

時空間を基軸とした百貨店ID付POSデータの分析

01506390 東京理科大学 *降 簾 徹 馬 FURIHATA Tetsuma
東京理科大学 西 野 和 美 NISHINO Kazumi
01308630 東京理科大学 寺 崎 康 博 TERASAKI Yasuhiro

1. はじめに

本稿では、ある百貨店のID付POSデータに対して、時間と空間という最も基本的な視点から売上の構造を調べるためにデータマイニングを行った結果を報告する。具体的には、顧客分布の時空間変動を把握する消費者密度推定手法を提示し、その変動を実際に把握すること、顧客の店舗内滞留時間と購買額の関係を明らかにすること、購買パターン分析から有力な購買パターンを抽出し、顧客分布および滞留時間と密接に関連していることを示すこと、といった分析結果を報告する。分析に用いた百貨店のID付POSデータは平成14年度データ解析コンペティションから提供されたもので、約1200万レコードの取引明細データ(期間は2001年1月~12月)である。

2. 顧客分布の時空間変動解析

小売店舗にとって、顧客がどこから来店しているのかという商圈の把握は、最も基本的で重要な分析である。近年のPOSシステムの発展は、こうした分析に必要な顧客情報の入手を容易にし、現状分析のみならば商圈モデルに頼らなくとも顧客の空間的な分布を推測できるようになってきている。ここでは、ノンパラメトリックな密度推定手法の一つであるカーネル密度推定[3]に基づいて顧客分布を直接的に推定する。特に、個々の地点ごとに異なる消費者の重み(購買額など)を考慮する拡張と、日、月、曜日ごとの空間変動を考慮する拡張を行う。

カーネル密度推定は、サンプルとして与えられる個々の地点にカーネルをおき、それを累積させて密度分布を推定するものである。密度分布の形状はパラメータ h のみに依存し、 h が大きいとなだらかな分布に、 h が小さいと凹凸の激しい分布となる。カーネルとしては正規分布が一般的に用いられる。サンプル $i(i=1, \dots, n)$ の座標を (x_i, y_i) 、その重みを w_i とすると、拡張したカーネルは

$$k_i(x, y, h_t) = \frac{w_i}{2\pi h_t^2} e^{-\frac{1}{2h_t^2}[(x-x_i)^2+(y-y_i)^2]} \quad (1)$$

と表現できる。この時点 t は日、月や曜日などの項目ごとに離散的にとらえることにする。このカーネルを重ねると、

$$k(x, y, h_t) = \frac{1}{n} \sum_{i=1}^n k_i(x, y, h_t) \quad (2)$$

という密度分布を得る。この密度分布はパラメータ h_t のみに依存するので、最尤法を用いたクロスバリデーションにより h_t を時点 t ごとに求め、分布を特定する。

百貨店のID付POSデータには顧客の居住位置を示す座標値(需要点)が含まれているため、各需要点ごとに購買額を合計して分析に用いることができる。時点 t は月ごと、曜日ごと、初回購買の時間帯ごとにそれぞれパラメータを推定した。パラメータ h_t の値は冬場と夏場に大きくなり、週末に小さくなる傾向を示した。初回購買の時間帯では日中と夕食時に小さくなり、17時台が最大になった。これら推定値を基に顧客の購買額密度分布を、月、曜日、時間帯ごとに動画として視覚化し、顧客の平均購買金額や平均距離(商圈の一尺度)と照合すると、冬場・夏場は中央部の密度が下り商圈が狭域になること、週末は商圈が狭域になるが中央部の密度が上がること、1日の営業時間帯内でも商圈が変動していること、平日と週末の商圈の違いと夕方以降の商圈拡大には通勤行動の影響が寄与している可能性が高いこと、などの知見を得た。

3. 顧客の店舗内滞留時間と購買額の関係

購買行動の種類では、購入予定にしがたい予定ブランドを購入する計画購買の割合は1割程度といわれており、これ以外は想起購買、関連購買、衝動購買などの非計画購買であるといわれている。計画購買は顧客の商品購買に対する意思決定が入店前であるのに対し、非計画購買は入店後である点に大きな違いがある。このため店舗側は店頭セールス・プロモーションによって非計画購買を高める経営努力を行っている。この非計画購買は顧客が店舗内で商品やサービスに接することによる情報伝達・普及効果の表れであり、店舗内での滞留時間と密接に関係していると考えられ、この滞留時間に関して考察することには十分な意義があると考えられる。

ID付POSデータには、購買時刻が記録されており、複数の売場で購買を行った場合には、最初と最後の購買時刻の差からおおよその店舗内滞留時間を把握することができる。平均滞留時間と平均購買額の関係を調

べると、滞留時間 100 分程度で平均購買額が最大になる性質を持っていた。この関係は性別、月別、曜日別などに区分して調べても、いずれも 100 分前後で購買額が最大となり、曲線の形状も類似していることが確認できた。このような関係は、滞留時間や購買額の差こそあれ、他の小売店舗でも観察されると想定される。この関係をモデル化すれば、ID 付 POS データが与えられなくとも少ないサンプルで推定することができる上、新たな店舗評価指標としての役割も期待できる。

そこで、この関係は Bass モデル [1] が表現する分布形に類似しているため、このモデルを参考に分析を行う。Bass モデルは、情報伝達理論、普及理論とも呼ばれ、新製品が外的影響（マスメディアの効果）と内的影響（口コミ）によって普及していく過程を表現したものである。ここでは、このモデルを店舗内滞留時間と購買額の関係にあてはめ考察する。モデルの基本型は、

$$n(t) = \frac{dN(t)}{dt} = p(m - N(t)) + \frac{q}{m} N(t)(m - N(t)) \quad (3)$$

であり、 $n(t)$ は購買額を、 $N(t)$ は累積的購買額を表す。 m は顧客の購買ポテンシャル（潜在的購買力）であり、 p は計画購買を中心とした影響を、 q は店舗内での普及効果の影響を表現しているものとして解釈する。特に、 q は計画購買によって来店した顧客が店舗内でのセールス・プロモーション等の情報伝達・普及効果により購買する部分を含んでいると考える。 $n(t)$ の解析的表現は、

$$n(t) = m \left[\frac{p(p+q)^2 e^{-(p+q)t}}{(p+q e^{-(p+q)t})^2} \right] \quad (4)$$

である。推定するパラメータは m 、 p 、 q の 3 つであるから、3 時点のデータさえあればパラメータを推定することができる。従来は、時点 i ごとに $N(t)$ を用いた回帰分析や最尤法によりパラメータを推定してきたが、多重共線性や誤差の把握の問題が指摘されており、非線形最小二乗法が推奨されている [2]。ここでは、滞留時間と購買額の間関係を考察の対象としているため、 $n(t)$ に基づいて非線形最小二乗法によりパラメータを推定する。

百貨店の ID 付 POS データに対し、データ全体、カード種別、月別、平日と休日、購買時間帯別にそれぞれパラメータを推定した。曲線の形状は、集計区分を変更しても類似していた通り、パラメータ p と q も類似した値となった。ただし、夕方以降の購買では q が大きくなり、店舗内での普及効果が高いことがわかった。購買ポテンシャル m は売上の変動と類似した推移を示した。興味深いのはカード種別で、ポイント・カードよりも割引カードの方が購買ポテンシャルが高くなっている点である。このような分析を各店舗ごとに実施

すれば、店舗内の普及効果や購買ポテンシャルを測る評価指標として活用できると考えられる。

4. 購買パターン分析

前節の通りに購買金額は滞留時間 100 分前後でピークとなる。この原因を探るため購買パターン分析を行った。購買した売場の組合わせを調べると 14357 パターンを抽出でき、最大は 11 で、食料品を中心とした単一売場選択が上位を占めている。これらパターンの購買額の分布を調べると、前節の滞留時間と購買額の関係と類似しており、滞留時間と購買額は密接に関係していることがわかった。さらに、売上順でデシル分析を行ったところ、デシル 1 だけで全体の売上の 94.6% を占めていた。有力な購買パターンを抽出するために、度数が 1000 以上で平均購買額が 1 万円を超えるパターンに絞り込むと、88 パターンとなった。これらは、単一売場、近接 2 売場、遠接 2 売場、3 売場の選択によって 4 クラスターに分類でき、1 売場 30 分が目安であり、滞留時間 100 分は 3 売場選択に相当することがわかった。この他にも月別・曜日別のパターンや順列購買パターンについても調べ、平日の売上を向上させる手がかりや購買順序が購買額に大きく影響しないこと、回遊性を高める購買パターンの基軸などを明らかにできた。さらには、有力な購買パターンによる売上を 2 節の手法で空間に展開すると、購買額密度のコアを形成しており、一見すると無関係と考えられがちな顧客の空間的分布と購買パターンは密接に関連していることも見いだせた。

5. おわりに

本稿では、時空間的な側面から、ある百貨店の ID 付 POS データに対してデータマイニングを行った結果を報告した。その結果、週末よりも平日の商圈の方が広く、週末の売上は平日の商圈を維持できれば向上させることができること、冬と夏には店舗近隣の居住者に対するアプローチが重要なこと、ポイントカードより割引カードの方が売上への貢献度が高いことや売上に貢献する購買パターンの性質についての知見を得た。

参考文献

- [1] Bass, M. F.: A new product growth for model consumer durables, *Management Science*, 15(5), 215-227, 1969.
- [2] Srinivasan, V. and C. H. Mason: Nonlinear least squares estimation of new product diffusion models, *Marketing Science*, 5(2), 169-178, 1986.
- [3] Silverman, B. W.: Density estimation for statistics and data analysis, Chapman & Hall: London, 34-95, 1986.