

## 論文の内容の要旨

論文題目 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

氏 名 野中 尚輝

商品の販売予測は、企業が意思決定を行う際に重要となる。近年、消費者は自身の嗜好にそった商品を購入するようになっており、消費が多様化している。そのため、商品の販売予測を行う場合に、個人の嗜好を考慮することの重要性が増している。また人気となる商品は、その時々流行に応じて変化するため、継続的に調査を行うことが重要となる。しかしながら、多様な個人の嗜好を捉えることや継続的な調査を行うことは、アンケート調査といった従来の手法を用いる場合、コストの面で問題があった。

一方で近年、インターネットやスマートフォンの普及が進んだことから、ソーシャルメディアが幅広く利用されるようになった。ソーシャルメディアの利用者数は全世界において拡大が続いており、オンラインレビューやオンライン百科事典、ソーシャルネットワークサービスなどさまざまな形で存在している。ソーシャルメディアの普及により、これまで情報を受け取る側であった消費者が、自身の考えや商品に対する感想といった情報を発信することが可能になった。各ユーザによりソーシャルメディアに投稿される内容は、各ユーザの嗜好を反映しており、これらを分析することで消費者の嗜好を捉えることができる可能性がある。また、ソーシャルメディアへの投稿は継続的に行われるため、人気の変化も捉えることが可能であると考えられる。

このような背景から、ソーシャルメディアを用いた商品の販売予測の研究が行われてきた。特に、多くの人々が鑑賞し、個人の嗜好により対象への評価が異なることからソーシャルメディアにおいて言及されることの多い、映画が予測の対象として扱われてきた。ソーシャルメディアによる映画の予測としては、Twitter のデータを用いる研究やプロ

グから得た情報を用いた研究, レビューの情報を用いた研究が知られている. 映画以外のドメインにおけるウェブ上のデータを用いた販売予測の研究としては, 入力される検索クエリの回数から自動車の販売数をはじめとする経済動向を推測した研究や検索クエリを入力回数, Twitter における言及数, Wikipedia のリンク構造を用いてアニメ, 漫画といったマルチメディアタイトルの人気予測を行っている研究が存在する. このように映画を中心として, 個人の嗜好が対象の評価に反映されるためソーシャルメディアに情報が上がりやすい傾向にある商品の予測が行われてきた. これらの先行研究において, ソーシャルメディアから得られた販売予測の際に用いられる特徴量には, 対象となる商品が言及された回数や対象の商品についてのレビューの投稿回数といった指標が用いられてきた. これらの言及された回数やレビューの投稿回数といった特徴量は, 予測を行う側のもつ外部知識に基づいて, ヒューリスティックに選択されている.

しかし, ヒューリスティックに特徴量を定めることが難しい場合がある. 例えば, 入力データが時系列の変動データや対象の商品名が出現しないような文章であるような場合, 用いるべき特徴量をヒューリスティックに定めることは難しい. ヒューリスティックに特徴量を選択できない場合, 得られる全特徴量を直接予測に用いることも考えられるが, 予測対象のデータの量が少ない場合, 過学習が起りやすくなり予測が難しくなる. このような背景のもと本研究では, ラベルデータの数が少なく, ソーシャルメディアから直接特徴量を獲得することが難しい場合における, ソーシャルメディアによる商品の販売予測の精度を向上させることを目的とした手法を提案した. 具体的には, まずソーシャルメディアから得られた素性を用いて, 商品の販売数に関連する指標を予測する問題を設定したその後, 設定した問題を解く上で有用な素性および特徴量を人気情報として保持し, 得られた人気情報をラベルデータの数が少ない商品の販売数の予測問題において用いる手法を提案した.

本研究では, 商品の販売予測の中でも, ソーシャルメディアが予測において有用であると考えられる商品を対象として販売予測を行った. ソーシャルメディアが予測において有用であると考えられる商品は, 大量生産されるため潜在的には誰もが入手可能であり, 各ユーザの嗜好が商品の評価に影響するためソーシャルメディア上に各ユーザの意見が投稿されやすく, また流行が存在するため継続的に調査を行う必要のある商品である. 本研究では, 予測においてソーシャルメディアが有用であると考えられる商品から, 予測対象として重要と考えられる自動車およびマルチメディアタイトルを対象として検証実験を行なった.

3章では, 日本の輸出総額に占める割合でトップの商品である自動車の販売予測を行なった. 実験では, 記述される内容に主観的な口コミ情報が含まれており, 記述内容に対応する商品が明確であるという理由からオンラインレビューのデータを用いた. まず, 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を行い, 感情予測タスクにおいて予測への寄与度の大きい単語素性を特定し, 人気情報とした. 続いて, 人気情報

とした単語素性を自動車の販売数予測タスクに用い、単語素性をを用いない場合、単語素性をランダムに選択した場合、レビュー文中に含まれる単語素性全てを用いた場合および既存の極性辞書をもとに単語を選択した場合の予測精度と比較を行なった。実験結果から、感情予測タスクにて予測に有用であった単語素性を用いることにより、予測精度が向上することを確認した。

次に、4章、5章および6章にて、低価格であることが多く、商品の販売予測を行う上で対象とした商品に対する注目の度合いが重要になると考えられるマルチメディアタイトルについての販売予測を行なった。実験では、マルチメディアタイトルの網羅性が高いという点から Wikipedia のデータを用いた。また、販売数のデータとして、公開される頻度が高く、複数のカテゴリの情報を収集できるという理由からオンラインショッピングサービスのデータを用いた。オンラインショッピングサービスにて公開される販売に関するデータは、販売ランキング上位のデータのみである。そのため入手できるラベルデータは、存在するページ数と同数のデータが得られる Wikipedia のラベルデータと比較して入手できるデータの量が少ない。また、オンラインショッピングサービスの販売データは、ユーザの個人の購買行動と紐づいていないため、ユーザの嗜好を反映できない。そこで本研究では、4章、Wikipedia の編集履歴からユーザの嗜好に関するベクトル表現を抽出する手法について検証し、5章において、データ数が少ない場合における予測精度を向上させる手法を提案した。続いて、4章および5章で得られた知見に基づいて、6章にて販売に関連すると考えられる Wikipedia のページ閲覧数を対象とした予測問題を解き、得られた人気情報を用いて販売予測を行なった。

最後に、上記の研究に基づいた提案手法の限界と応用可能性について考察し、結論を述べる。