

令和 3 年 6 月 14 日現在

機関番号：35302

研究種目：基盤研究(C)（一般）

研究期間：2017～2020

課題番号：17K04029

研究課題名（和文）レビューデータを用いたマーケット・セグメンテーション手法の開発

研究課題名（英文）Development of Market Segmentation Method using Review Data

研究代表者

水谷 直樹（Mizutani, Naoki）

岡山理科大学・経営学部・准教授

研究者番号：30330533

交付決定額（研究期間全体）：（直接経費） 1,500,000円

研究成果の概要（和文）：商品やサービスに関して消費者から発信されるレビュー情報を収集し、収集した大規模データからマーケティングに有用なセグメンテーション情報を獲得する手法を提案した。提案した手法は、消費者からのレビュー情報から得られる商品間の関係性をもとに、ネットワーク分析の手法を用いて、商品および消費者をグループ分類するものである。また、時間とともに変動するセグメンテーションの規模に関して、規模の変動を表現する数理モデルを提案し、実データから数理モデルのパラメータを推定する方法を確立した。

研究成果の学術的意義や社会的意義

マーケット・セグメンテーション手法の確立と、セグメンテーションの時系列的変化を表現する数理モデルの導出によって、マーケットの状況把握が容易に実施できるようになるとともに、効果的なマーケティング戦略を適時に実施可能になると考えられる。その際に必要となるデータは、商品やサービスに関して消費者から発信されたレビュー情報やSNSへの投稿情報であり、容易に入手可能で情報入手のためのコストは安価である。アンケートよりも偏りが少ないことなど、企業におけるマーケティング活動において有用な方法である。

研究成果の概要（英文）：We proposed a method of acquiring market segmentation useful for marketing activities, analyzing review-information of products and services posted on web-sites from consumers. The proposed method classifies products and consumers into groups using a network analysis method based on the relationships between products obtained from review information from consumers. We also proposed a mathematical model that expresses the segmentation scale with time fluctuation. And we established a method for estimating the parameters of the mathematical model from actual data.

研究分野：商学

キーワード：マーケット・セグメンテーション レビュー・データ ネットワーク分析 最適化

1. 研究開始当初の背景

マーケティング活動を効果的に展開するために、消費者を何らかの意味で同質的な消費者グループに分けて、そのいずれかにマーケティング・ターゲットを絞ることが行われる。このセグメンテーションを行なう方法としては、デモグラフィック特性など事前に決めた基準を適用してセグメント分けを行なう方法と、多次元的な変数を基準としてサンプルを集めることによってセグメントを構成する方法との大きく2つに分けられる[1],[2]。

前者の方法では、消費者の表面的な属性をもとにしてセグメンテーションを行なうため、ライバル企業との差別化を行なう上で有利な方法と言えない[3]。また、後者の方法においては主としてアンケートの実施によってサンプルを集めるが、セグメンテーションを行なう基準をあらかじめ選定できていないとアンケートを設計できないので、この方法も事前に基準を決める必要がある。そして、アンケートの設計は勘や経験に頼らざるを得ないのが実情で、その良し悪しがセグメンテーションの結果に大きく影響してしまう。アンケート以外の方法としては、情報技術の進展によって、IDつき POS データを用いる研究が進められているが、POS データから得られる情報には限りがあり、顧客の好みを推測するには情報が少なすぎる。

本研究は、上記の欠点を補うために、Web 上に投稿された商品・サービスに対する膨大なレビュー情報を消費者情報として用いることが特徴である。Web 上に投稿されたレビュー情報を活用する研究[4][5]は、主として工学分野で行われているが、機械学習等の人工知能要素技術の開拓を志向した研究である。これらの研究成果が経営学分野へ寄与するには至っておらず、経営学分野からのアプローチが必要であり、大きな課題となっている。

2. 研究の目的

本研究の目的は、Web 上に投稿されたレビュー情報を使って、消費者の嗜好の多様性を反映した精度の高いマーケット・セグメンテーション手法を確立することである。本研究の特徴は、セグメンテーションのためのデータとして、消費者が Web 上に投稿した商品・サービスに対する膨大なレビュー情報を活用することである。レビュー情報には信頼性の低い情報が混入しているので、まずレビューデータの信頼性の検証手法を開発する。次に、レビューデータのうち、商品・サービスに対する総合評定値および各種属性についての評定値といった数値データに対して統計的手法を適用し、コメントとして記述されたテキスト文に対して機械学習の手法を適用して、それらを統合したマーケット・セグメンテーション手法を提案・構築し、手法の評価を行なう。

これまでに、映画のレビューデータを対象に統計的にレビューデータを分析して、その特徴を精査してきた。さらに、消費者モデルを推定する研究を行ってきたが、これはマーケット・セグメンテーション手法を実現するための重要な要素となるものである。これら研究では、個人 ID 情報を伴うレビューデータを用いてきたが、積極的に消費者それぞれの嗜好を分析してこなかった。その理由としては、有意な分析を行なうには 1 個人あたりのレビューデータ数を数十件以上確保する必要があり、大量のレビューデータが必要になるためである。

レビューデータとして Web に蓄積されているデータは、映画、耐久消費財などのカテゴリごとに百万件以上に及ぶことを確認している。それらの大規模データを入手できれば、これまでの研究成果を適用し、消費者ごとの分析を行なうことにより、レビューデータによるマーケット・セグメンテーションが実現でき、これまでにないマーケット・セグメンテーション手法となる。

本研究の学術的な特色および独創的な点は、事例紹介や抽象的議論によって展開されることが多かった本課題に対し、大規模データを用いて実用レベルで実現させて評価するところにある。大規模データの利用は近年になって盛んになった新しい分野であり、経営学の分野では未だに活用事例が少ない。したがって本研究は、経営学分野での大規模データ利用として、アプローチ手法の独創性が高い。また、経営学においても企業の情報通信技術活用に対応して学問領域を拡充し、経営学分野の学際的展開を進めていく必要があるが、本研究は情報学および工学的な知識や技術を導入して経営戦略を考察することになり、経営学の学際的発展に貢献すると考えられる。

市場の変化は年々スピードを増しており、これまでに提案されていた手法では顧客の嗜好変化の把握が追いつかない場合も出てきている。本研究は次々と投稿されるレビューデータをもとに市場を把握するものであり、スピードの速い市場変化に適応した手法として有用性に富むものである。また、現実に存在するデータを利用して成果を示すため、信頼性の高い研究内容になるとともに、応用価値の高い研究成果を得ることができる。

3. 研究の方法

本研究では、以下に示す第1課題～第3課題を設定し、それぞれを段階的に推進することによって目標を達成した。また、その前段階として第0課題が加わる。

第0課題: レビューデータの収集および前分析

投稿者個人の嗜好を分析し分類するには、相当数のレビューデータが必要となる。これまでの収集実績から、収集には多大な時間が必要となることがわかっている。それをふまえ、投稿を100件以上行った投稿者が1000名以上になるところまでレビューデータを順次収集する予定であり、レビュー件数にして100万件の収集が必要と予測している。ただし、第2課題のセグメンテーションが良好に実行可能と判断できた段階で収集作業を打ち切る。

第1課題: レビューデータの信頼性検証

これまでの研究実績において、いかなるレビュー対象に対しても同一の評点をつける投稿者の存在や採点基準が投稿者によって大きく異なることを確認している。このように、精度の高いマーケット・セグメンテーションを得るには不適当なレビューデータが混入しているため、それらのデータを発見して、分析対象から外す、あるいは値を補正することが肝要である。これまでの研究成果や最新の研究動向をフォローしながら、フィルタリング手法の提案を行なう。

第2課題: レビューデータからのマーケティング・セグメンテーション手法の提案

信頼性の高い大量の投稿者IDつきレビューデータに対して、レビュー中の評定値など数値データに対しては統計的手法を適用し、テキスト文に対しては機械学習的方法を適用して、セグメントを導出するアルゴリズム開発を行なう。それら手法を試行し評価しながら手法を具体化する。

第3課題: 実データにもとづく提案手法の評価

本研究で開発したマーケット・セグメンテーション手法の有用性を検証するために、過去のレビューデータを対象に本手法を適用して評価する。ビジネスの現場では、次々と新たなセグメンテーションが発見されて新カテゴリーの商品が創出されている。提案手法によって、それら商品の対象セグメントが予測できるか確認することを通じて、手法の有効性検証とする。

4. 研究成果

(1) 消費者から発信された情報の収集によるデータベースの構築

本研究では、商品やサービスに関して消費者から発信されたレビュー情報を収集し、収集した大規模データからマーケティングに有用な各種の情報を得ることを目指す。商品やサービスに関して消費者がソーシャル・ネットワーク・サービス(SNS)等に発信する情報は有益であり、活用されるべきであるが、有効に活用されることは少なかった。一方、消費者が、SNSからの情報によって購買行動に大きな影響を受けていることが指摘されており、消費者から発信される情報に容易にアクセスする手段をもつことが、マーケティング分野の関係者にとって重要な時代となっている。本研究は、そのデータ収集から分析の方法論を明確にすることが目的である。まず、消費者から発信された情報を大量に収集して大規模データとして蓄積する必要があり、研究の初年度は、それら情報の収集方法の確立と、データ収集に充てた。

データの収集先として、映画のレビュー情報、グルメサイトの店舗レビュー情報、価格比較サイトでの商品レビュー情報、CtoC物販サイトの商品取引情報、一般のSNSに投稿された記事などを対象に、投稿された情報を網羅的に収集する方策を検討した。まず、レビュー情報をまとめるサイトが数多く存在する。それらサイトでは、商品や店舗などの評価を等級づけしたレーティングに関するデータを入手できる。評価の情報を数値として得るには、レビュー情報を用いることが有効である。一方、一般のSNSサイトへは、日々多数の投稿が行われており、特定の商品や店舗に関する投稿について、投稿件数で見るとレビューサイトよりも遥かに多くの投稿が行われており、有力な情報収集媒体である。特に、商品等の話題性の変化を早期に計測する手段として、一般のSNSは非常に有力と考える。

本研究の初年度は、それらレビューサイト、SNSサイトからのデータ収集を行った。分析以降の方法が確立していない段階であり、また、多くの種類の入手先からデータを収集して分析試行を行って比較したかったため、収集方法として、Webスクレイピングを主として用いた。そのため、サイトごとに収集プログラムを作成し、投稿されたレビューデータ、SNS投稿記事を収集し、データベース化を行った。収集にかかった期間は1年半ほどで、消費者から発信された情報として、延べ数百万件の情報が収集できた。

(2) レビューデータの信頼性を高めるためのデータ・フィルタリング

レビューデータの投稿においては、いわゆるサクラ投稿を行う投稿者の存在やプログラムによって多数のレビュー投稿を行う自動投稿BOTの存在があり、これらによってレビューデータの件数には歪みが存在する。これらの行為は、1つの投稿者IDを使って多数の投稿が行われるか、投稿者IDを頻繁に変えながら行われることが多いと考えられる。そこで、投稿者IDが入手

できるレビューデータの場合、上記の特性を考慮して、好ましくないレビューデータを判別して調査対象としない方策について考察した。例えば、投稿件数が1件だけの投稿者のデータの信頼性は低い。また、大量に投稿している投稿者IDについて、同一内容の投稿が多いならば、その投稿者について、投稿内容の信頼性は低いと考え、レビューデータの信頼性を高めるためにデータを取捨選別した。

(3) レビューデータからのマーケティング・セグメンテーション手法の提案 [6]

大量の投稿者IDつきレビューデータを用いて、レビュー中の評価値など数値データに統計的手法を適用し、マーケティング・セグメントを導出するアルゴリズム開発を行なった。個人の好みのグルーピング化をレビューサイトに投稿されたレビューデータによって行う方法の提案である。レビューサイトによっては、個人をIDレベルで特定できるサイトがある。投稿者IDをもとにすれば、各個人がどの商品を好ましく思っているかを抽出することが可能となる。そこで、個人の商品選択履歴に関する単純データをもとに、関連性の高い商品間関係を抽出するとともに、商品をグループ分類する方法を提案した。特定の個人について、商品Aと商品Bを好ましく思うといった情報から、それらの関係をネットワーク表現し、ネットワーク分析の手法を導入することによって、商品をグループ分類、消費者をグループ分類する方法である。

研究対象としたのは、2016~2018年に公開された映画タイトルに対して投稿されたレビューデータである。映画レビューを投稿できるレビューサイトの中で、投稿者IDが特定できるYahoo!映画 (<https://movies.yahoo.co.jp/>) から収集したレビューデータを用いて検証した。Yahoo!映画のレビュー投稿ページでは、投稿者が固有のIDを利用してレビュー投稿を行うため、どの投稿者がどの映画タイトルに対してレビューを投稿したか分析可能である。したがって、個人が自発的に選択した映画タイトルの組み合わせを抽出することができる。

ネットワーク分析は、対象間相互の関連性を分析する有力な手段であるが、まずそれぞれの対象間に存在する関連性指標を決めてネットワークとして表現する必要がある。個人に選ばれやすい2本の映画タイトルの組み合わせを抽出し、それらの間に関連性が存在すると決める方法を採用した。また、ネットワーク分析方法としては、モジュラリティ指標を用いたコミュニティ分析を導入した。

異なる2本の映画タイトルの組み合わせが、何人の個人に選ばれたかという件数をもとに、選ばれやすい組み合わせを抽出する方法について考察するが、人気のある映画タイトル、人気のない映画タイトルが混在するため、映画タイトルの組み合わせが選ばれた件数の多寡だけで選ばれやすい組み合わせを抽出することはできない。そのため、まず投稿件数が1件だけの投稿者のデータから、映画タイトルが単独に選ばれる割合を得て、個人にとって映画タイトルの間に関連性なく選ばれると想定した場合の特定の2本の映画タイトルの組み合わせが選ばれる確率理論値を計算する。そして、映画タイトルの組み合わせとして実際に選ばれた割合が上記理論値を大きく上回った場合に、選ばれやすい組み合わせとして判定することにした。さらなる工夫として、投稿件数が2件だけの投稿者のレビューデータを用いて、映画タイトルの間に関連なく非復元抽出によって映画タイトルが選ばれた場合の、映画タイトルが単独に選ばれる確率の推定を試みた。その確率を用いて非復元抽出によって2本の映画タイトルが選ばれた場合の、各映画タイトルが選ばれる件数の計算値が実測値に一致するような確率を求め、図1に示すように仮定の4つを選択肢とした場合にA、B、C、Dがそれぞれ選ばれる確率を4本の式からなる連立方程式を解いて求める。求めた確率を用いて、2本の映画タイトルの組み合わせが選ばれる件数理論値を計算し、実際に選ばれた件数が理論値を大きく上回った場合に、選ばれやすい組み合わせとして判定する方法を提案した。

映画タイトルのグループ分類は、映画タイトル間の関連性分析の結果をネットワーク分析することによって実施する。具体的には、ネットワーク全体の中から高密度のサブネットワークを抽出する方法として、モジュラリティQという指標[6]を導入し、Qの値が高い分割を探索する方法を用いる。モジュラリティQとは、ネットワーク全体のリンク数に対するサブネットワーク内の実際のリンク数の割合から、ネットワーク全体のリンク数を確保しつつランダムにリンクを張った場合のサブネットワーク内のリンク数の割合を引いた値を、すべてのサブネットワークについて合計したものである。このモジュラリティQが最も大きくなるようなサブネットワーク分割を探索する。なお、この分析は、ネットワーク分析ソフトウェア Pajek (<http://vlado.fmf.unilj.si/pub/networks/pajek/>) を用いて行った。

図2には、32本の映画タイトルをノードとし、関連性の高い映画タイトル間にリンクを張った映画タイトルのネットワーク表現とともに、モジュラリティQ値が最も大きくなる分割を合わせて示した。その結果、映画タイトルが4つのグループに分類されることが示された。

(4) マーケティング・セグメンテーション規模の時系列変化の推定に関する研究 [7],[8]

映画やファッションなどの分野では、商品の人気が時間とともに大きく変動する場合がある。いわゆる流行であるが、SNSなど、インターネットでの情報流通が盛んになり、それによって流行が短期的に大きく変動することもある。本課題では、映画やファッション、食品などの人間社会的流行に対して、レビューサイトやSNSに投稿された消費者から発信された情報を用いて、流行の時系列変化の推定に関する数理モデルを提案する。レビューサイトやSNSに投稿される消費者からの情報には、商品に関する情報も多々発信される。発信件数は、その時点での商品の流行の大きさを示すとらえることができる。したがって、特定商品に関する投稿件数の時系列データ

ータから、商品の流行の規模変化パターンを導出できる可能性があり、商品販売の戦略立案に有効な情報になると考えた。そこで、投稿件数の時系列データをもとに、商品ごとの流行の数理モデルの導出方法について提案した。

提案する数理モデルは、マーケティングにおけるイノベータ理論を考慮し、新しい形の流行モデルとして提案した。収集した SNS 投稿データの中から、一過性で終わるような流行、及び再び盛り上がりを見せるような流行などの投稿データを用いて、ベイズ推定のアプローチによる提案モデルのパラメータ推定を行い、さらに、投稿数データへのフィッティングを行った。本研究で提案されたモデルによって、再起や微細な振動を表現することが可能となり、実データへのフィッティングにおいて、その精度を従来手法よりも大幅に高めることができた。特に、再起タイプの流行である「KitKat」や「本麒麟」の SNS への商品に関する投稿データ数においては、多少のずれはみられるものの、これまでの手法では表現できない再起や微細な振動が再現された。これらの結果から、提案モデルは人間社会的流行の変遷を説明するための便利なツールであることが結論づけられた。

本研究では、ファッションや映画、食品などの人間社会的流行に対して、先行研究で提案された数理モデルを基に、疫学的流行を記述する数理モデルの基礎となった SIR モデル、及びマーケティングにおけるイノベータ理論を考慮し、新しい形の流行モデルを提案した。また、一過性タイプ、及び再起タイプの流行の実データを用いて、ベイズ推定のアプローチによる提案モデルのパラメータ推定を行い、さらに、実データへのフィッティングを行った。

なお、モデルにおけるパラメータ推定には、ベイズ推定 (MCMC, M-H 法) を導入し、実データからの推定を容易に実施できる方法として提案した。流行規模の変化には種々の要因が考えられるが、推定されたモデルのパラメータ値は、要因を推定するための大きな手がかりになる。推定されたパラメータの大小によって、普及させるために有効な方策を選択することができ、さらに施策によって流行の状況がどのように変化するかシミュレーションすることも可能となる。

本研究の成果としての、マーケット・セグメンテーション手法の実施と、その時系列的な変化に対する数理モデルの導出によって、マーケットの状況把握が容易に実施できるようになるとともに、効果的なマーケティング戦略を適時に実施可能になると考えられる。その際に必要なデータは、商品やサービスに関して消費者から発信されたレビュー情報や SNS への投稿情報であり、それを収集して用いることが特徴である。これら情報は、常に発信されており、容易に入手可能で情報入手のためのコストが安いこと、アンケートよりも偏りが少ないことなど、企業におけるマーケティング活動において有用な情報である。その有効活用について提言できた。

【参考文献】

- [1] Myers, J. H. and E.Tauber (1977) 'Market Structure Analysis', American Marketing Association.
- [2] Wind, Y. (1978) Issues and Advances in Segmentation Research, Journal of Marketing Research, XV, 317-337.
- [3] 朝野熙彦 (1998) 消費者行動の予測を目的としたマーケット・セグメンテーション, 『マーケティング・サイエンス』, 6/2, 45-66.
- [4] Pang, B., Lee, L. and Vaithyamathan, S.(2002) Thumbs up? Sentiment classification using machine learning techniques. Proceedings of the 2002 Conference on EMNLP, 79-86.
- [5] Dave, K., Lawrence, S. and Pennock, D.M.(2003) Mining the peanut gallery: Opinion extraction and semantic classification of product reviews, WWW, 519-528.
- [6] 水谷直樹 (2020) 商品選択履歴データをもとにした商品グループ分類手法の提案, 情報経営, Vol.80, 121-124.
- [7] 大田靖, 水谷直樹 (2021) 人間社会的流行における数理モデルの提案 - イノベータ理論と時間遅れの方程式を用いて -, 日本オペレーションズリサーチ学会和文論文誌, Vol.64, 22-45.
- [8] Ota, Y. and Mizutani, N.(2021) Estimating Parameters in Mathematical Model for Societal Booms through Bayesian Inference Approach, Mathematical and Computational Applications, Vol.25.

5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 2件/うち国際共著 0件/うちオープンアクセス 3件）

1. 著者名 水谷 直樹	4. 巻 80
2. 論文標題 商品選択履歴データをもとにした商品グループ分類手法の提案	5. 発行年 2020年
3. 雑誌名 情報経営	6. 最初と最後の頁 121 ~ 124
掲載論文のDOI (デジタルオブジェクト識別子) 10.20627/jsimconf.80.0_121	査読の有無 無
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 大田 靖 水谷 直樹	4. 巻 64
2. 論文標題 人間社会的流行における数理モデルの提案 - イノベータ理論と時間遅れの方程式を用いて -	5. 発行年 2021年
3. 雑誌名 日本オペレーションズ・リサーチ学会和文論文誌	6. 最初と最後の頁 22 ~ 45
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 Ota Yasushi, Mizutani Naoki	4. 巻 25
2. 論文標題 Estimating Parameters in Mathematical Model for Societal Booms through Bayesian Inference Approach	5. 発行年 2020年
3. 雑誌名 Mathematical and Computational Applications	6. 最初と最後の頁 42 ~ 42
掲載論文のDOI (デジタルオブジェクト識別子) 10.3390/mca25030042	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計2件（うち招待講演 0件/うち国際学会 0件）

1. 発表者名 伊藤颯大、大田 靖、水谷直樹
2. 発表標題 マーケティングリサーチにおけるネットワーク分析手法の活用
3. 学会等名 第69回数理社会学会大会
4. 発表年 2020年

1. 発表者名 内藤将大, 水谷直樹, 大田靖
2. 発表標題 流行モデルにおけるパラメータ推定の方法 - ベイズ推定の手法を用いて -
3. 学会等名 日本オペレーションズ・リサーチ学会 2019年春季研究発表会
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------