな意味を理解するのを助けることができる．

本研究は，語義の箱埋め込みを学習することを目的とする．語義の箱埋め込みとは，ベクトル空間における領域によって表される語義の抽象表現である．語義を箱埋め込みで表現することで，語義を単一のベクトルで表現していた従来の WSD の研究とは異なり，語義の概念的な大きさを表現できる．それにより，ある文脈における単語が新語義であるのかを判定したり，新語義の直接の上位概念 (上位語義) を推定することが可能となる．本論文では，既存の WSD の手法である MetricWSD を語義の箱埋め込みを学習するよう拡張する．さらに，モデルの損失を計算するために用いる小さいデータセットを作成する 2 通りの方法を提案する．提案手法を WSD，新語義の判定，新語義の上位語義の推定の 3 つのタスクに適用し，語義を単一のベクトルで表現する従来手法と実験的に比較する．

箱埋め込みは，箱の中心と辺の長さを表すベクトルの組から構成される．本研究で学習するモデルは，入力文の各単語に対し，その単語の文中における意味を表す箱埋め込みを出力する．以下，これを「文脈箱埋め込み」と呼ぶ．語義の箱埋め込みは，各語義が付与された単語の文脈箱埋め込みの平均により得られる．箱埋め込みの平均とは，箱の中心と辺の長さを表すベクトルのそれぞれについて平均ベクトルを計算することで求められる箱埋め込みである．

MetricWSD では，対象単語の文脈埋め込みを得るためのモデルとして Bidirectional Encoder Representations from Transformers (BERT) を用いていた．本研究では，対象単語の文脈箱埋め込みを得るために，BERT の最終層の後に全結合層をつなげたモデルを用いる．この層の入力は BERT の対象単語の最終層のベクトル表現であり，出力は文脈箱埋め込みである．全結合層の出力を半分に分割し，それぞれ箱の中心と辺の長さを表すベクトルとする．

語義の箱埋め込みを出力するモデルを学習するための損失関数は 2 つの語義の箱埋め込みの重なりから計算する．直感的には，ある語義の箱埋め込みは，自身もしくは上位概念の語義の箱埋め込みと重なるように，それ以外とは重ならないように学習される．損失を計算するためのデータセット (エピソード) を各語義ごとに用意し，各エピソードでは以下の手順でモデルを学習する．(1) 訓練データに出現する語義の集合の中から N_C 個の語義を選択する．(2) それぞれの語義について，その語義の用例をランダムに N_S 個選択し，サポートセットとする．(3) 対象語義について，サポートセットに含まれていない用例の中からランダムに N_Q 個の用例を選択し，クエリセットとする．(4) サポートセットに対するモデルの出力の平均から，各語義のプロトタイプ表現を得る．(5) クエリセットに対するモデルの出力とプロトタイプ表現を基に損失を計算する．(6) 損失を最小化するようにモデルのパラメータを更新する．

上記の手順 (1) において，N_C 個の語義を選択する 2 つの戦略を提案する．戦略 S_r では，まず対象語義とその上位語義を選び，残りをランダムに選ぶ．戦略 S_n では，対象語義とその上位語義に加えて下位語義と兄弟関係にある語義を選び，その後残りの語義をランダムに選ぶ．戦略 S_n では，ある語義の箱埋め込みが，特に下位語義や兄弟関係にある語義の箱埋め込みと重ならないことを重視している．

提案手法を WSD，新語義の判定，新語義の上位語義の推定の 3 つのタスクに適用し，その結果を評価する．実験に用いるデータセットとして，箱埋め込みを学習する語義の数による違いを調査するために，既存の WSD のデータセットから $\mathcal{D}_{\text{living\_thing.n.01}}$，$\mathcal{D}_{\text{artifact.n.01}}$，$\mathcal{D}_{\text{entity.n.01}}$ という 3 つのデータセットを作成する．living_thing.n.01，artifact.n.01，entity.n.01 は WordNet における語義 (synset) であり，各データセットはこれらの synset より下位の概念をもつ用例から構成される．$\mathcal{D}_{\text{entity.n.01}}$ は名詞全体であり，$\mathcal{D}_{\text{living\_thing.n.01}}$ と $\mathcal{D}_{\text{artifact.n.01}}$ は名詞のサブセットである．ベースラインとして，事前学習済みの BERT である BERT-NN と MetricWSD の 2 つを用意する．MetricWSD は，ベースモデルとして BERT を用いているため，BERT-NN をファインチューニングした手法といえる．

WSD の実験では，一部の名詞から構成されデータセットに出現する語義の数が少ないデータセット ($\mathcal{D}_{\text{living\_thing.n.01}}$ と $\mathcal{D}_{\text{artifact.n.01}}$) では，提案手法はベースラインを上回った．しかし，名詞全体から構成され語義の数が多いデータセット ($\mathcal{D}_{\text{entity.n.01}}$) では，提案手法はベースラインを下回った．これは，語義の数が多くなると，語義の箱埋め込みが必ずしも適切に学習されないことを示唆するものであった．新語義の判定の実験では，WSD の結果とは異なり，名詞全体から構成され語義の数が多いデータセットでは，提案手法はベースラインを上回った．新語義の上位語義推定の実験では，提案手法は全体的にベースラインを上回った．上位語義の候補を絞り込む際に用いる閾値は 0.5，0.7，0.9 のいずれかと設定したが，データセットや戦略によって最良の結果が得られる閾値の設定は異なることがわかった．以上の結果から，提案手法によって学習された語義の箱埋め込みは，実験条件によっては単一のベクトルで表現された語義の埋め込み表現よりも優れていること，ま

た語義の上位下位関係を適切に表現できる能力を持つことが確認された,