

称号及び氏名	博士(理学) 濱野 慎一
学位授与の日付	平成31年3月31日
論文名	「Semantic Analysis Models of Association Rules」
論文審査委員	主査 佐藤 優子 副査 石井 伸郎 副査 馬野 元秀 副査 寶珍 輝尚 副査 向内 康人

論文要旨

コンピュータの飛躍的な発展により、科学実験によるデータや顧客データが大量に蓄積されるようになり、それらのデータから知識・規則の発見を試みる研究が盛んに行われている。その中でも、意思決定システムや **Data Mining** が注目され、決定木やクラスタリング、結合ルール等の手法の提案、その効率化に関する多くの手法が提唱されている。その背景には、経営戦略決定や科学的発見に繋がる知見を得ようという大きな潜在的な要求があるものと思われる。

本論文では、大量に蓄えられたデータから有益な情報を抽出する情報技術である **Data Mining** を扱う。**Agrawal** 等によって提唱された **Association rule mining**(Agrawal et al.'93)は、小売店舗情報からマーケティングの専門家でも知り得なかった有益な顧客行動パターンを抽出することに成功した。代表的な例として挙げられるのは“おむつ”と“ビール”の同時購買パターンを抽出したもので、**Data Mining** が隠れた顧客の購買パターンを抽出、分析する手法として一躍脚光を浴びるようになった。

しかし、従来の **Association rule mining** による顧客行動分析では、抽出される **rule** が既知の顧客行動パターンであったり、全く解釈を行えない顧客行動パターンであったりと、興味深い **rule** を抽出することが困難であった。本研究では、大規模データから局所的な顧客行動パターンを抽出するべく、**Association rule** による顧客行動分析の拡張モデルとして異なるアプローチにより2種類のモデルを提唱する。局所的な顧客行動パターンとは、ある与えられた条件の下で成立する顧客行動パターンを表し、この条件とは顧客間に共通する、ある行動パターンを含む事を意味する。

(1)**Association Rule Analysis Model via Mediator**: 同時に出現する確率が低い商品対に、**Mediator** と呼ばれる各商品と強い従属性を持つ頻出商品集合を介在させることにより、商品対の間接的な関係性を **rule** として抽出し、分析を行うモデルを提唱する。

(2)**Association Rule Analysis Model via Item Response Theory**: 顧客行動に影響を与える顧客の嗜好を、潜在変数として **Association rule** の分析に導入し、潜在的な顧客行動パターンを抽出し、分析を行うモデルを提唱する。

本研究では、潜在特性モデルの1つである項目反応モデル(Lord et al.'68)を適用する。この2種類のモデルに共通するのは、顧客間に共通する行動パターンを基に、各顧客の特有のパターン(局所的な顧客行動パターン)を抽出し、比較分析を行う点である。

(1) Association Rule Analysis Model via Mediator

Association rule mining に代表される顧客の同時購買行動の分析だけでなく、“なぜこの商品は同時購買されないのか？”などといった新たな顧客行動パターンの分析のために Negative Association rule(Savasere et al.'98)が提唱された。この Negative Association rule の枠組により既存の Association rule mining からは抽出不可能であった新たな顧客行動パターンから有益な仮説を構築することが可能となった。しかし、有益なパターンが抽出される一方、抽出される Negative Association rule の数が膨大となる為、利用者にとって全ての Negative Association rule を分析することは非常に困難であった。

効果的に Negative Association rule の抽出及び分析を行うモデルとして Indirect Association rule mining(Tan et al.'00)の概念が Tan 等によって提唱された。

Indirect Association rule mining とは、同時に購入される割合が低い商品対(a,b) が Mediator と呼ばれる商品集合 M を介在させることにより、間接的な従属性を高い割合でもつ rule の抽出技術である。Tan 等は、商品と Mediator の間の従属性を表す指標として、IS と呼ばれる指標を用いている。この指標は、ある条件の下では、相関係数に近似できる指標であるが、従属性に関して方向性がなく、品目間に統計的な独立性が成り立つときでも大きな値をとる。本研究では、Indirect Association Rule の指標として、新たに μ 指標を導入する。 μ 指標は、Association rule 及び Negative Association rule の興味深さを同時に測定・分析することが可能であり、この指標は Shapiro や Zhang('00)によってそれぞれ提唱された“良い指標の条件”をすべて満たすことを示す。そして、Indirect Association Rule の数値実験として実際のビジネスデータであるドラッグストアの POS データを解析する。そこで発見された rule はメーカーのマーケターなどにとって興味深く、判りやすい rule となっている。この顧客行動レベルでのスイッチング分析、及び競合分析の結果、そこで利用する μ 指標の有益性を示す。

(2) Association Rule Analysis Model via Item Response Theory

Marketing Science では、これまで消費者行動の分析が様々なモデルに対して行われてきた。特に顧客の嗜好を潜在変数として取り入れた潜在因子分析は、現在まで有益な結果をもたらしてきた。しかしその一方で大量のデータを取り扱うことは困難であり、POS データ等の個人の顧客履歴の分析などには適用が困難であった。本研究では、Data Mining の技術と Marketing Science の潜在因子の概念を融合させることにより、現在まで誰も挑戦しなかった Association rule の潜在特性の抽出及び興味深さを測定するモデルの提唱を行った。

様々な潜在因子モデルが存在するが、本研究では項目反応モデル(Item Response Model)を採用した。項目反応モデルは、テストの作成・運用・評価が可能な数理モデルであり、その解釈のしやすさや応用のしやすさなどで近年注目を集めている。特に情報技術の発達により潜在因子の推定が比較的 low cost で行えるようになった為、医学や心理学などの分野への応用も盛んに行われ、有益な結果をもたらすようになってきた。

項目反応モデルの Data Mining、特に Indirect Association rule mining の枠組みへの適用には、以下の様な問題点が存在する。

(i) 既存の Indirect Association rule mining の枠組みでは、2 選択肢間での顧客行動パターンを抽出・分析できるが、実際には多選択肢間で顧客の意思決定が行われており、真の顧客行動パターンを表現できていない。Multinomial Logit Model に代表されるマーケティングモデルのように、多選択肢を想定したモデルの構築が必要である。

(ii) 項目反応モデルを潜在因子モデルとして適用する為には、局所独立の条件を満たす必要がある。現在の多くの研究者は、モデルが局所独立の条件を満たしているという仮定の下で分析を行

っている。

(iii) Association rule mining 等の枠組みにおける既存の手法では相関係数が小さい為に、rule 抽出の際に多くの商品対が枝刈りされていた。

上記の問題点を解決する為に、本研究では選択肢集合を生成し、ここから顧客の行動パターンを抽出・分析・解釈を行うことのできるモデルを提唱する。

また頻出商品集合は顧客行動の局所的なパターンの抽出の際に妨げになってきたが、顧客の共通する行動パターンとして頻出商品集合を抽出し、選択肢集合との反応パターンを項目反応モデルの枠組みにおける潜在因子推定を行うことにより、潜在的な顧客行動パターンを抽出することに成功した。

Association rule や Indirect Association rule などは基本的に 2 値データを取り扱っており、マーケティングにおいて重要な価格やプロモーション等の情報（コーザルデータ）を同時に扱うのが困難であった。本研究では、これらのデータを取り扱い有益なパターンを抽出することのできるモデルとして、段階反応モデルと Causal rule mining(Zhang et al.'02)の枠組みの融合を試みた。

Samejima によって提唱された段階反応モデル('69)は複数の値が順序関係を示しているデータを分析する枠組みで、現在では心理学（YG 性格検査等）の分野で非常によく利用されている。Causal rule は量的データを取り扱ったパターン抽出の技術で、因果関係を明瞭に表現することのできる rule mining の技術の 1 つである。

段階反応モデルと Causal rule mining を融合したモデルの提唱により RFM 分析など多値データの利用が可能となり、既存の rule mining 技術から抽出された rule よりも、より多彩な顧客行動パターンを表現した rule を抽出することが可能となった。また、潜在変数を rule の興味深さの指標として導入したことにより、データに依存されない統一的な指標で、抽出された rule の興味深さを評価することが可能となった

本研究では、局所的ではあるが興味深い顧客行動パターンを rule として抽出し、分析する為に、2 種類のモデルを提唱した。

審査結果の要旨

本論文では、大量のデータから顧客行動パターンの抽出を目的に、結合規則(Association rule)による顧客行動分析の拡張モデルとして2つの独創的なモデルを提唱している。1つは、間接結合規則の解析モデルである。間接結合規則とは、同時に購入される割合が低い商品対が Mediator と呼ばれる商品集合を介在させることにより、間接的な従属性を高い割合でもつ規則である。この研究では、間接結合規則の指標として、新たに μ 指標を導入している。 μ 指標は、Mediator と商品の間の相関を反映した指標で結合規則及び否定型結合規則の興味深さを同時に測定・分析することを可能とする。この指標に関して、Shapiro や Zhang によって提唱された“良い指標の条件”のすべてを満たす等の理論的結果も得ている。さらに、実際のビジネスデータであるドラッグストアの POS データにこのモデルを適用し、従来の結合規則の枠組では得られなかった興味深く、かつ、判りやすい間接結合規則を発見している。

2つ目は項目反応理論に基づく結合規則解析モデルである。データマイニングにマーケティング科学の潜在因子の概念を用いることにより、結合規則の潜在特性の抽出及び興味深さを測定する新しいモデルの提唱を行っている。Samejima によって提唱された段階反応モデルは、複数の値が順序関係を示しているデータを分析する枠組みで、現在では心理学等の分野でよく利用されている。この研究では、段階反応モデルをコーザル規則マイニングに適用している。コーザル規則は、マーケティングにおいて重要な価格やプロモーション等の情報(コーザルデータ)や量的データを取り扱うことができ、因果関係を明瞭に表現するという特徴がある。段階反応モデルをコーザル規則マイニングに適用したモデルの提唱により RFM 分析などの多値データの利用が可能となり、既存の結合規則マイニング技術により抽出された規則よりも多彩な顧客行動パターンを表現した規則を抽出することが可能となっている。さらに、潜在変数を規則の興味深さの指標として導入したことにより、データに依存しない統一的な指標で、抽出された規則の興味深さの評価も同時に可能とした。

以上の通り、この論文で提起された2つのモデルはいずれも斬新な解釈を可能とする結合規則を導き、従来のデータマイニングの枠組みでは得られなかった顧客行動パターンの抽出に成功するなど、今後の新しい展開が大いに期待できる成果である。学位論文審査委員会は、学位論文の審査ならびに最終試験の結果から、申請者に対して博士(理学)の学位を授与することを適当と認める。