

論文 / 著書情報  
Article / Book Information

題目(和文)	単語埋め込み表現の語彙知識への意味適応
Title(English)	
著者(和文)	水木 栄
Author(English)	Sakae Mizuki
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第12629号, 授与年月日:2023年12月31日, 学位の種類:課程博士, 審査員:岡崎 直観,佐久間 淳,篠田 浩一,徳永 健伸,村田 剛志
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第12629号, Conferred date:2023/12/31, Degree Type:Course doctor, Examiner:,,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	審査の要旨
Type(English)	Exam Summary

## 論文審査の要旨及び審査員

報告番号	甲第		号	学位申請者氏名	水木 栄	
		氏名	職名		氏名	職名
論文審査 審査員	主査	岡崎 直観	教授	審査員	村田 剛志	教授
	審査員	佐久間 淳	教授			
		篠田 浩一	教授			
		徳永 健伸	教授			

## 論文審査の要旨 (2000 字程度)

本論文は、「単語埋め込み表現の語彙知識への意味適応」と題し、和文 5 章から構成されている。自然言語処理において、単語の意味の取り扱いが中核的課題の一つである。単語の意味を表現する方法論は、単語埋め込みに基づく方法論と人間の知識（語彙資源）に基づく方法論に大別される。前者の方法論は 2010 年頃からの深層学習に基づく自然言語処理で広く採用されているが、対義語や同位語、上位下位関係を表現しにくい。後者の方法論は単語の意味を辞書やシソーラス、オントロジーとして体系化するものであるが、網羅性が不十分である、単語から語義への対応付けを別途考える必要があるなど、実用上の課題が多い。本論文では、単語埋め込みと語彙資源を統合し、お互いの長所を活かすことを目指す。具体的には、深層学習を共通基盤とした手法を確立するため、語彙資源から得られる意味に関する知識を教師信号として、事前学習済みの単語埋め込みを変換または更新する「意味適応」に取り組んでいる。

第 1 章「序論」では、単語埋め込みや語彙資源など、単語の意味の取り扱いに関する研究背景を説明し、意味適応による両者の統合の狙いおよび既存手法の課題を述べている。その後、本論文で取り組むタスクとして、概念階層への適応による上位下位関係識別と、意味ネットワークへの適応による語義曖昧性解消の二つを挙げ、これらのタスクで意味適応に取り組むことの新規性および貢献を説明している。

第 2 章「関連研究」では、単語埋め込みおよび語彙資源について、それぞれの特徴、およびどのように意味を表現しているのかを詳述している。その後、意味適応によって単語埋め込みと語彙資源を統合する既存研究を説明している。また、上位下位関係識別および語義曖昧性解消タスクの学習や評価に用いられる言語資源を説明している。

第 3 章「概念階層への適応による上位下位関係識別」では、静的単語埋め込みを概念の階層構造に適応させる手法を提案している。具体的には、上位下位語ペアを教師信号として、単語埋め込みを階層コード表現に変換するモデルを学習する。階層コードは先頭桁の一致によって包含関係を定義可能な離散コードの体系であり、変換したコードの包含関係を調べることで、階層構造の特徴である推移性と反対称性に沿った上位下位関係の推論を狙う。しかし埋め込みからコードへの変換モデルを最適化するには、コード対の包含関係を微分可能な連続変数で計量する必要がある。そこで、本研究では連続緩和したコードを用い、包含関係が成立する事象の期待値を定義することにより、連続変数による計量を可能にした。提案手法で得られた単語のコード表現同士の包含関係を用いて上位下位関係を推論したところ、特に分類タスクにおいて既存手法の精度を上回った。これにより、概念階層への適応が上位下位関係識別に有効であることが実証された。

第 4 章「意味ネットワークへの適応による語義曖昧性解消」では、用例文内の単語および語義の語積文を表す埋め込み、すなわち対象語および語義の埋め込みを変換して、意味ネットワーク構造の特徴に沿って単語と語義、および語義同士の類似度を変更する手法を提案している。具体的には、隣接する語義対および語義・単語対は近付けるが、単語を介して結ばれる語義対および、つながりのない語義対は遠ざけるように埋め込みを変換して適応させるモデルを訓練する。ただし複数の語義と隣接する多義語は、単語が出現した文脈に最も近い語義を選んで近づける。提案手法の狙いは、ネットワーク上の繋がり方を埋め込みの類似度に反映させることで、意味的に関連する単語や語義が互いに近づき、関連しないものは遠ざかることである。訓練したモデルによって適応させた埋め込みを用いて、対象語に最も近い語義を選択する最近傍法で語義曖昧性解消タスクを解いたところ、語彙資源のみを用いる知識ベース手法における最高精度を更新した。これにより、意味ネットワークへの適応が語義曖昧性解消に有効であることが実証された。

第 5 章「結論」では、本論文のまとめと今後の展望を述べている。

本論文では、深層学習で獲得した単語埋め込みを人間が構築した語彙資源に適応させることで、計算機による単語の意味の取り扱いが改善するかという問いに対して、埋め込みを階層構造やネットワーク構造の特徴に適応させる二つの手法を提案し、その効果を実証した。本論文の成果は、単語の意味を計算機でどのように表現すべきかという学術的な問いだけでなく、意味検索や自然言語理解（対話においてユーザの意図を認識するタスク）などの応用に繋がることから、工学の発展にも寄与する。よって、本論文は博士（工学）の学位論文として十分価値あるものと認める。

注意：「論文審査の要旨及び審査員」は、東工大リサーチポジトリ(T2R2)にてインターネット公表されますので、公表可能な範囲の内容で作成してください。