

潜在クラスモデルによる流通量販店舗の 来店人数予測の精度改善の評価

Improvement of Prediction Accuracy of the Number of Customers by Latent Class Model

石垣司*1*2 竹中毅*2 本村陽一*2
Tsukasa Ishigaki Takeshi Takenaka Yoichi Motomura

*1東北大学大学院 経済学研究科

Graduate School of Economics and Management, Tohoku University

*2産業技術総合研究所 サービス工学研究センター

Center for Service Research, National Institute of Advanced Industrial Science and Technology

The present paper describes an improvement of prediction accuracy of the number of customers in retail service. We focus on the customers' lifestyle in order to understand of their behaviors. Here, the lifestyles of customers are given by a multilayer latent class model using large scale ID-POS data and customer questionnaire data. For the prediction of the number of customers in daily stores, we employ a linear regression model based on the various situations such as weather or events. The models are applied to the number of all customers' prediction and the number of a part of customers characterized their lifestyles. The result shows that the lifestyle- characterized one is superior to the all customer's one.

1. はじめに

適切な在庫管理による廃棄率の低下や過剰在庫の抑制は小売業や外食産業の生産性に大きな影響を与える。スーパーマーケットなどの流通量販店では全体の来店人数を予測し、その予測結果に基づいて商品の発注や従業員のシフト管理を行っている。そのため、来店人数の予測は日常の業務において重要な意味を持つ。しかしながら流通量販店における需要予測は、前年度の同月同曜日における売上との比較や、各店舗の店長の経験と勘により行われている現状がある [竹中 11]。

現在では生活者の価値観やライフスタイルが多様化し、時間・状況・気分などの各個人の動的な変数によりその行動が変化する一人十色の時代へと突入している。そのため、単なる顧客の欲求充足の視点を越えた生活者起点の発想がサービス業には求められている。ライフスタイルの異なる消費者が異なる状況においてどのように異なる行動を起こすのかを知ることは生活者を理解するうえで重要な事象である [本村 09, 中村 09]。

そこで本報告では、流通量販店における顧客ライフスタイルを考慮した来店人数予測とその精度評価を行う。ここでは、ID-POS データと顧客アンケートデータを利用した多層潜在クラスモデルによる顧客のライフスタイル分類を行い、全顧客に対する来店予測精度とライフスタイル別に行った来店予測精度を比較する。

2. 来店人数予測ベースモデル

本報告における来店人数の予測には各日の状況（曜日、祝日、雨量、気温、イベントなど）を説明変数とし、各日の来店人数を目的変数とする線形回帰モデルにより予測を行う。このモデルをベースモデルと呼び、詳細は [竹中 11] に記されている。ここでは説明変数として、降水量 10mm 以上、降水量 30mm 以上、曜日（月曜日を基準変数）、祝日、連休最終日、飛び連休の最終日、1月1日、1月2日、1月3日、年末3日間、ク

リスマスイブ、気温（6段階、1段階を基準変数）を扱い、全て2値のダミー変数として利用する。また、予測モデルにはステップワイズ法による変数選択を行い、採用された変数を来店人数予測に使用する。

3. 顧客ライフスタイル分類

3.1 ID-POS データと顧客ライフスタイルアンケート

本章では多層潜在クラスモデルを用いた顧客ライフスタイルの分類について述べる。

ここでは、兵庫県を事業エリアに約150店舗を展開する流通量販店で記録された2008年10月1日から2009年9月30日の期間におけるID-POSデータを利用する。全データ中のID観測可能な割合は約87%である。また、同流通量販店において、顧客のライフスタイルやパーソナリティを把握するためのアンケート調査を実施した。2009年12月に同流通量販店の会員約17,000人に対しダイレクトメールによりアンケートを送付し、その内3,965名から回答を得た。アンケート回答顧客とその購買履歴は対照することが可能となっている。詳細は [石垣 10] に記す。アンケート分析の結果、各顧客は以下の6タイプのライフスタイルに分類できることが分かった。

- 1 こだわり消費派：高くても健康に良いものを選び、産地への関心、こだわりのブランドがある
- 2 家庭生活充実派：料理が好きで食事も生活も充実している。気分も安定している
- 3 アクティブ消費派：外向的で、新商品や話題の商品は試しに買ってみる。ただ無駄遣いは多い
- 4 節約消費派：チラシを見てお得な商品を買う。安ければ少々遠い店にも行く。高い商品は買わない
- 5 堅実生活派：几帳面で家計簿をつけ、無駄遣いはしない。毎日の献立はスーパーに行く前に決める
- 6 パパッと消費派：スーパーでの買い物はできるだけ早くすませたい。お弁当を作ることがある

連絡先： 産業技術総合研究所 サービス工学研究センター，〒135-0064 東京都江東区青海 2-3-26, ishigaki-tsukasa@aist.go.jp

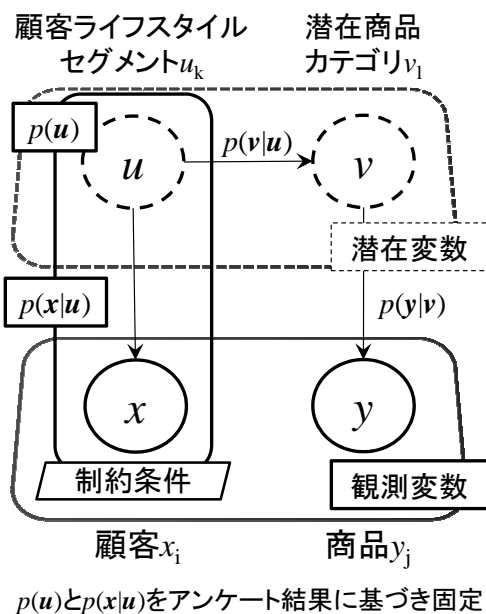


図 1: 多層潜在クラスモデル

アンケート回答率は約 23%であり、7 割強の顧客のライフスタイルを特定することはできない。そのため、ここでは多層潜在クラスモデルを利用して、非アンケート回答顧客のライフスタイルを推定する。

3.2 多層潜在クラスモデル

X 人の顧客と Y 個の商品を対象とし、顧客 i と商品 j を表す変数をそれぞれ $x_i (i = 1, \dots, X)$ と $y_j (j = 1, \dots, Y)$ とする。また、潜在顧客セグメント数を U 、潜在商品カテゴリ数を V とし潜在顧客カテゴリ k と潜在商品カテゴリ l を表す変数をそれぞれ $u_k (k = 1, \dots, U)$ と $v_l (l = 1, \dots, V)$ とする。顧客 i の商品 j の購買数を N_{ij} とする。ここでは顧客、商品、潜在カテゴリ間の関係を

$$p(x_i, y_j, u_k, v_l) = p(u_k)p(x_i|u_k)p(v_l|u_k)p(y_j|v_l) \quad (1)$$

としてモデル化する。その関係を図 1 に示す。このモデルでは EM アルゴリズムを用いて各確率値を最尤推定することができる。またここでは、 $X = 3965$ (アンケート回答顧客)、 $Y = 1000$ (売上個数上位 1000 商品)、 $U = 6$ (アンケート因子分析に基づく)、 $V = 12$ (赤池情報量基準に基づく) とした。その詳細は [Ishigaki 10, 石垣 10] に記す。

3.3 非アンケート回答顧客のライフスタイル推定

3.3.1 商品のためのライフスタイル購買傾向の付与

前節で推定した各確率値より、各商品へのライフスタイルによる購買傾向の違いを付与する。 $p(u_k)$ と $p(v_l|u_k)$ は前節で推定されているので、各商品が与えられたときのライフスタイル傾向は

$$p(u_k|y_j) = \alpha p(u_k)p(y_j|u_k) \quad (2)$$

として求めることができる。ただし、 α は正規化項とする。

3.3.2 非アンケート回答顧客のライフスタイル推定

ID 情報のある非アンケート回答顧客は商品の購買履歴 N_{ij} のみが存在する。購買履歴を用いて非アンケート回答顧客のライフスタイルを推定する。ここでは、 $p(u_k|y_j)$ の加重平均を

用いる。顧客 i の k 番目のライフスタイル確率を $p(u_{ik})$ とし、その推定値を

$$p(u_{ik}) = \beta \sum_j^Y N_{ij} p(u_k|y_j) \quad (3)$$

として定義する。ただし、 β は正規化項とする。これにより、非アンケート回答顧客に対しても、その購買履歴から各々のライフスタイルが推定可能となる。

4. 来店人数予測と評価

4.1 来店人数の予測

ここでは、以下の 3 手法により来店人数の予測を行った。

1. 前年同月同曜日予測：前年の同月同曜日のデータを予測値として使用 (現状の予測方法)
2. ベースモデル予測：2 章に示した曜日、イベントなどを説明変数とした線形回帰モデル
3. ライフスタイル毎予測：多層潜在クラスモデルに基づき顧客ライフスタイルを分類した顧客層ごとにベースモデルを構築

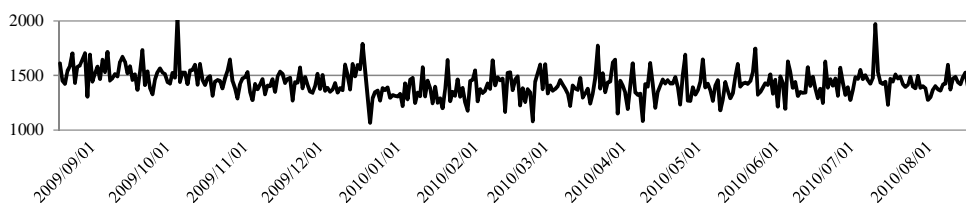
使用したデータは以下の通りである。訓練データ：2009 年 9 月 1 日火曜日から 2010 年 8 月 31 日火曜日の 12 カ月分。テストデータ：2010 年 9 月 1 日水曜日から 2010 年 9 月 30 日木曜日の 1 カ月分。前年同月同曜日データ：2009 年 9 月 2 日水曜日から 2010 年 10 月 1 日木曜日。ここでは、同流通量販店が経営する神戸西宮地区のある店舗 (A 店) における来店者数の予測を行う。この店舗は他の店舗と比べ、ベースモデル予測による再現結果と実測値の相関係数が低かった。つまり、線形予測での予測が比較的難しい店舗である。図 2 に A 店の約 1 年間の全来店人数の推移と、ライフスタイル別顧客の来店人数の推移を示す。

4.2 来店人数予測の評価

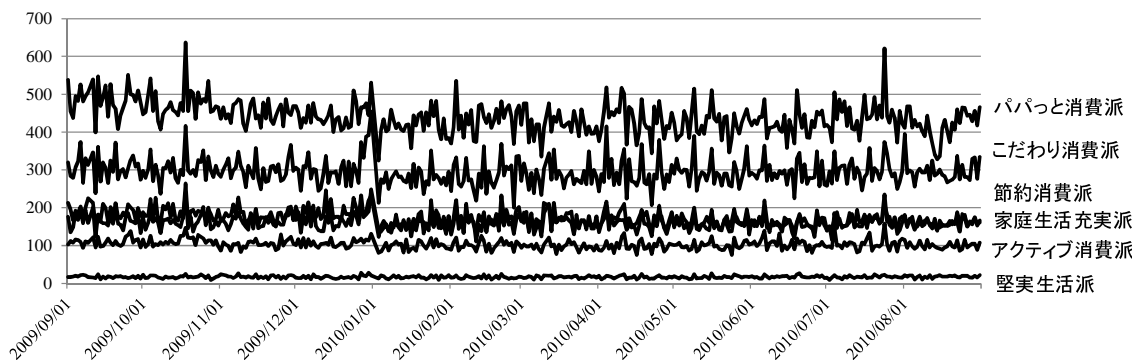
ここでは、予測値とテストデータの実数値との残差を予測精度の指標とする。前年同月同曜日予測では、前年同月同曜日のデータを予測値として使用し、テストデータとの残差を見る。ベースモデル予測ではベースモデルに基づく予測結果を予測値として使用し、テストデータとの残差を見る。ライフスタイル毎予測では各ライフスタイル毎にベースモデルを作成し、その予測結果の合計値を予測値として、テストデータとの残差を見る。また、ここでは、残差の合計と絶対残差の合計の両方で評価する。その結果を表 1 に、日ごとの残差を図 3 に、日ごとの絶対残差の合計を図 4 にそれぞれ示す。残差の合計、絶対残差の合計の両方で顧客ライフスタイル毎の予測結果が最も良い結果を示している。

4.3 変数選択の結果

表 2 にステップワイズ法による変数選択の結果を示す。が正の係数を持つ要素、が負の係数を持つ要素、-が選択されなかった係数を示す。また、表中の L1 から L6 はそれぞれ、こだわり消費派、家庭生活充実派、アクティブ消費派、節約消費派、堅実消費派、パパッと消費派のライフスタイル分類を示す。表より、各ライフスタイルの来店人数について影響を与えている変数を見ることができる。例えば、雨、1 月 3 日、気温 1 (低温) の日は、全顧客について来店人数が下がることが読み取れる。また、火曜日、水曜日、金曜日、日曜日は月曜日と



(a) A店の全来店人数



(b) A店の顧客ライフスタイルセグメント毎の来店人数

図 2: A 店舗の来店人数

表 1: 予測残差の合計

	1. 前年同月同曜日予測	2. 線形予測	3. ライフスタイル毎予測	1vs3 の改善率	2vs3 の改善率
残差合計	-2835 人	-762 人	-516 人	81.8%	32.2%
絶対残差合計	3285 人	1599 人	1542 人	53.1%	3.6%

手法 B の手法 A からの改善率=1-(手法 A の残差合計÷手法 B の残差合計)

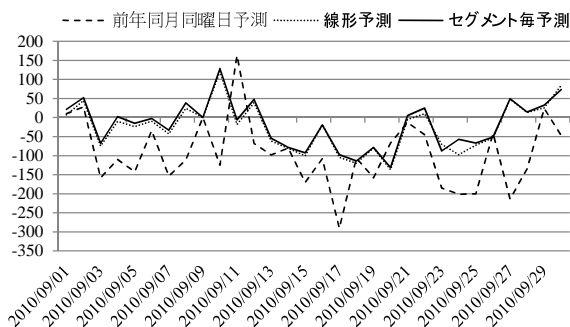


図 3: 予測残差

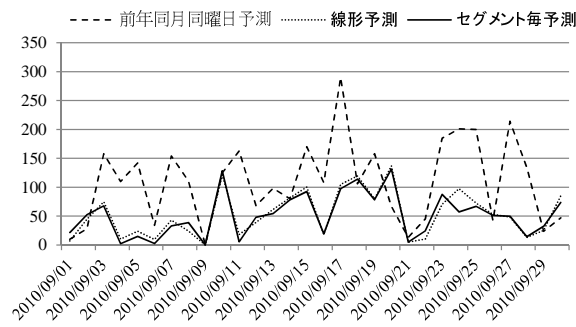


図 4: 絶対予測残差

比べて来店人数が増えることがわかる。また、各ライフスタイルにのみ影響を与える変数も読み取れる。例えば、パパッと消費派のみ祝日や連休最終日に来店人数が増え、気温が高いと来店人数が減少することが読み取れる。このように、全ライフスタイルによって共通な変数と各ライフスタイル特有の変数が存在することが分かる。

5. 考察

線形回帰モデルを利用した予測モデルでは、理想的には予測残差の分布はガウス分布となる。そのため、ベースモデルの予測残差と各ライフスタイル毎の予測残差が線形モデルで十分に

説明可能な状態である場合、ガウス分布の再生性より、ベースモデルの予測残差の分散とライフスタイル毎予測を合計した予測残差の分散は等しくなる。しかしながら、本報告における予測精度と変数選択の結果では各ライフスタイルにおける行動の特徴が抽出されているため、顧客ライフスタイルは流通量販店への来店行動に何らかの影響を与えていることが読み取れる。そのため、予測に対する変数の影響をモデル化することで、何故その予測が当たらないのかという「予測の外れ方」を検討することができる。加えて、各状況による外れ具合を階層的にモデリングすることでより詳細な来店人数予測を実行できる可能性がある [本村 11]。

表 2: 変数選択の結果

状況	L1	L2	L3	L4	L5	L6
雨 10						
雨 30						
火						
水						
木	-	-		-	-	-
金						
土		-			-	-
日						
祝	-	-	-	-	-	
連最終	-	-	-	-	-	
連中日		-	-	-	-	-
正月 1	-	-	-	-	-	-
正月 2	-	-	-	-	-	-
正月 3						
年末						
イブ					-	
気温 1						
気温 2	-		-	-	-	-
気温 3	-	-	-	-	-	-
気温 4					-	
気温 5	-	-	-			

[中村 09] 中村, 寺本, 矢野: “顧客視点の商品マスター (商品 DNA) の可能性”, 流通情報, No.477, pp.22-33

[竹中 11] 竹中, 石垣, 本村: “生活者行動に着目したサービス需要予測技術の検討”, 2011 年度人工知能学会全国大会 (第 25 回) 論文集, 2011

[山口 04] 山口, 土屋, 樋口: “状態空間モデルを用いた飲食店売上の要因分解”, オペレーションズ・リサーチ, Vol.49, No.5, pp.316-324, 2004

6. むすび

本発表では, 多層潜在クラスモデルによる顧客ライフスタイルの推定と, そのライフスタイル別の来店人数予測モデルについて述べた. 本手法を用いた店舗支援システムの構築は現在進行中の課題である. また, [山口 04] に示されるようなベースモデル自体の改善についても今後の課題とする.

謝辞

本報告で用いた ID-POS データと顧客アンケートデータは生活協同組合コープこうべから提供を受けた. ここに感謝の意を表す. また, 本研究は経済産業省委託費事業「IT とサービスの融合による新市場創出促進事業 (サービス工学研究開発事業)」の支援を受けている.

参考文献

[石垣 10] 石垣, 竹中, 本村: “大規模 ID-POS データ活用のための顧客の計算モデル化 ~ 小売サービスにおける生産性向上の試み”, 第 80 回人工知能学会人工知能基本問題研究会, pp. 15-18, 2010

[Ishigaki 10] T. Ishigaki, T. Takenaka and Y. Motomura: “Category Mining by Heterogeneous Data Fusion Using PdLSI Model in a Retail Service”, Proc. IEEE ICDM 2010, pp. 857-862, 2010

[本村 09] 本村, 石垣: “サービス工学における計算論的モデル”, システム/制御/情報, Vol. 53, No.9, pp.167-172, 2009

[本村 11] 本村, 竹中, 石垣: “条件付層別差分モデルによる需要予測の高精度化”, 2011 年度人工知能学会全国大会 (第 25 回) 論文集, 2011