

審査の結果の要旨

氏名 丸井 淳己

本論文は、商品や人物のようにウェブ上に大規模なテキストデータとして表出している実世界のエンティティ情報について、それらを効率的に分析し、多様なアプリケーションにおいて利活用可能にするための新たな分散表現の学習手法を提案している。特に、エンティティに付与されているカテゴリ情報に着目し、従来の分散表現モデルにカテゴリを組み込んだ新たなフレームワークを提案し、実際の大規模な EC（電子商取引）サイトデータならびにソーシャルメディアデータに適用して実験を行い、その有効性を示している。

第 1 章では、研究の背景、目的等について述べている。ウェブの普及に伴い、大規模 EC サイトにおける商品情報やソーシャルメディアにおける人物情報等、実世界のエンティティが大規模なテキストデータとしてウェブ上に急激に増加していることをまず述べている。そして、それらの大規模データに対して、効率的にエンティティの情報を表現し、検索や推薦等の用途で利活用するための基盤技術の必要性を背景として述べている。大規模なエンティティ情報を計算機上で適切に表現し、処理するための手法として、従来、分散表現が多く用いられている。テキスト中の文脈上における表層言語情報に基づいてエンティティの潜在表現を学習するという一般的な分散表現に対して、大規模なウェブ情報を構造化し、エンティティの属性や関係性等の知識構造を利用して分散表現学習を行う研究も進められているが、特に大規模なエンティティの分散表現学習においては、これら表層的な言語情報と構造的な知識を適切に統合し、検索や推薦等のさまざまな用途に利活用が可能な柔軟かつ頑健な分散表現モデルが必要となる。本論文では、カテゴリ情報という、人が対象を認識する際の構造的知識に着目し、それらをエンティティの情報を表現するための分散表現の学習手法に組み込んだ新たなフレームワークを提案している。

第 2 章では、本論文の関連研究について述べている。まず、テキスト処理における単語や文書の従来の表現手法として、**Bag-of-words** モデルがある。近年では、大規模なテキストデータと計算機資源の高度化に伴い、テキストのより潜在的な表現を獲得する手法として分散表現に関する研究が盛んになっている。特

に、深層学習に基づく分散表現の学習手法である **Word2Vec** やその拡張としての段落ベクトルモデルは、高速かつ高精度な手法として着目されている。

第 3 章では、本論文の提案手法について述べている。まず、本論文では、人が対象を認識しながら適切に分割したエンティティの集合をカテゴリと定義し、従来の段落ベクトルモデルを拡張し、エンティティに紐づくテキストの文脈からエンティティの分散表現を学習するエンティティベクトルモデル、ならびにエンティティに付与されたカテゴリも文脈として分散表現を学習するカテゴリベクトルモデルを提案している。カテゴリベクトルモデルの特徴はカテゴリに基づくモデルの柔軟性と拡張性であり、異なるカテゴリを組み合わせた結合モデルや複数の言語のカテゴリを統合した多言語モデルの構築が可能となる。

第 4 章では、提案手法であるカテゴリを用いた分散表現学習手法を大規模な **EC** サイトのデータに適用し、その有効性を評価している。具体的には、日米の楽天 **EC** サイトの数百万の商品とそれらの数万のカテゴリ情報を学習データに用いて商品の分散表現を学習した上で、テストデータの数万の商品のカテゴリをそれらの説明文から推定するタスクを行った。その結果、日本語、英語の両方のデータについて従来手法を上回る精度で商品のカテゴリを推定できることを示している。

第 5 章では、提案手法を大規模なソーシャルメディアのデータに適用し、その有効性を評価している。具体的には、**Twitter** から抽出したコミュニティをユーザの属するカテゴリと見なし、学習データの数百万のユーザのつぶやき内容からユーザの分散表現を学習した上で、テストデータのユーザのコミュニティを推定するタスクを行った。その結果、従来手法を上回る精度で **Twitter** ユーザのコミュニティを推定できることを示している。さらに、第 6 章においては、提案手法で獲得されたコミュニティについて、その分散表現の表す文脈の性質を定性的、定量的に論じている。

第 7 章では、議論として、提案手法の限界や応用について述べている。特に、提案手法の分散表現学習における収束性、最適なパラメータの選択、そして他のアプローチによる分散表現モデルとの差異等を通して、提案手法の限界や複数の応用の可能性を論じている。また、本論文の学術的貢献として、大規模なテキストデータを処理するための情報工学分野の方法論、技術経営学分野における大規模データ分析に基づいた戦略的な意思決定支援への応用、という 2 つの視点から提案手法の意義を述べている。

第 8 章では、結論を述べている。本論文は、現在爆発的に増加しているウェブ上の大規模なテキストデータ分析と利活用の基盤技術として、特に商品や人物のような実世界のエンティティ情報について、カテゴリ情報という人が対象を認識する際の構造的知識に着目し、大規模で多種多様なエンティティ情報を柔

軟かつ頑健に処理可能な新たな分散表現の学習手法を提案した。また、本論文では、実際の大規模データを用いた実験を通して、提案手法の有効性を検証し、さらに多面的な観点から応用の可能性を論じたとしている。

以上のことから、本論文は、情報工学分野ならび技術経営学分野の両分野の研究として独創性、新規性が高いものであり、先端の人工知能技術とビッグデータ分析を利活用した効率的かつ戦略的な意思決定の判断が求められる技術経営の現場に対し、基礎と応用の両面で有用な知見や手法を提供するものであると評価出来る。

よって本論文は博士（工学）の学位請求論文として合格と認められる。