【学位論文審査の要旨】

（論文審査の要旨）

自然言語処理における表現学習は，単語や文の文法や意味などに関する情報をデータからベクトルやモデルとして学習することである．単語の情報を表現することを単語分散表現，文の情報を表現することを言語表現と呼ぶ．生データは人手で情報を付与していないデータであり，容易に大規模なデータを収集できるため，表現学習の学習データとして用いられることが多い．深層学習が主流である昨今の自然言語処理において，表現学習は基盤的な役割を果たしており，盛んに研究されている．既存の表現学習の多くは単語や文の共起から文法や意味に関する有益な情報を学習している．一方で，生データは単語や文の共起が偏っており，このバイアスが文法や意味の表現学習に悪影響を与えることが知られている．そのため，本研究では表現学習における文法と意味に関するバイアスの問題に取り組む．

(1) 文法バイアス：表現学習に使われているコーパスのほとんどは母語話者により書かれており，文法的に誤ったテキストはほとんど含まれていない．そのため，コーパスで学習した単語分散表現や言語表現は文法的に誤った表現を考慮することができない．従って，文法的に誤ったテキストを扱うタスクにおいて最適な単語分散表現や言語表現になっていないため，問題である．文法誤りを含むテキストを扱うタスクとしては，入力文の文法的に誤った箇所を検出する文法誤り検出や入力文の文法的に誤った箇所を文法的に正しく書き換える文法誤り訂正などがある．これらは言語教育や言語学習の支援に用いることができる．

(2) 意味バイアス：表現学習では偏った単語の共起から，差別的な意味表現を学習してしまうことが知られている．例えば，コーパスでは単語 doctor は単語 she より単語 he と共起し，単語 nurse は単語 he より単語 she と共起する．共起を基に学習した単語分散表現では，共起する単語同士は類似度が高くなるように学習される．そのため，単語 doctor は単語 he と類似度が高く，単語 nurse は単語 she と類似度が高くなる．このように単語分散表現は意味バイアスから差別的な単語の関連性を学習することがある．そして，性差別的な情報を含んだ単語分散表現を用いることで，下流タスクにも性差別的な影響を与えることが知られている．これは公平性や倫理的な観点から問題である．

本研究では，まず，文法バイアスに関しては，単語分散表現の学習に擬似的な文法誤りと文法的に正しいテキストの両方を学習に用いる手法を提案する．そして，文法誤り検出により言語表現に事前学習された情報を保持しながら文法誤りを考慮する手法を提案する．文法誤り検出と文法誤り訂正の結果から，提案手法は単語分散表現と言語表現に文法誤りを効果的に考慮できることを示した．さらに，言語表現の文法情報を効果的に活用するために，言語表現の各層に対してアテンション（どのベクトルを重要視するかを学習する仕組み）を計算する手法を提案する．提案手法を用いることで文法誤り検

出と文法誤り訂正の 2 つのタスクにおいて世界最高精度を達成した.

次に，意味バイアスに関しては，自己符号化器と回帰モデルを用いて性差別バイアスを除去する手法を提案する.性別単語とステレオタイプ単語の単語分散表現間の類似度，事前学習された単語分散表現を用いた共参照解析モデルの性別単語とステレオタイプ単語を含む文の予測の均一さ，類似性やアナロジーに関するベンチマークデータセットの結果から，性差別情報を除去する提案手法は差別的ではない性別情報を保持しながら，既存手法と比較して最も性差別バイアスを除去できていることを示した.

以上のように，本論文で提案する手法は，テキストに存在する文法・意味バイアスを軽減することができ，あらゆる自然言語処理の基礎技術となることが期待されるため，情報科学において重要な意義があると考えられる．よって，本論文は博士（情報科学）の学位を授与するに十分な価値があるものと認められる.

（最終試験又は試験の結果）

本学の学位規則に従い，最終試験を行った．公開の席上（オンライン）で論文発表を行い，学内外の教員による質疑応答を行った．また，論文審査委員により本論文及び関連分野に関する試問を行った．これらの結果を総合的に判断した結果，専門科目についても十分な学力があるものと認め，合格と判定した.