

# ID 付 POS データによる購買行動の季節変化の分析と可視化

Analysis and Visualization of ID-POS data how purchasing behavior changes along with annual seasons.

原田 奈弥<sup>1,2\*</sup> 山下 和也<sup>2</sup> 本村 陽一<sup>2</sup>

Nami Harada<sup>1,2</sup>, Kazuya Yamashita<sup>2</sup>, and Yoichi Motomura<sup>2</sup>

<sup>1</sup>豊田自動織機

<sup>1</sup>TOYOTA INDUSTRIES CORPORATION

<sup>2</sup>産業技術総合研究所 人工知能研究センター

<sup>2</sup>Artificial Intelligence Research Center,

National Institute of Advanced Industrial Science and Technology

**Abstract:** These days ID-POS data are widely used with AI to understand consumers and improve retailing services. We suggest a way to make longitudinal research of ID-POS data to understand how and why any change in people's purchasing along with seasons with pLSA and visualized our result. Many people change what they buy in each seasons, but others don't in some seasons. So we investigated the reason why some people don't change their purchasing in some seasons with Bayesian network. The result told us that the seasonal purchasing changes are effected from consumers' family members and their preference. The result implicates that some changes in consumers' family members can cause some change of their purchasing behavior. Visualizing the result of AI's data analysis will help us understand what AI tell us and decide what we do to improve our businesses or lives.

## 1. 序論

### 1) 背景

昨今日本では、ライフスタイルの多様化が進んでいる。その要因の一つは内閣が進めている一億総活躍社会である。性別や年齢、家族構成、ハンディキャップの有無などのバックグラウンドを問わず社会の中で役割をもって貢献する人が今後、一層増えることで、仕事やプライベートの時間帯や過ごし方が多様化している。ライフスタイルの多様化のもう一つの要因は少子高齢化である。一人暮らし世帯や老老介護世帯などが増え、世代毎に典型的なライフスタイルやライフステージを当てはめることが妥当ではなくなってきている。

ライフスタイルの多様化に伴い、購買行動や需要も多様化している。世代や職業、家族構成の情報からこれまでの典型的なライフスタイルと、そこから推測される需要を当てはめては、人々がそれぞ

れに生活の中で必要とする商品などを予測するには不十分である。

また、ライフスタイルがあまりにも多様化しているため、データサイエンティストがそれぞれのライフスタイルに応じた消費行動の予測を行うのは非現実的である。年齢や家族構成毎のライフスタイルがあまりにも多様であるばかりでなく、データサイエンティストの深刻な人材不足が叫ばれる昨今では実現が困難である。

したがって、ライフスタイルの多様化に伴う消費や購買の多様化に応えるには、人工知能を活用した需要の予測や調達が望まれる。

### 2) ID 付 POS データとその活用

POS データとは、小売店等のレジを通して商品を購入するときに通常収集できるデータである。商品名や価格、購入した個数を、レジを通過した日時や店舗ともに記録されている。

多くの小売店などでは、お客様一人ひとりに ID 番

号を割り当てて番号で個人を識別できるようにしている。ID 番号が照合できる ID 付カードと POS データを連動させることで、購入者と購入商品や商品を購入した時間などのデータを照合させることができる。これが ID 付 POS データである。

ID 付 POS データ活用の先行事例については、石垣(2011)[1]に詳しく述べられている。ID 番号と共に、家族構成や職業、年齢、消費行動などに関する嗜好性のアンケートへの回答を収集するものが既に多数運用されている。例えば、個人的なプロフィールや嗜好性をデータとして収集することで、お客様一人ひとりに対してそれぞれに、ニーズや日々の困りごとなどを推測し、その人に合っていると考えられる販売施策を個別に、アンケートなどと一緒に回答された住所やメールアドレス、アプリを通してスマートフォンなどに対して行うことができる。川島(2016)[3]では、ID 付 POS データから、各 ID のお客様の、来店した店舗と来店時間、購入商品について pLSA でクラスタリングを行った。お客様がどのような店舗で、いつ、どのような商品を購入するかについて、店舗がある地域や、駐車場の有無、各お客様のお買い物の嗜好性と確率的因果推論を、ベイジアンネットワークを用いて行い、それぞれの地域や店舗別で、その店舗のお客様の望まれる設備や仕入れを整えていけるような AI を活用したシステムを提案している。

購買した日時も含めたモデリングを行っている事例としては、流通量販店のデータを扱った石垣ら(2011)[1]や、百貨店のデータを扱った石垣ら(2011)[2]があるが、これらの事例では、季節や時間はそれぞれのお客様の購買行動に影響する要因として扱われている。

これまでの研究では、季節や年齢、ライフステージや家族構成の変化に影響を受けた、時間軸上の変化のモデリングや可視化ができていないことが課題である。

### 3) 目的

本研究では、人工知能を用いて購買行動の時間に伴う変化や、その変化を誘発していると推測される要因を可視化し、人々が季節や年齢の変化に伴い、なぜ、どのように購買行動が変わっているかをいかにして知ることができるか、その手段の検討を目的とする。ID 付 POS データを用いて、購買行動の時系列上の変化をモデリングし、可視化と、その変化の因果推論を行っていく。

人工知能の確率モデリングによるアプローチを用いることで、購買行動の変化について、確率的に推測される因果も含めて可視化することができる。人

工知能による、事象や推測される因果関係をモデリングによって可視化することで、人工知能と人の相互理解に努めることも本研究の目的である。すなわち、ID 付 POS データを、人工知能による学習やモデリングを通して、人工知能が人の購買や消費の行動を理解する。その人工知能の出力結果から、人は人工知能の算出したモデルを理解し、課題解決や改善を行うことができる。ID 付 POS データの分析を通じて人と人工知能が相互に理解をすることで、販売戦略の向上により、必要な商品を、必要な人に、必要なときに届けられ、購買や消費によって欲求に応えたり日々の困りごとを解決したりできる、豊かな社会を作り出すことが、ID 付 POS データに対する人工知能の確率モデリングのアプローチを行う究極的な目標である。

### 4) 本稿の流れ

本稿では、2 章で、取り組んだ方法やプロセス、用いた AI 手法、用いたデータの紹介を行う。3 章で、結果を述べ、その結果から得られる考察を 4 章で述べる。5 章で、本研究で得られた知見などを踏まえた今後の課題について説明し、6 章の結論にて、本研究をまとめ、社会において AI の活用を進めるうえでの本研究の意義について考える。

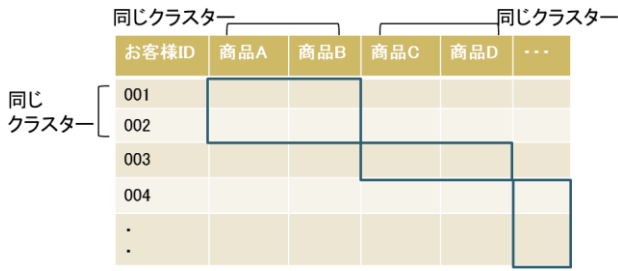
## 2. プロセス

本研究では、ID 付 POS データに対して、人工知能の確率モデリングによるアプローチによって、時間軸上の変化と、その確率的な因果推論の算出を行った。

本研究では、これまでできてこなかった、お客様それぞれの、購買行動の時間に伴う変化のモデリングと可視化を試みる。

### 1) pLSA とは

pLSA とは、Probabilistic Latent Semantic Analysis(確率的潜在意味解析法)の略称である。pLSA では、図 1 のように 2 つの変数が同時に観測される、共起確率が高いものを共通の潜在的な発生要因があると考ええる。



<図1：pLSA とは>

pLSA のクラスタリングの算出方法としては、図2のように、2つの変数  $x, y$  間である値を同時に引き起こす潜在的な要因  $z$  を仮定したときの尤度関数  $L$  を最大にするものを求める。



<図2：潜在的な要因の考え方>

$$p(x, y) = \sum p(y | z) p(x | z) p(z)$$

$$L = \sum \sum N(x, y) \log p(x, y)$$

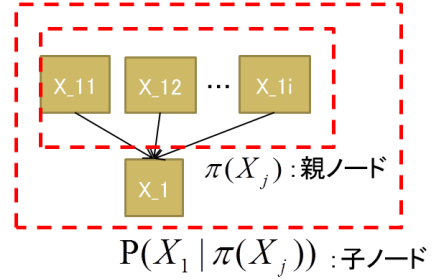
共起確率を考えることにはいくつかメリットがある。観測される変数の分布に対する仮定が不要になる。離散変数だけでなく、離散化することで連続変数でも pLSA を行うことができる。異なる分布や形式の変数を、同時に、厳しい仮定や制約なく扱うことができる。

本研究では、pLSA の算出には産業技術総合研究所が所有する分析ツールである、APOSTOOL を用いて行った。

## 2) ベイジアンネットワークとは

ベイジアンネットワークとは、現象同士の階層的な確率的因果関係をグラフィカルに表現する手法である。データが観測された現象について、ベイズの定理から得られる確率的な因果関係を最も当てはまりよく説明できるモデルを、観測値の条件付確率から探索的に求める。

モデリング算出の考え方について述べる。階層構造の親ノードと子ノードの関係について、図3のように示される。



<図3：ベイジアンネットワークの考え方>

図3のモデルの中で、親ノード  $x_{11}, \dots, x_{1i}$  が観測さ

れた元での子ノード  $x_1$  が観測される条件付確率を最大にするモデルをベイジアンネットワークでは探索している。

本研究では、ベイジアンネットワークのモデル構築には、株式会社NTTデータ数理システムのソフトウェアである Bayonet を用いた。

## 3) プロセス

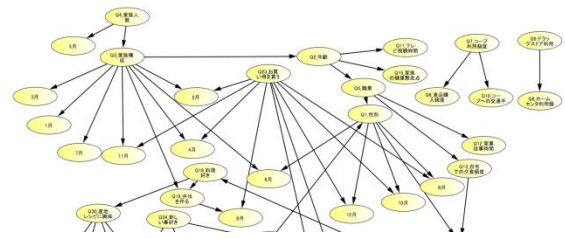
本研究では、ID 付 POS データに対して、お客様の各時点での購買行動の経時的な変化を捉えるため、同じお客様を示す ID を、測定時点毎に別々に発生させた。具体的には例えば、ID 番号が 1234 の人に対して、各月で、1234 Jan., 1234 Feb., ... といったように、同じ ID でも、それぞれの時点での購買行動を区別できるようにデータを加工した。ID の要素に対してこの加工を行ったデータと、それぞれの時点でそれぞれの ID に対応するお客様の購買の POS データを組み合わせたデータに対して、pLSA とベイジアンネットワークを行った。本研究では、測定時点を区切る間隔は月とした。その理由は、季節の変化に伴う購買行動の変化を可視化するのには月が適切な間隔だと考えたためである。

月の識別を付けた ID と、その月にその ID が購入した商品とで pLSA を行った。このプロセスによって、それぞれのお客様が、それぞれの月でどのような商品を購入したかによって商品とお客様をクラスタリングすることができる。

pLSA の結果として出力された、各 ID が各月で属するクラスターや、各商品の属するクラスターと、各 ID がアンケートで答えた家族構成や職業などのプロフィール情報や、購買や消費に関する嗜好性についてのアンケート回答を結合したデータに対してベイジアンネットワークを行った。各 ID 番号に対応したお客様が、各時点で各クラスターに属することについて、それぞれのお客様のプロフィールや嗜好性との確率的な因果をモデリングし、可視化した。

#### 4) 分析を行ったデータ

本研究では、兵庫県を主な事業エリアに約 150 店舗を展開する流通量販店の、2009 年 9 月から 2010 年の 8 月までの 1 年間の ID 付 POS データに対して分析を行った。アンケート結果と組み合わせて分析を行うため、該当の 1 年間の POS データのうち、プロフィールや購買行動に関するアンケートに回答した 3,965 名のお客様のデータを抽出し、それぞれのお客様の、月別の購買行動の変化のモデリングや、変化をもたらす要因についての因果推論を行った。



<図 5：各月の所属クラスターの分析>

図 5 の左上の家族構成と、1~4 月に季節感のないクラスターに所属する確率との関係が推測できる。同様に、弁当を作るかどうかと 9 月の所属クラスター、お買い得を買うかどうかと 10~12 月の所属クラスターとの関係も、このベイジアンネットワークの出力より推測できる。

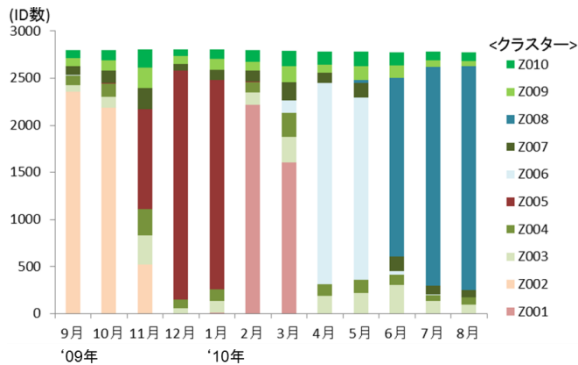
各 ID のお客様が、1 年間のうち何か月、季節感のないクラスターに属するかというデータと、プロフィールや嗜好性との関係も同様にベイジアンネットワークを用いてモデリングを行った。

### 3. 結果

前章で述べたプロセスに従い、結果を示す。

#### 1) クラスタ構成の推移

pLSA の出力より、クラスター数別の情報量基準の変化からクラスター数 10 のモデルを本研究では採択した。各クラスターの、1 年間の各月毎の、ID の所属数の推移を図 4 に示す。



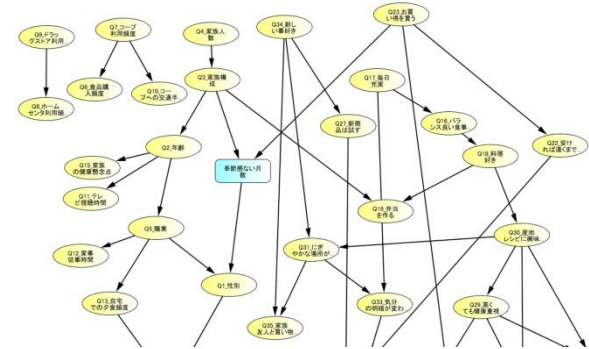
<図 4：各クラスターの所属 ID 数推移>

季節の変化に伴い、各月で最も大きな割合を占めるクラスターも変化しているのが分かる。一方で、毎月一定数のお客様が属しているクラスターもある。

#### 2) クラスタ間移動の誘発要因

1 年間、毎月一定数のお客様が属している、図 4 のクラスター Z003, Z004, Z007, Z009, Z010 を、季節感のないクラスターとして着目した。

各月に季節感のないクラスターに属する要因について、アンケートの回答とのベイジアンネットワークを行った出力結果を図 5 に示す。



<図 6：季節感のないクラスター所属の月数の分析>

この出力結果より、アンケート回答の中で、家族構成とお買い得を買うかどうかと、季節感のないクラスターに属する月数との関係が推測できる。また、季節感のないクラスターに属する月数を事前に条件として与えると、男女の差が大きくなっていることが分かる。

### 4. 考察

本研究のテーマである、季節や時間の変化に伴うお客様の購買行動の変化について、pLSA とベイジアンネットワークの出力結果から考察を行う。

#### 1) 季節の推移について

図4のpLSAの出力結果より、季節の変化に伴い、最も多く売れる商品のクラスターが変化していることが分かる。それぞれの季節で最も所属するID数が多いクラスターは、それぞれの季節の旬の食材やイベントに関係しているものが多いことが、各クラスターの商品構成を見てみると分かった。例えば、9、10月に最も多いZ002クラスターには、秋の旬の食材であるリンゴや梨、柿が属しており、11～1月に最も多いZ005には、手袋やコートなどの冬用の衣類や迎春用品が所属している。

<表1：季節感のあるクラスター>

クラスター	季節	主な商品
Z002	9, 10月	リンゴ、梨、柿などの秋の食材など
Z005	11～1月	冬用の衣類や迎春用品など
Z001	2, 3月	節分向け太巻きや伊予柑など
Z006	4, 5月	甘夏や新玉ねぎ、新ジャガイモなど
Z008	6～8月	アイスクリームや水着、浴衣など

一方で季節感のないクラスターには、通年で需要があると考えられる肉類や文房具、ペットフード、お惣菜などがそれぞれのクラスターに属していた。

<表2：季節感のないクラスター>

クラスター	主な商品
Z003	豚モモ肉や鶏胸肉などの肉類など
Z004	化粧品や文具、酒類など
Z007	ペットフードなど
Z009	和牛やブランド肉など
Z010	惣菜や加工食品、菓子など

## 2) クラスター間の移動について

図5、図6で示したベイジアンネットワークの出力から、お客様は季節の変化の中で、家族構成や購買行動の嗜好性との関係によって、季節感のあるクラスターとないクラスターの間を移動していることが推測できた。

月毎の所属クラスターの変化は、家族構成やライフスタイルと関係しているのではないかと考えられる。1～4月はお子様のいるお客様がより季節感のないクラスターに所属しにくいのは、この期間はお正月や卒業式、入学式など、家族で準備したり祝った

りするイベントが多いことが影響しているのではないかと考えられる。同様に、9月の行楽や運動会の季節には、お弁当を作る人が、より季節感のないクラスターに属する可能性が低い。10～12月は、冬物衣料やお節料理の準備など、他の季節に比べて単価の高い季節商品が多いため、お買い得を買う、節約志向の強いお客様は季節感のないクラスターに属しやすいのではないかと考えた。

図6で示したベイジアンネットワークの出力結果から、どのようなお客様が、傾向的により季節感のない購買行動へと変わりやすいかを推測することができた。家族構成として、季節感のないクラスターに属しにくいのは、小学生以下のお子様のいるお客様で、大学生以上のお子様のいるお客様、夫婦二人、単身のお客様の順に、季節感のないクラスターに属する確率が高くなっていく。この結果から、お客様の経時的な購買傾向の変化として、幼いお子様のいる家庭では、お子様の成長や独立などの時間の経過に従い、季節感のないクラスターに属する可能性が徐々に上がっていくのではないかと推測できる。

また、お買い得を買う節約志向の高いお客様ほど季節感のないクラスターに属しやすいという結果より、例えば定年退職などのライフステージの変化に伴う収入の変化と共に購買傾向が変化して、季節感のない購買行動に変わっていくことが考えられる。

## 5. 今後の課題

### 1) 地域軸を加えたモデリング

本研究では、該当の期間に該当のお客様が行った商品の購入を一括でモデリングを行ったが、ID付POSデータでは、データを収集した店舗を区別して分析することができる。店舗が所在する場所の違いによってその店舗のお客様に多く見られる世代や職業、ライフスタイルが異なることが容易に考えられる。本研究では時間軸上での変化のモデリングを行ったが、地域差による違いも加えてより細分化されたモデリングを行うことで、より詳細で緻密な結果を得られたり、精度の高い予測を行ったりすることができるようになるのではないかと考えられる。

### 2) 傾向的な変化のモデリング

本研究では、1年間の中で月の変化に伴う購買行動の変化についての分析を行った結果、周期的な季節変化についての推測を行うことができた。一方で、お客様のお子様の独立などの家族との離別や、高齢

化に伴う健康状態の変化などの影響を受けた、傾向的なライフスタイルの変化に伴い、より長期的に傾向的な購買行動の変化があることも考えられる。このような変化をいかにモデリングするかについても、今後検討を行い、数年分などのより長期的なデータに適用して有用性を検証していきたい。

特に、長期的に ID 付 POS データの分析を通じてお客様へのサービスの向上を行っていく中で、お客様のお子様の独立やペットの購入、死別といったライフステージや家族構成の変化を捉えることに関心を持っている。一般的には、アンケートを一定の頻度で行わなければ、お客様のライフステージや家族構成の変化を知ることは難しい。アンケートの作成や回答の回収、集計には莫大なコストや時間がかかるという問題がある。ID 付 POS データを通してお客様の購買行動を経時的に観測して変化を捉えることで確率的にライフステージや家族構成の変化を予測することができれば、少ないコストとタイムラグで、お客様に必要なサービスを提供できるよう、業務の質を向上することができる。

### 3) 計算の効率化

本研究では、各測定時点で、測定時点が分かるように ID を生成しなおしている。測定時点別で異なる ID 識別を行うことで、データ内の個体数が膨大になり、計算の時間や負荷が増大する。今後、モデルを探索、学習するアルゴリズムの検討などによって、計算の効率化を図っていきたい。

### 4) 汎用化に向けて

本研究では ID 付 POS データに対して、時間軸上の変化のモデリングを行ったが、この時間軸上での変化のモデリングや可視化のニーズは、小売りや流通以外の他分野でも大いにある。例えば、他のサービス業でのお客様のニーズの変化や、工業での、生産や品質の管理などである。他の分野でも汎用的に、より使いやすい分析プロセスを今後構築していきたい。そのためにも、前述の 1) ~ 3) の課題の解決が、今後急がれる。

## 6. 結論

### 1) まとめ

本研究では、ID 付 POS データから、購買行動が月の経過とともにどのように変化をしているかをモ

デリングし、可視化を行った。その結果、多くのお客様は季節毎に、季節と関係の強い旬の食材や季節のイベントに関する商品を購入しているが、各月で一定数のお客様は、季節感のない購買傾向にあること、同じお客様でも季節感のある購買行動とない購買行動が季節の間で変化をすることが分かった。購買行動の中の季節感の有無の変化は、お客様の家族構成などのライフスタイルや、節約指向によって引き起こされているのではないかと推測ができた。

これらの結果から、お客様個人個人のライフステージや健康状態などの変化に伴い今後起こるであろう傾向的な変化についても仮説を得ることができた。お客様のお子様の成長や独立といったライフステージの変化や、家計の収入の減少による節約指向上昇とともに、お客様がより季節感のない購買傾向へと変化していくのではないかと考えられる。このような傾向的な変化のモデリングなどの、今後の課題を洗い出すことができた。

## 2) 本研究の意義

本研究を通じて、これまでされてこなかった、ID 付 POS データに対して、時間軸上の変化について、モデリングと可視化を行った。これにより、購買行動の周期的な季節変化をモデリングすることができた。この、時間軸上の変化をモデリングするプロセスの提案は、時間軸上でマネジメントやサービスの対象が変化する様々な分野での問題解決の助けになるのではないかと考えられる。様々な分野で活用できるよう、計算の効率化や汎用化について、今後取り組んでいきたい。

人工知能の応用については、購買行動の時間軸上の変化を可視化し、変化を引き起こす要因について確率的なアプローチからの因果推論と可視化を行うことで、人と人工知能の相互理解と相互補完を促すことができた。すなわち、人が時間軸上の購買行動の変化の可視化や因果推論を行うという目的を明確に持つことで、人工知能をいかに活用するかの方向付けを行うことができる。人工知能は人の目的に応じて、観測された現象について、確率的なアプローチからの記述による可視化と因果推論を行うことで、どのような現象が、なぜ起きているのかについての推測を出力する。人工知能の出力を人が、因果構造も含めて理解することで、解決すべき課題や現状からの改善についてのアイディアを得ることができる。この、人と人工知能が相互理解するプロセスを繰り返すことで、人工知能が行うモデリングはより妥当なものとなり、人は人工知能を活用することで、より豊かな生活を送ることができる。

本研究が、ID 付 POS データの更なる活用のみならず

らず、このような人工知能と人の、情報のやり取りを通じた相互作用による豊かな社会の構築の一助になれば幸いである。

## 7. 謝辞

本研究は NEDO 委託事業「人間と相互理解できる次世代人工知能技術の研究開発」の支援を受けて行いました。

私に、産業技術総合研究所 人工知能研究センターという素晴らしい環境で、人工知能とその活用について学ぶ機会を与えていただいている豊田自動織機に感謝申し上げます。

## 8. 参考文献

- [1] 石垣司, 竹中毅, 本村陽一: 日常購買行動に関する大規模データの融合による顧客行動予測システム, 人工知能学会論文誌, 26 巻, 6 号, D, pp.670-681, (2011)
- [2] 石垣司, 竹中毅, 本村陽一: 百貨店 ID 付き POS データからのカテゴリ別状況依存的変数間関係の自動抽出法, オペレーションズ・リサーチ, 2011 年 2 月号, (9), pp.77-83, (2011)
- [3] 川島健佑: 価値観とライフスタイルを考慮した確率的潜在意味構造モデルの活用技術 平成 27 年度東京工業大学知能システム科学専攻修士論文 (2016)
- [4] 鈴木譲: ベイジアンネットワーク入門 確率的知識情報処理の基礎 培風館 (2009)
- [5] 本村陽一: ベイジアンネットワークによる確率的人間行動モデリング 電気通信大学大学院電気通信学研究科学位論文 (2008)

---

\* 連絡先: 産業技術総合研究所 人工知能研究センター  
〒135-0064 東京都江東区青梅 2-3-26  
E-mail: nami.harada@aist.go.jp