

審査の結果の要旨

氏名 野中 尚輝

この学位請求論文「ソーシャルメディアからの人気情報を用いた商品の販売予測に関する研究」では、ソーシャルメディアを使って、マンガやアニメなどのコンテンツ作品、また自動車などの耐久消費財の販売を予測する手法を提案し、その有効性を検証している。従来研究と比較し、データ数が少ない場合に対するさまざまな工夫を行い、複数の提案手法それぞれに対して評価実験を行い、得られた知見をまとめている。

近年、インターネットやスマートフォンの普及とともに、Twitter や SNS, Wikipedia などのソーシャルメディアが幅広く利用されている。ソーシャルメディアを用いた社会動向の予測はこれまでもさまざまな研究が行われており、例えば、代表的なものとして、映画の人気予測や選挙結果の予測などが挙げられる。しかし、こういった研究において予測のために用いる素性はドメインごとにヒューリスティックに決められることが多い。また、データが少ない場合には、素性の数を絞るために、人間がドメインの知識を用いて有効な素性を絞る必要があった。

データが少ない場合に対応するためにはいくつかのアプローチが考えられる。本研究では、いくつかのアプローチによる手法を提案している。まず、3章では、ソーシャルメディアから予測したい問題と別の問題を設定し、その問題において有用な素性をもともと予測したい問題に適用する手法を提案した。これは転移学習のひとつと考えることができる。レビューサイトに投稿されたレビュー文章に対して、感情分析という問題を設定し、このタスクに対して寄与の大きい単語素性を選択した。そして、これをもともとの対象問題であった自動車の販売数予測タスクに適用し、転移した素性により予測精度が向上することを示した。

4章では、Wikipedia のデータを用い、ユーザの行動履歴の系列データからコンテンツ作品に関する嗜好を抽出し、コンテンツ作品の人気予測を行う手法を提案した。Wikipedia の編集履歴は誰がどのコンテンツを編集したかがデータとして蓄積されており、これを系列データと捉えることで、近年幅広く用い

られているニューラルネットワークによるベクトル表現である **word2vec** を用いることができる。編集履歴からコンテンツ作品のベクトル表現を得て、これを用いてコンテンツ作品の人気予測を行った。コンテンツ作品の人気を代替する指標として、**Wikipedia** の被リンク数を用いた。結果として、編集履歴から構築したコンテンツのベクトル表現を用いることで予測精度が向上することを示した。

5章では、**Wikiedia** によるコンテンツ作品の人気予測を、地域をまたがって拡張する。**Wikipedia** のデータは、異なる地域でデータ量に差がある。データが少ない地域における予測問題を、データが多い地域における予測問題で補うことができれば、さまざまな地域での予測の精度を上げることが可能になる。提案手法では、データが多い地域のデータを用いてモデルの事前学習を行い、少ない地域のデータからの予測する手法を構築した。評価実験によって、予測精度が平均で10%程度向上することを示した。

6章は、3章で扱った自動車の販売予測と、4、5章の **Wikipedia** からの予測を統合するという位置づけであり、**Wikipedia** のデータを用いて自動車の販売予測を行う。閲覧数を予測するという別の問題を設定し事前学習を行うことで、コンテンツ作品の販売予測の精度が向上することを示した。また、コンテンツ作品に関するベクトル表現を工夫することで、販売予測の精度がさらに向上することを示した。

本研究は、従来から行われているソーシャルメディアからの社会動向の予測を拡張し、深層学習や転移学習に関連する新しい技術を取り入れ、その有効性を示した。データが少ない問題に対して転移学習、事前学習を行うことの有効性を示した。また、**Wikipedia** の閲覧データ等の効果的な活用法を提示しており、さまざまな商品に適用可能な手法の可能性を示した。本論文で得られた知見は、ソーシャルメディアの活用がますます進む上で、学術コミュニティにとって重要な知見を提供している。同時に、産業界においても、本論文の手法は商品の販売予測等のさまざまな活用の用途がある。以上、ソーシャルメディアの活用に関して、産業上と学術上の両面から必要性の高い課題とその解法を設定しており、その意義は高く評価できる。

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。