

2012年度 S-PLUS学生奨励賞

# 購買行動トランザクションデータによる 顧客クラスタリングとその活用

---

静岡大学大学院情学研究科 原野 朱加

# 目次

1. Executive Summary
2. 背景と目的
3. 使用するショッパー行動データ
4. 分析の流れ
5. トランザクションデータ処理
6. クラスタリング
7. クラスタプロファイリング
8. 商品購入要因分析
9. マーケティングアプローチの提案
10. 終わりに

# 1. Executive Summary

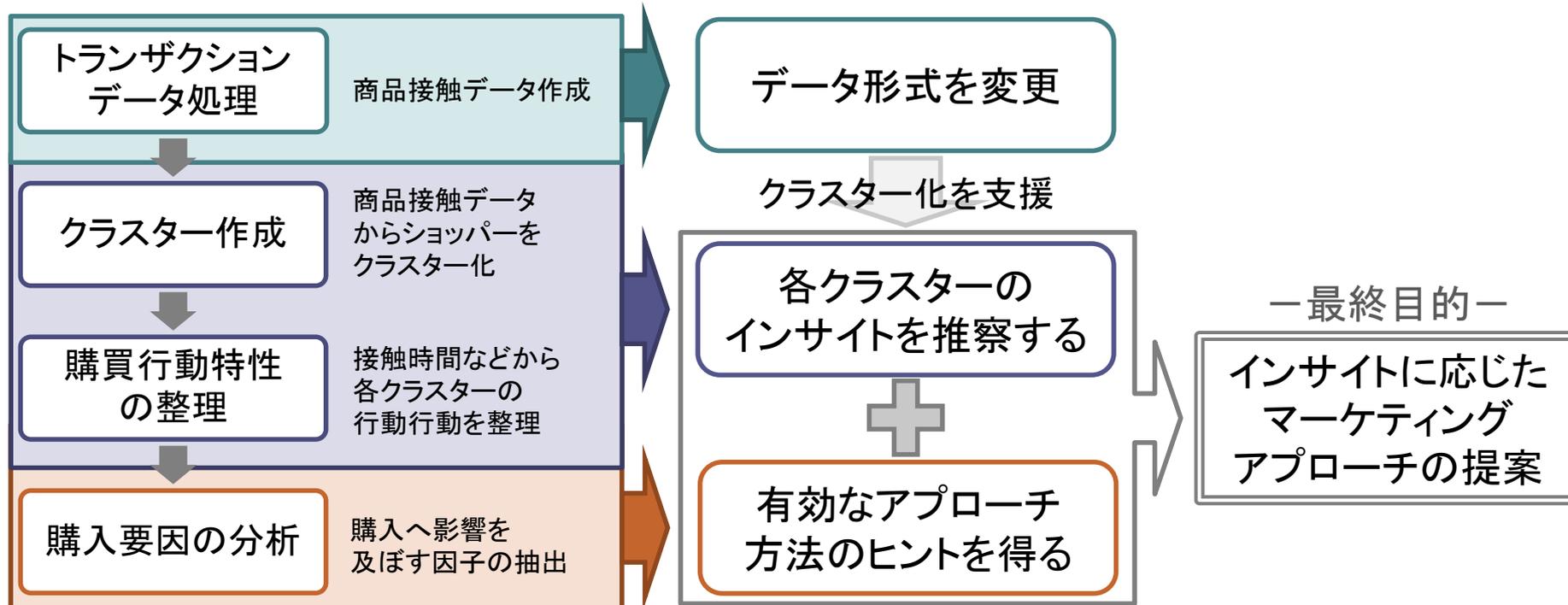


ショッパー行動データ

店舗に現れたショッパーの接触した商品や時間、ブランド、価格など様々な情報が得られる。

課題

ショッパーの**購買行動と購買心理(インサイト)**が知りたい  
**トランザクションデータ**のためレコード数が多様で分析が複雑化



## 2. 背景と目的(1/2)

### ショッパーマーケティング

マーケティングの対象を、コンシューマー（消費者）ではなく、あるチャネルにおけるショッパー（購買者）とし、「ショッパーの行動に関する深い理解に基づいて開発され、ブランドエクイティを構築し、ショッパーを惹きつけ、購買決定に導くために計画されたすべてのマーケティング刺激からなる活動」である。（Grocery Marketing Association, GMA/Delloite, 2007）

日本でも近年注目が高まっており、今後拡大が期待される

方法

ショッパー・インサイトを捉える

ショッパー・インサイト: 購買者の売り場での行動とその心理のこと

行動データ

どう行動し何を買ったか？



心理データ

何を考えて購買したか？

目的

ニーズを捉えた店頭マーケティングを実施する

品揃え／陳列場所・方法／店頭コミュニケーションツールの設置など

## 2. 背景と目的(2/2)

### ▼ショッパー・インサイトを捉えるためのデータ

#### 行動データ

##### ・店内回遊データ

店内に座標を設定し、無線LANやRFIDなどで取得

##### ・商品接触データ

店内カメラや同行調査によって取得

##### ・情報接触データ

アイカメラやマイクによって取得

##### ・商品購入データ

IDPOSやアンケートによって取得

#### 心理データ

##### ・知覚／記憶／感情／関与データ

インタビューや店頭・インターネットアンケートを用いて取得

2つのデータの組み合わせ方や  
活用方法に工夫が必要

ただし多くのコストと時間がかかる

一つのデータから購買行動と購買心理(ショッパー・インサイト)を推測できないか？

#### 目的

ショッパー行動データから購買行動だけでなく購買心理も読み取ることで  
ショッパー・インサイトを推測し、店頭マーケティングへ結びつける

### 3. 使用するショッパー行動データ(1/2)

ドラッグストア(化粧品コーナー)における商品接触データ

提供元: 株式会社Midee

化粧品コーナーへ訪れてから立ち去るまでの行動を店内カメラで撮影

ショッパーA0101

化粧品コーナー



商品を手にする



悩む



商品を戻す

(上記の作業を繰り返す)



化粧品コーナーへ  
訪れる

1つ~複数の商品を手にする

化粧品コーナー  
から離れる

トランザクション型ショッパー行動データ



トランザクション型であるため、分析が複雑であるが、多くの情報を得られる

### 3. 使用するショッパー行動データ(2/2)

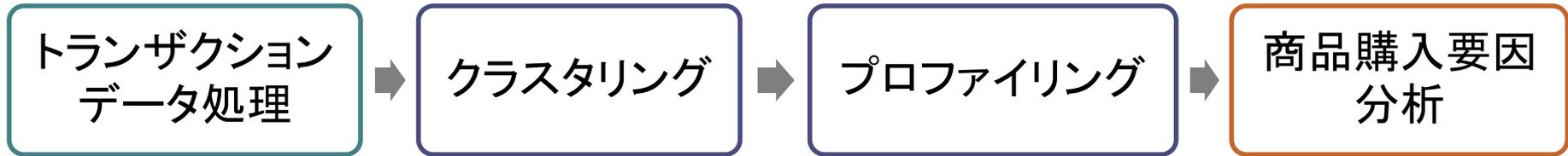
#### データ形式

Shopper	商品名	ブランド名	価格	接触時間	棚前滞在時間	商品接触までの時間	接触商品点数
306	肌研 極潤ヒアルロンミスト つけ替え 45ml	ハダラボ	798	0:00:08	0:01:37	0:00:46	3
306	肌研 極潤ヒアルロンミスト 45ml	ハダラボ	838	0:00:02	0:01:37	0:00:54	3
306	肌研 極潤ヒアルロン美容液 30g	ハダラボ	2100	0:00:04	0:01:37	0:00:59	3
307	肌研 極潤ヒアルロンミスト 45ml	ハダラボ	838	0:00:24	0:01:16	0:00:09	3
307	肌研 極潤ヒアルロンミスト つけ替え 45ml	ハダラボ	798	0:00:04	0:01:16	0:00:40	3
307	BR モイストクリーム 40g	バリアリペア	1280	0:00:07	0:01:16	0:01:07	3
308	うるおい屋 しっとりクリーム 47g	ナリスアップ	880	0:00:04	0:00:24	0:00:11	1
309	うるり 高保湿液しっとり 200ml	ウルリ	880	0:00:09	0:01:15	0:00:02	1
310	モイスタージュ Eローション さっぱり 210ml	モイスタージュ	698	0:00:10	0:04:10	0:00:43	2
310	肌研 極潤ヒアルロン液 170ml	ハダラボ	980	0:00:05	0:04:10	0:00:50	2

#### データ内容詳細

- ・ブランド名:各商品のブランド名
- ・価格:各商品の店頭表示価格
- ・接触時間:ショッパーがその商品を手に取り接触していた時間
- ・棚前滞在時間:1ショッパーが化粧品コーナーに現れてから立ち去るまでの時間
- ・商品接触までの時間:棚を訪れるor前の商品を離してからその商品に接触するまでの時間
- ・接触商品点数:1ショッパーが1回のコーナー立ち寄りの中で接触した商品の合計数

## 4. 分析の流れ



### ①トランザクションデータ処理

トランザクション型のショッパー行動データから、ショッパーが接触した商品を把握するためのデータを作成する

### ②クラスタリング

作成したデータから、ショッパーを商品接触パターンによりクラスタリングする

### ③クラスタープロファイリング

クラスターごと、購入割合、検討時間、購買点数、接触点数などの分布を確認し、それぞれ購買行動の特性を整理する。

### ④商品購入要因分析

クラスターごと、検討時間、接触点数、商品接触までの時間がどのように購入有無へ影響しているかを分析する。

# 5. トランザクションデータ処理(1/3)

トランザクション型ショッパー行動データから  
商品接触パターンを捉える

—商品接触パターン—  
ショッパーの選好グループの傾向  
価格帯選好で商品を選んでいるのか？  
ブランド選好で商品を選んでいるのか？など....

しかし、パターンが多様すぎて捉えづらい

トランザクション型ショッパー行動データ

—一人一人が接触した全商品のブランドや  
価格が商品ごと記録してあるため  
多様な商品接触行動を観察することができる

**問題** ショッパーごとレコード数が違う

ショッパー行動データを利用し  
商品接触行動を捉えたいが、レ  
コード数がばらつき扱いづらい



データを1ショッパー1レコードと  
して記録した新たな商品接触  
データを作成し、商品接触パ  
ターンを捉えたクラスタリングを  
可能にする。

▼作成する商品接触データ  
Data1:接触ブランド数  
Data2:価格帯のばらつき  
Data3:最も接触している価格帯

# 5. トランザクションデータ処理(2/3)

## ▼商品接触データ作成手順

### 前処理: 価格帯の設定

各ショッパーが接触した商品それぞれの価格データ(量的データ)をクラスタリング用に価格帯データ(質的データ)に変換する.

価格帯データのラベル設定の定義は以下のとおりである.

▽低価格帯(low): 688円以下

▽中価格帯(middle): 689円~1259円

▽高価格帯(high): 1260円以上

※設定基準は全商品を価格別で昇順に並べ、下位30%の商品=低価格帯、上位30%の商品=高価格帯、それ以外=中価格帯とした.

### Data1: 接触ブランド数

各ショッパー $S_i$ が接触した商品は、何種類のブランドで構成されていたか

例)  $S_i$ の接触商品リスト $s_i = \{X_1, X_2, X_3\}$ であるとき、それぞれの商品のブランドをブランドリスト $s_i' = \{M, M, L\}$ として格納する.

このブランドリストの重複するブランドを省いた、ユニークなブランドリストをブランドリスト'とすると、ブランドリスト $s_i' = \{M, L\}$ となる.

このとき、接触ブランド数 $s_i = \text{length}(\text{ブランドリスト}')$ とする.

## 5. トランザクションデータ処理(3/3)

### Data2:価格帯のばらつき

各ショッパー $S_i$ が接触した商品は、特定の価格帯に集中しているのか、ばらついているのか

例)  $S_i$ の接触商品リスト $S_i = \{X_1, X_2, X_3\}$ , それぞれの商品の価格帯を商品価格帯リスト $S_i = \{\text{high}, \text{low}, \text{high}\}$ として格納する. 価格帯の商品数を価格帯別カウントリスト $\text{Price}_{S_i} = \{\text{Price}_{\text{high}}, \text{Price}_{\text{middle}}, \text{Price}_{\text{low}}\}$ とすると, このとき $\text{Price}_{\text{high}}=2$ ,  $\text{Price}_{\text{middle}}=0$ ,  $\text{Price}_{\text{low}}=1$ となる.

このとき, 価格帯のばらつき $S_i = \sqrt{\frac{1}{3} \sum (\text{Price}_{S_i} - \overline{\text{Price}_{S_i}})^2}$  とする.

### Data3:最も接触している価格帯

各ショッパー $S_i$ が接触した商品のうち, 最も多く接触した価格帯はどれか

Data2と同じく接触商品の価格帯別カウントリスト $\text{Price}_{S_i}$ を設定する.

このとき, 最も接触している価格帯 $= \max(\text{Price}_{S_i})$ とする.

※ $\max(\text{Price}_{S_i})$ が複数存在する場合 (例: $\text{Price}_{\text{high}}=2$ ,  $\text{Price}_{\text{middle}}=2$ のとき)

接触商品の各価格帯の接触時間 $\text{Time}_{S_i} = \{\text{Time}_{\text{high}}, \text{Time}_{\text{middle}}, \text{Time}_{\text{low}}\}$ を計算し,

最も接触している価格帯 $S_i = \max(\text{Time}_{S_i})$ とする.  $\max(\text{Time}_{S_i})$ も複数存在する場合はランダムとする.

## 6. クラスタリング(1/2)

作成した商品接触データを利用し、  
商品接触パターンを捉えたクラスタリングを行う

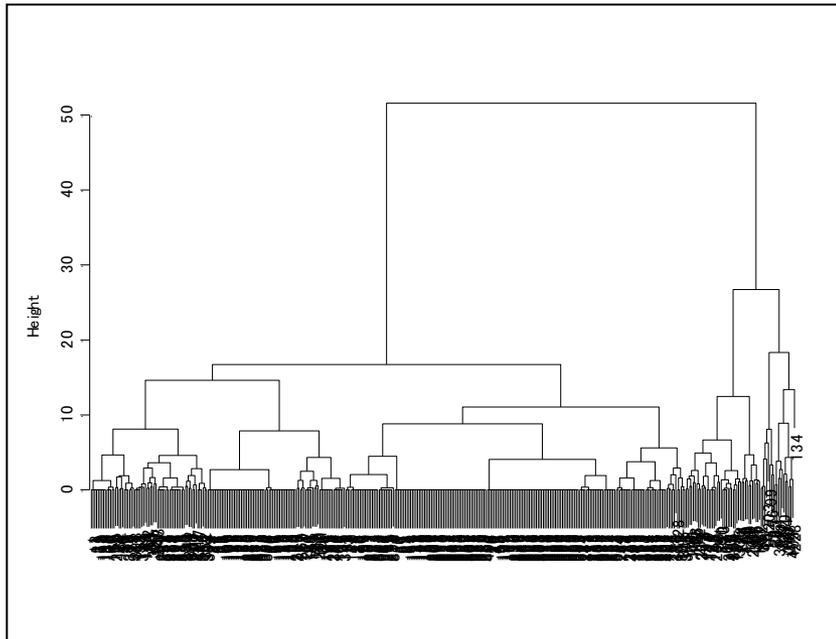


図1: クラスタードエンドログラム

分析方法 階層クラスタ分析

変数 接触ブランド数  
価格帯のばらつき  
最も接触している価格帯

凝縮型階層クラスタ分析を行い、クラスタリングを行う。

距離にはユークリッド平方距離、手法にはWard法を用いた。

デンドログラムより5クラスターを採用

## 6. クラスタリング(2/2)

### クラスター分布

設定した各クラスターの合計ショッパー人数, 平均接触ブランド数, 価格帯のばらつきの平均, 各価格帯の人数分布を以下に示す.

表1: クラスタ分布

	合計数	ブランド数	ばらつき	低価格帯	中価格帯	高価格帯
Class1	195	1.282	0.840	91	103	1
Class2	146	1.527	1.612	15	36	95
Class3	8	5.250	9.323	3	3	2
Class4	43	3.233	4.166	6	24	13
Class5	12	9.250	4.974	6	4	2

————→ 表1よりクラスターの定義づけを行う【商品接触パターン】

Class1	低～中価格帯の単ブランド内で選好する群
Class2	中～高価格帯の単ブランド内で選好する群
Class3	価格帯に全くこだわらず選好する群
Class4	価格帯にこだわらず, 平均3ブランドで選好する群
Class5	価格帯・ブランドに全くこだわらず選好する群

# 7. クラスタープロファイリング

各クラスターの購買行動パターンを把握する

表2: クラスタープロファイリング結果

	購入確率	検討時間	接触点数	購入点数
Class1	0.534	102.747	3.288	1.256
Class2	0.333	458.083	16.833	1.750
Class3	0.628	231.256	9.279	1.481
Class4	0.538	60.990	1.764	1.086
Class5	0.750	605.750	21.500	1.333

## 評価データ詳細

### ①購入確率

各クラスター内で購入に至った人数の割合

### ②平均検討時間

1ショッパーが棚前に滞在していた時間の平均

### ③平均接触商品数

1ショッパーが接触した商品数の平均

### ④平均購入商品数

購入に至ったショッパーが購入した商品数の平均

## ◎各クラスターの購買行動パターン

Class1

数点の商品に接触し、5割の人が購入している

Class2

長時間悩むも購入に至る確率は低いが、購入する際は複数購入している

Class3

複数の商品を手に取り、ある程度購入している

Class4

立ち寄っただけで買わない、もしくは決め買いですぐに立ち去っている

Class5

長時間複数の商品を検討し、いずれかの商品を購入する

## 8. 商品購入要因分析(1/2)

クラスターごと購入へ至るために影響している要因を明らかにすることで、購入へ導くマーケティングアプローチのヒントを得る

分析方法 二値ロジスティック回帰分析(パラメータ推定:最尤法)

$$\log \frac{p}{1-p} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon \quad (\text{式1})$$

$$P(Y = 1) = p, P(Y = 0) = 1 - p$$

Y=1 : 購入

Y=0 : 非購入

$\beta_1$  : 検討時間(1ショッパーが棚前に滞在していた時間)

$\beta_2$  : 接触点数(1ショッパーが接触した商品数)

$\beta_3$  : 商品接触までの時間

(ショッパーが棚前に訪れてから、1商品目に接触するまでの時間)

※分析には、関数glm()を用いた。P値推定がS-PLUSでは行えなかったため、P値推定のみ、統計ソフトRを用いて行った。

## 8. 商品購入要因分析(2/2)

### ▼分析結果

表3: ロジスティック回帰分析結果

	Class1	Class2	Class3
切片	-0.499 (0.436)	-0.769 (0.749)	0.525 (0.579)
検討時間	<b>-0.013 (0.006)**</b>	-0.002 (0.001)	-0.011 (0.008)
接触点数	<b>0.355 (0.189)*</b>	0.060 (0.053)	-0.064 (0.327)
接触までの時間	0.018 (0.014)	<b>0.027 (0.014)*</b>	0.003 (0.012)
	Class4	Class5	
切片	0.106 (0.490)	0.955 (0.606)	
検討時間	-0.002 (0.002)	<b>-0.036 (0.015)**</b>	表内のデータは、 偏回帰係数(標準誤差)を示し、 '*':10%有意, '**':5%有意とする。
接触点数	0.168 (0.104)	0.339 (0.393)	
接触までの時間	-0.003 (0.006)	<b>0.039 (0.021)*</b>	

Class1⇒検討時間が短く、接触商品数が多いほど購入確率は上がる

Class2⇒商品接触までの時間が長いほど購入確率は上がる

Class5⇒検討時間が短く、商品接触までの時間が長いほど購入確率は上がる

Class3とClass4については、購入に影響する要因の抽出はされなかった

## 9. マーケティングアプローチの提案(1/3)

インサイトを推察・マーケティングアプローチを提案する

▼分析結果を利用したアプローチ提案方法



この方法に基づき例としてClass2とClass5をターゲットとしたときのマーケティングアプローチの提案を行う

## 9. マーケティングアプローチの提案(2/3)

### ▼Class2へのアプローチ

—Shopper Insight—  
高価格ブランド目当て  
現物を確かめたい

接触商品パターン

中～高価格帯の単ブランド内で選好

購買行動パターン

長時間悩むも購入に至る確率は低い

—Approach Hint—  
棚から得られる情報が  
多いほど購入する

購入要因

商品接触までの時間が長い

価格が高いためすぐに決断できず、購入する人も少ないが、**棚から得られる情報が多い＝棚が多く情報を発信している**と購入確率が上がる。

Approach

デジタルサイネージやPOPなどを設置し、商品を手にする前に棚全体からブランドの特徴や品質を伝える。

検討時間は長いため、ビューティーアドバイザーを配置し、具体的な使用方法などの相談に乗ることで購入意欲を高める。

## 9. マーケティングアプローチの提案(3/3)

### ▼Class5へのアプローチ

—Shopper Insight—  
実際に手に取って  
いいものを探したい

接触商品パターン

価格帯・ブランドに全くこだわらず選好

購買行動パターン

長時間複数検討→購入

—Approach Hint—  
棚前情報で誘導し  
素早く惹きつける

購入要因

検討時間が短く、接触までの時間が長い

事前情報は特になく、売り場から考慮集合を形成していると考えられるため、店舗から与えた情報で考慮集合を誘導しやすい。

検討時間に余裕を持っているショッパーであるが、時間をかけすぎることなく商品を理解してもらうことが重要。

Approach

一押し商品を強調したディスプレイや巨大パネルの設置などを行うことで、購入対象商品へ誘導する。

また、テスターを配置することで、商品の良さをすぐに実感してもらう。

# 10. 終わりに

トランザクション型ショッパー行動データを利用して  
ショッパーマーケティングへ有効な示唆を行うことができた

## ▼トランザクションデータ処理

トランザクション型ショッパー行動データを利用することで、複雑な商品接触パターンを捉えるためのデータを作成することができた。

## ▼ショッパーマーケティングへの活用

作成した商品接触データを用いてクラスター化することで、商品接触パターンを捉えたクラスタリングを行うことができた。またクラスターの購買行動パターンや購入要因を分析することで、有効なマーケティングアプローチの示唆へつなげる方法を提案することができた。

## ▼今後の課題

今回はClass3とClass4に対する購入要因の特定ができず、モデルや使用する変数にまだ検討の余地があると考えられる。

また今後はより明確にインサイトを推察するため、他の指標や方法を取り入れ、最適なクラスタリング手法を追及していきたい。

# 参考文献・謝辞

## 【参考文献】

- 財団法人流通経済研究所, ショッパー・マーケティング, 日本経済新聞出版社, 2011
- 財団法人流通経済研究所, インストア・マーチャンダイジング, 日本経済新聞出版社, 2011
- 平久保仲人, 消費者行動論—なぜ、消費者はAではなくBを選ぶのか?, ダイヤモンド社, 2005

## 【謝辞】

本研究は株式会社Mideeより提供していただいたデータを使用したものである。また、株式会社アイズファクトリー片所強氏より分析等に関して手厚いご指導をいただいた。

最後に両者のご協力に、心から御礼申し上げ、謝辞と代えさせていただきます。