論 文 内 容 の 要 旨

博士論文題目 Semantically Readable Distributed Representation Learning and Its Expandability Using a Word Semantic Vector Dictionary

氏　　名　芥子 育雄

要　　旨

あらゆる人間活動が非構造化テキストとして蓄積されようとしている．これらビッグデータから価値を取り出すことが，ニューラルネットワークを用いた分散表現学習に期待されている．課題は，Twitter などの短文テキストは単語のスパース性があること，およびニューラルネットワークにより抽出された特徴量の理解が困難なことである．単語のスパース性とは，解析対象テキスト中の単語の多くが，非構造化テキストから分散表現を学習した語彙に含まれないことを指す．

ニューラルネットワークにより獲得される分散表現に対して，本論文では，辞書編纂の専門家が構築した単語意味ベクトル辞書を導入する．本辞書では，単語は 266 種類の概念分類から成る特徴単語との関係（関係あり，なし）により意味を表現する．専門家が特徴単語との関係を付与した単語を基本単語と呼ぶ．

最初に本辞書を利用して，百科事典から抽出された 10 万語以上の単語意味ベクトルを半自動的に学習する方法を提案する．まず百科事典の頻度の高い 3700 語を基本単語として，専門家が 266 種類の特徴単語との関係を付与した．次に百科事典を用いたブートストラップアルゴリズムにより，10 万語以上の単語意味ベクトルを学習した．この単語意味ベクトルを用いて，画像に付与されたキーワードに対して，百科事典の知識を利用した画像の連想検索を実現し，有効性を検証した．

本論文で提案する日本語 Twitter を対象とした評判情報抽出のための極性分析ベンチマークは存在しないため，単一製品を対象とした小規模なベンチマークを 2 種類，および 8 カテゴリーの製品などを対象とした多様性のあるベンチマークを 1 種類構築した．これらは共通ベンチマークとして公開を予定している．

次に Twitter における単語のスパース性を解消することを目的に本辞書を用いてツイート中の基本単語を特徴単語に拡張し，ニューラルネットワークで学習する統合化手法を提案する．本辞書では，専門家が構築した約 2 万語の基本単語を用いる．本統合化手法を用いて，Twitter からの評判情報抽出を実現した．提案手

法は単語のスパース性を解消し，極性分析の精度が向上することを示した．

我々は更にニューラルネットワークの隠れ層ノードに特定の意味（特徴単語）を割り当て，本辞書を元に約2万語の基本単語に初期重みを設定することにより，学習後も重みの大きな隠れ層ノードの意味が維持されることを示す．ユーザテストの結果，教師なし学習によるツイートの分散表現に対して，重みが大きな上位5ノードの52.4%がツイートと関連があることが示された．また，分散表現学習を前提に辞書を改良し特徴単語を264種類とした．前述の多様性のあるベンチマーク，および本辞書を用いて提案手法の拡張性評価を行った．さらにWikipediaコーパスを用いて，提案手法のドメイン独立性評価を行った．これら客観的，主観的評価により，隠れ層ノードは特定の意味を維持することを確認し，提案手法は分散表現学習の可読性を改善することを示した．

氏　名	芥子　育雄

（論文審査結果の要旨）

大量のテキストデータから価値を取り出すために，ニューラルネットワークにより獲得される分散表現が用いられている．本論文では，分散表現を可視化し，人間に解釈可能，さらに，人間により発展させることが可能な処理法の実現を目指している．そのため，まず，辞書編纂の専門家が構築した単語意味ベクトル辞書を導入する．この本辞書では，それぞれの単語は 266 種類の概念分類から成る特徴単語との関係（関係あり，なし）により意味が表現される．この辞書をベースに約 2 万語（当初 3700）語の基本単語を百科事典や新聞記事から抽出し，それをベースに分散表現を学習させる．これを画像の連想検索，さらには，Twitter の評判情報解析に適用し，性能の向上ができることを明らかにした．

さらに，ニューラルネットワークの隠れ層に特徴単語を割り当て，約 2 万語の基本単語の初期重みを設定し，学習することにより学習後も重みの大きいノードの意味が維持されることを示した．これら客観的，主観的評価により，隠れ層ノードは特定の意味を維持することを確認し，提案手法は分散表現学習の可読性を改善することを示した．

これらの成果は，従来技術では本質的に解決困難であった問題に対して，解決策を示しており，本研究に関し３編の学術論文、１編の査読付き国際会議論文として発表しており，これまでの種々の研究業績を総括すると，10 編の学術論文，7 編の査読付き国際会議論文，43 編の国内外の査読無し発表を行っていることから、非常に高く評価できる．以上より，平成２９年７月２０日に開催した公聴会の結果も参考にして，本博士論文の審査を行い，本論文は，博士論文（工学）として十分な価値があるものと判断した．