

# 季節性商品への嗜好を考慮した顧客クラスタリング手法に関する研究

1XC12088-9 仁ノ平 将人  
指導教員 後藤 正幸

## 1 研究背景・目的

近年、多くのスーパーマーケットにおいて、ICカードポイント付与システムを通じて得られた購買履歴データを解析し、各顧客がどのような商品（以下、アイテム）を購入し易いかといった購買傾向を把握することが可能となっている。実際に、多くのスーパーマーケットは様々なマーケティング施策の立案のために、購買履歴データを解析して、顧客が嗜好するアイテムの特徴の把握を試みている。顧客の嗜好するアイテムを把握する手法として、顧客の購買行動は潜在的嗜好に従うと仮定した潜在クラスモデルを用いた分析方法が多く提案されている。特にその中の1つである Aspect Model[1] は顧客のアイテムへの嗜好を適切に表現できるとして一定の評価を得て幅広く適用されている。

一方、スーパーマーケットで販売されている様々なアイテムの購買額に着目すると、農産・水産品といった生鮮食品の売上が高い割合を占めているという特徴がある。一般に、生鮮食品は他の商品に比べ、特定の期間に多く購買されている旬のあるアイテム（以下、季節性アイテム）と1年間を通じ安定して購買されている旬のないアイテム（以下、非季節性アイテム）が混在していることが知られている。

また、顧客ごとの購買アイテムに着目すると、季節性アイテムを好んで購買する顧客や、年間を通じ非季節性アイテムを中心に購買する顧客といったように、顧客ごとに季節性・非季節性アイテムへの嗜好は異なることが考えられる。よって、顧客別の1年間を通じたマーケティング施策を考える上で、季節性・非季節アイテムへの嗜好特性という観点から顧客をクラスタリングし、ターゲットを明確にすると共に、クラスタ別に施策を検討することが有効である。しかし、従来の Aspect Model では顧客の購買アイテムの季節性を考慮していないため、顧客の季節性・非季節アイテムへの嗜好を考慮した分析には適さない。

そこで本研究では、季節性・非季節性アイテムへの嗜好という観点から購買行動の分析を可能にする潜在クラスモデルを提案する。また、実際のスーパーマーケットにおける1年間の購買履歴データに提案モデルを適用し、分析結果を示す。

## 2 準備

### 2.1 事前分析

本研究では、経営科学系研究部会連合協議会主催、平成27年度データ解析コンペティションで提供された、株式会社アイディーズのi-codeデータサービスによる、某スーパーマーケットの2014年1年間分の購買履歴データを分析対象とする。当該データの事前分析により以下の三点の特徴が確認された。まず、アイテムごとの売上額に着目すると、農産・水産品アイテムの売上額が大きい割合を占めている。また、各月ごとに購買額の大きいアイテムを集計すると、1年間を通じ安定して購買されているアイテム（非季節性アイテム）と特定の月に多く購買されるアイテム（季節性アイテム）が混在していることが確認された。さらに、顧客ごとの購買アイテムに着目すると、季節性アイテムを好む顧客と季節性アイテムを好まない顧客が存在することも示唆された。

### 2.2 潜在クラスモデル

顧客の購買行動はアイテムへの潜在的嗜好に従うとして顧客とアイテムの間に潜在クラスを仮定し、購買傾向をモデル化する様々な潜在クラスモデルがマーケティングモデルとして提案されている。その1つである Aspect Model[1]

は、大量の購買履歴データに対しても、容易に適用可能であり、各顧客のアイテムへの嗜好を適切に表現出来るとして、多くの事例に適用されている。このモデルは似た購買傾向を持つ顧客とこれらの顧客に購買され易いアイテムは同じ潜在クラスに属することを仮定している。また、このモデルでは顧客とアイテムが複数の異なる潜在クラスに所属することを許容するという特徴を有する。いま、 $I$ 人からなる顧客集合を  $\mathcal{U} = \{u_i : 1 \leq i \leq I\}$ 、 $N$ 個のアイテム集合を  $\mathcal{C} = \{c_n : 1 \leq n \leq N\}$ 、 $K$ 個の潜在クラス集合を  $\mathcal{Z} = \{z_k : 1 \leq k \leq K\}$  とする。このとき、顧客  $u_i$  がアイテム  $c_n$  を購入する事象  $(u_i, c_n)$  の確率は以下の式 (1) で表現される。

$$P(u_i, c_n) = \sum_{k=1}^K P(z_k)P(u_i|z_k)P(c_n|z_k) \quad (1)$$

式 (1) のパラメータ  $P(z_k)$ 、 $P(u_i|z_k)$ 、 $P(c_n|z_k)$  は EM アルゴリズム [2] により推定することができる。Aspect Model では顧客とアイテムの共起関係を特徴量としている。しかし、その購買アイテムがどのような季節性を持つかを特徴量としていないため、季節性アイテムへの嗜好を考慮したモデルを得ることはできない。

## 3 提案モデル

### 3.1 提案モデルの概要

本研究では顧客の季節性・非季節性アイテムへの嗜好の差異を表現するための潜在クラスモデルを提案する。

その際に、調査対象期間を複数の期間に細分化し、アイテムの各期間の売上数量を要素としたベクトルでアイテムの特徴量を表現する。これにより、顧客がアイテムを購入する事象は、顧客と購買アイテムの売上特徴ベクトルとの共起と捉えられる。さらに、この共起を事象とし、所属するアイテムの各期間の売上数量を要素とするベクトルの生起確率に対して潜在クラスごとに多項分布を仮定した潜在モデルを定式化する。

### 3.2 提案モデルの定式化

顧客が1アイテムを購入する事象を1購買データとし、全解析対象データ数を  $L$  とする。また、 $L$  個の購買データで出現する顧客集合を  $\mathcal{X} = \{x_l : 1 \leq l \leq L\}$  とする ( $x_l \in \mathcal{U}$ )。ここで、顧客  $x_l$  の購買データのアイテムの特徴量を、調査対象期間を  $M$  期に等分した内の任意の  $m$  期目 ( $1 \leq m \leq M$ ) の売上個数  $N_{lm}$  を要素とした  $M$  次元ベクトル  $\mathbf{y}_l = (N_{l1}, \dots, N_{lm}, \dots, N_{lM})$  で表すことにする。さらに、 $K$  個の潜在クラス集合を  $\mathcal{Z} = \{z_k : 1 \leq k \leq K\}$  とし、各潜在クラス  $z_k$  ごとに、 $m$  期目の売上数量の割合  $\theta_{km}$  を要素とした多項分布のパラメータを  $\boldsymbol{\theta}_k = (\theta_{k1}, \dots, \theta_{km}, \dots, \theta_{kM})$  と定義する ( $\sum_{m=1}^M \theta_{km} = 1$ )。このとき、顧客  $x_l$  と  $M$  次元ベクトル  $\mathbf{y}_l$  で表される  $l$  番目の購買データの共起確率  $P(x_l, \mathbf{y}_l)$  は以下の式 (2) で表現される。

$$P(x_l, \mathbf{y}_l) = \sum_{k=1}^K P(z_k)P(x_l|z_k)P(\mathbf{y}_l|z_k) \\ = \sum_{k=1}^K P(z_k)P(x_l|z_k) \frac{\left( \sum_{m=1}^M N_{lm}! \right)}{\prod_{m=1}^M N_{lm}!} \prod_{m=1}^M \theta_{km}^{N_{lm}} \quad (2)$$

### 3.3 EM アルゴリズムによるパラメータ推定

式 (2) に対する対数尤度関数  $LL$  は次式で与えられる.

$$LL = \sum_{l=1}^L \log P(x_l, \mathbf{y}_l) \quad (3)$$

式 (3) を最大化するパラメータ  $P(z_k)$ ,  $P(u_i|z_k)$ ,  $\theta_{km}$  は、以下の EM アルゴリズムを用いて推定することができる.

[E-step]

$$P(z_k|x_l, \mathbf{y}_l) = \frac{P(z_k)P(x_l|z_k)P(\mathbf{y}_l|z_k)}{\sum_{k=1}^K P(z_k)P(x_l|z_k)P(\mathbf{y}_l|z_k)} \quad (4)$$

[M-step]

$$P(z_k) = \frac{\sum_{l=1}^L P(z_k)P(x_l|z_k)P(\mathbf{y}_l|z_k)}{\sum_{l=1}^L \sum_{k=1}^K P(z_k)P(x_l|z_k)P(\mathbf{y}_l|z_k)} \quad (5)$$

$$P(u_i|z_k) = \frac{\sum_{l=1}^L P(z_k)P(x_l|z_k)P(\mathbf{y}_l|z_k)\delta(x_l = u_i)}{\sum_{l=1}^L \sum_{i=1}^I P(z_k)P(x_l|z_k)P(\mathbf{y}_l|z_k)\delta(x_l = u_i)} \quad (6)$$

$$\theta_{km} = \frac{\sum_{l=1}^L P(z_k)P(x_l|z_k)P(\mathbf{y}_l|z_k)N_{lm}}{\sum_{l=1}^L \sum_{m=1}^M P(z_k)P(x_l|z_k)P(\mathbf{y}_l|z_k)N_{lm}} \quad (7)$$

ただし、 $\delta(x_l = u_i)$  は  $x_l = u_i$  ならば 1、それ以外は 0 をとるインジケータ関数である. E-step, M-step を式 (3) が収束するまで繰り返し、パラメータ推定を行う.

## 4 解析

### 4.1 解析データ

提案モデルの有効性を検討するため、2.1 節の事前分析で用いた某スーパーマーケット 1 店舗における 2014 年 1 月から 2014 年 12 月の 1 年間の購入履歴データを提案モデルを用いて解析する. 本解析では、アイテム間で旬に大きく差異があると予想される農産・水産カテゴリの 323 アイテムの購入履歴データに着目して解析を行う. 顧客数は  $I = 27,184$ 、データ数は  $L = 2,256,271$  であり、事前分析により、潜在クラス数  $K = 4$  と設定した. また、 $M = 12$  とし、 $N_{lm}$  を  $l$  番目のデータで購買されたアイテムの  $m$  月における売上個数とする.

### 4.2 解析結果と考察

解析の結果として、各潜在クラス  $z_k$  ごとに所属アイテムの季節性を表す期間  $m$  の売上個数の割合  $\theta_{km}$  を要素とした多項分布のパラメータ  $\theta_k = (\theta_{k1}, \dots, \theta_{km}, \dots, \theta_{k12})$  を図 1、各潜在クラスの生起確率  $P(z_k)$  を表 1 に示す.

表 1: 各潜在クラスの生起確率 (%)

$k$	1	2	3	4
$P(z_k)$	1.41	8.22	3.42	86.95

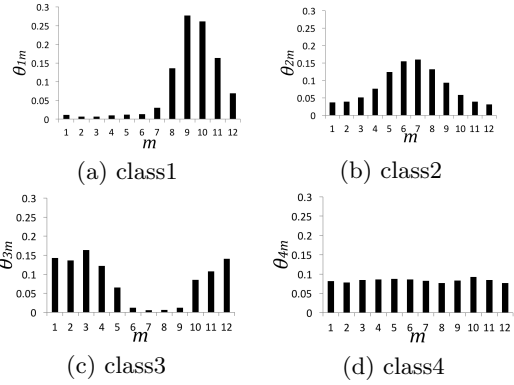


図 1: 各潜在クラスの  $\theta_k$

図 1 より、class1 には 8 月～11 月に旬を持つ季節性アイテムを嗜好する顧客、class4 には旬を持たない非季節性アイテムを嗜好する顧客といったように、提案モデルにより季節性・非季節性アイテムへの嗜好の特徴を捉えた潜在クラスを求めることができていたことが確認された.

ここで、季節性・非季節性アイテムへの嗜好により 1 会計あたりの購買金額の期待値 (以下、1 会計購買額期待値) に差異があるか確かめるため、 $P(z_k|u_i) > \alpha$  を満たす顧客の 1 会計購買額期待値について、 $\alpha$  を変化させたときの挙動を図 2 に示す ( $0 \leq \alpha \leq 1$ ). なお、1 会計購買額期待値は、 $E(u_i)$  を顧客  $u_i$  の農産・水産品以外のアイテムも含む 1 会計期待購買額とし、 $\sum_{i=1}^I P(u_i|z_k)E(u_i)$  で求めた.

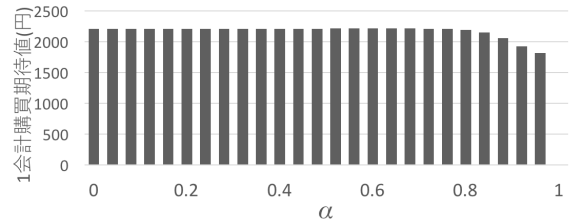


図 2:  $\alpha$  と 1 会計購買額期待値の関係

図 2 を見ると、 $\alpha$  が 0.8 より大きいときに、1 会計購買額期待値が減少していることが分かる. これは、class4 への所属確率が高い顧客、つまり季節性アイテムを好まない顧客は 1 会計購買額期待値が小さいと解釈できる. これよりスーパーマーケットは、このような顧客に季節性アイテムを購買させるようなアプローチを取ることで売上向上が期待出来ると考えられる.

## 5 まとめと今後の課題

本研究では顧客のアイテムの季節性への嗜好を表現するための新たな潜在クラスモデルを提案し、実際の購買履歴データを用いて提案モデルの有効性を示した. また解析の結果より季節性のアイテムを好む顧客の方が非季節性を好む顧客に比べ、1 会計あたりの期待購買額が大きいことが明らかになった. 今後の課題として、各クラスに属する顧客のさらなる分析や、本提案モデルの適用により得られた知見を用いた施策の考案などが挙げられる.

### 参考文献

- [1] T.Hofmann, "Probabilistic Latent Semantic Analysis," *Proc. of UAI'99*, pp.289-296, 1999.
- [2] 宮川雅巳, "EM アルゴリズムとその周辺," 応用統計学, Vol.16, No.1, pp.1-21, 1987.