

あるホームセンターの売上予測と店舗グループ化

2005MM035 町田佳子 2005MM040 三浦有理沙

指導教員：田中豊

1 はじめに

近年、ホームセンター業界の売上は日々大きく変化している。売上は、その店舗を取り巻く競合店や周囲の人口などに影響を受けている。現在あるホームセンターでは、新規店舗を出店する際や競合店の出店による売上予測を、担当部署・担当者が経験則に基づいて行っている。

本研究では、その不確定な予測を改善するために、新規店舗の売上や競合店が出店した場合の売上を統計的な根拠を元に予測することを目的としている。また、各店舗を自店情報・商圈情報・競合店情報に基づいてグループ化する。これらの情報については後述する。さらに小分類別売上データや商圈情報に基づいたグループ化も行い、色々な側面からみた各店舗の特徴付けを行った。

2 データについて

2.1 各店舗のデータについて

店舗の数は122店舗である。それぞれの店舗は、27個の変数を持つ。売上実績以外の26変数は、自店情報・商圈情報・競合店情報に分けられる。

これらのデータは全て2007年度のものである。ここでいう2007年度とは、2007年3月5日～2008年3月2日のことである。

2.2 商品分類について

商品は約30万種類ほど存在する。それらは一般食品部門や園芸部門、ペット用品などの27種類の部門に分かれている。さらに、その中に約40種類の中分類、約450種類の小分類と呼ばれるグループが存在する。各部門と各小分類の売上のデータも使用する。

3 分析方法

3.1 売上予測モデルの作成

全体売上予測モデルの作成には、重回帰分析を用いる。全体売上モデルの作成方法と同様に、重回帰分析を用いて27部門それぞれについて売上予測モデルを作成する。

3.2 店舗のグループ化

店舗のグループ化は3種類の方法で行う。1つ目は、重回帰分析の結果から取り出した影響指標を用いる方法である。クラスター分析によって店舗をグループに分け、各グループの意味付けを行う。2つ目は、小分類別売上データを用いる方法である。クラスター分析によって店舗をグループに分け、主成分分析も行って意味付けを行う。3つ目は、商圈情報を用いる方法である。クラスター分析によって店舗をグループに分け、主成分分析や数量化III類を用いて意味付けを行う。

4 全体売上予測モデル

4.1 運行時間優先・面積優先の全体売上予測モデル

4.1.1 重回帰分析

代表競合店の選び方には、自店からの運行時間が短い店舗を代表競合店とする運行時間優先と、より売り場面積が大きい競合店を代表競合店とする面積優先の2通りがある。それぞれ122店舗・27変数のデータを用いて重回帰分析を行う。売上実績を目的変数とし、それ以外の26変数を説明変数とする。自由度調整済み決定係数(以下 R^2 とする)は、運行時間優先では0.7112となり、面積優先では0.7075となった。次にAIC基準で変数選択を行う。このとき、ステップワイズ法(減増法)を用いる。そして、残った変数は共に同じく、OPEN日、売り場面積、人口総数10分、人口65歳以上割合、第3次産業就業者割合、他社競合店からの距離、他社競合店面積和、自社競合店のOPEN、自社競合店売り場面積、自社競合店面積和の10変数となった。 R^2 は運行時間優先では0.7387となり、面積優先では0.7335となった。

4.1.2 回帰診断

運行時間優先の変数選択後の全体売上予測モデルについて、回帰診断を行った。等分散性・独立性・正規性・外れ値のチェックを行うと、261番店、208番店、268番店の3店舗の残差が大きいので、これらの店舗が外れ値と考えられる。また、Cookの距離が0.5を超えていないので、異常な値と言い切ることはできない。ホームセンターと協議した結果、ホームセンター側の意向を含め、外れ値とみなさないこととした。

また、偏回帰プロット図・偏残差プロット図についても問題は見られなかった。

面積優先の変数選択後の全体売上予測モデルについても同様に回帰診断を行った結果、モデルが妥当であることが確認できた[1]。

4.2 競合店を合わせた場合の全体売上予測モデル

運行時間優先・面積優先で異なっていた代表他社競合店を一致させることを考える。この時、各解析結果の残差を比べ、より残差が小さい方の店舗を代表競合店として採用する。それらの競合店は、ホームセンターがより強力な競合店と考えている店舗と一致した。重回帰分析を行った結果得られた、変数選択後のモデルを表1に示す。このモデルは現在考えられる最良のモデルであり、 R^2 は0.7405となった。この R^2 は、昨年度の研究[2]で得られた最良の予測モデルの R^2 より、0.0318ポイント上昇している。得られた回帰モデルで回帰診断を行った結果、特に問題はなかった。

変数	回帰係数	p 値
切片	748465909	0.021850 *
OPEN 日	28631	0.011832 *
売り場面積	164107	< 2e-16 ***
人口総数 10 分	9758	1.58e-12 ***
人口 65 歳以上割合	-16127839	0.100955
第 3 次産業就業者割合	-8396574	0.100458
他社競合店からの距離	28569379	0.006884 **
他社競合店面積和	-11858	0.000232 ***
自社競合店の OPEN	-75949	0.002549 **
自社競合店売り場面積	33259	0.021968 *
自社競合店面積和	-39488	0.000207 ***

予測値にマイナスの値がみられたため、 \hat{y} の Box-Cox 変換を試みた。パラメータ λ の 95 % 信頼区間を求めると $0.2 < \lambda < 0.5$ で $\lambda = 0$ は含まれていないがマイナスの予測値をなくすために \log 変換を行った。そして、マイナスの予測値をなくすことができた。

4.3 部門別売上予測モデル

全体の売上予測モデルを作成したときと同じように重回帰分析をかけ、変数選択を行い、各部門別の売上予測モデルを作成した。たとえば売上構成比が 1 番高い部門 2 では 10 変数が残った。変数の数としては、全体の売上予測モデルと変わらないが、残った変数が変わっている。例えば、全体の売上モデルには、第 3 次産業就業者割合が残っているが、部門 2 には、第 3 次産業就業者割合はなく、他社競合店舗総数が、変数として残っている。部門によって、変数選択を行った後に残った変数は違っていた。また、全体の売上モデルと同じ変数を用いて、部門別の売上を予測してみたところ表 2 のようになった。

表 2 部門別売上予測モデルと部門別で全体売上予測モデルの変数を用いたモデルの各予測精度を比較 (一部抜粋)

部門	部門別予測精度 (%)	全体売上予測精度 (%)	予測精度差
部門 2	69.25	68.44	0.81
部門 12*	64.97	65.24	-0.27
部門 8	61.55	61.55	0.00
部門 3	74.65	74.10	0.55
部門 18*	64.35	64.48	-0.13
部門 14	74.15	73.80	0.35
部門 17	70.00	69.85	0.15
部門 1	73.44	72.46	0.98
部門 22*	66.14	66.20	-0.06

表 2 の部門の横にある * についてだが、* がついている部門は、部門別売上予測モデルの予測精度と全体売上予測モデルの予測精度を比べたときに、全体売上予測モデルの予測精度の方が良いものである。

表 2 より、部門別の変数を用いたときの売上予測モデルと全体の売上予測モデルの R^2 を比べると、どの部門も大きな差はみられなかった、よって、各部門で予測モ

デルを作成するのではなく、全体の売上予測モデルの変数選択結果を用いることができるとわかった。

5 影響指標による店舗グループ化

5.1 影響指標について

次に、前節までに作成した全体売上予測モデルから影響指標を算出した。影響指標とは、下の数式で算出した数値のことである。

$$\hat{y} = \alpha_0 + \alpha_1 x_1 + \cdots + \alpha_n x_n + \alpha_{n+1} x_{n+1} + \cdots + \alpha_p x_p + \alpha_{p+1} x_{p+1} + \cdots + \alpha_q x_q \quad (1)$$

$\alpha_1 x_1 + \cdots + \alpha_n x_n$ が自店影響指標、 $\alpha_{n+1} x_{n+1} + \cdots + \alpha_p x_p$ が商圈影響指標情報、 $\alpha_{p+1} x_{p+1} + \cdots + \alpha_q x_q$ が競合店影響指標であり、 \hat{y} に及ぼす影響の大きさを表すものである。そこで、これらを自店情報、商圈情報、競合店情報の影響の大きさを表す指標として考えるとする。これらの影響指標の偏差値を用いる方法と、相対的な影響指標の偏差値を用いる方法の 2 通りで解析を行った。

5.2 影響指標の偏差値を用いた解析

5.2.1 データについて

前小節で定義した影響指標を全体売上予測モデルから部門別売上予測モデルに拡張し、部門別売上構成比が高い 22 部門の各々に対する店舗別の影響指標 (自店情報・競合店情報・商圈情報) を算出しそれぞれ偏差値を求めた。解析に使用するデータは個体を店舗 (120 店舗)、変数を全体売り上げに対する影響指標と各部門 (22 部門) の売上に対する影響指標の 23 変数として、解析を行う。

自店情報・商圈情報・競合店情報の影響指標の偏差値について、それぞれ全体と各部門の売上に対する影響指標の偏差値の相関係数を求めると、部門 11 の相関係数は 0.7548 と他の部門より低くなっていることがわかる。それ以外の相関係数は自店情報・商圈情報・競合店情報の全てにおいて 0.9 を超えており、非常に相関が高い。

自店情報・商圈情報・競合店情報の影響指標の偏差値について主成分分析を行う。どの場合も第一主成分で累積寄与率は 0.9 を超える。また、各主成分の第 1 主成分の係数はほぼ等しく、各店舗が全体や各部門の売上に対して受けている影響の傾向は似ていることがわかる。

よって、全体の売上に対する影響指標の偏差値を用いた場合の店舗のグループの分け方を、部門 11 以外の部門にも適用できると考えることができる。

5.3 クラスタ分析結果

5.2.1 で説明したデータを用いて、分散 1、平均 0 に標準化を行い、ward 法を用いてクラスタ分析を行った。6 個の群に分け、図 1 のような自店情報・商圈情報・競合店情報の偏差値についてのプロット図を書いて意味付けをした。また、図 1 は、横軸が自店の影響指標、縦軸は、商圈の影響指標である。第 1 群は、「平均的に影響を受けている群」。第 2 群は、「どの要因からも影響を受けていないが、特に自店に影響を受けていない群」。第 3 群は、

「商圈と競合店に大きく影響を受けているが、特に商圈に影響を受けている群」。第4群は、「自店に大きく影響を受けている群」。第5群は、「自店にやや影響を受けている群」。第6群は、「競合店にやや影響を受けている群」となる。

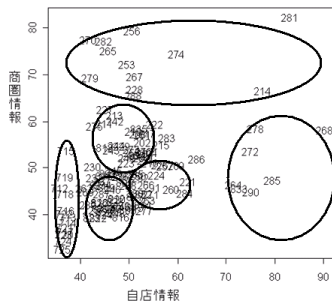


図1 横軸：自店情報、縦軸：商圈情報

5.4 相対的な影響指標の偏差値を用いた解析

相対的な影響指標である、店別の全体と各部門の売上に対する影響指標を \hat{y} で割った比の偏差値を用いて分析を行った。

全体売上や各部門の売上に対する自店情報・商圈情報・競合店の影響指標の割合を見ると、偏差値が小さすぎたり大きすぎる店舗がある。中には偏差値が0より小さい店舗や、100より大きい店舗もある。また、自店情報割合の影響指標の偏差値のデータは第6主成分、商圈情報割合の影響指標の偏差値のデータは第7主成分、競合店情報割合の偏差値のデータは第5主成分で累積寄与率が0.8を超える。それでは説明がしづらく、データのばらつきも明らかになったため、相対的な影響指標は用いないこととした。

6 小分類の主成分分析・クラスター分析

6.1 120小分類の主成分分析・クラスター分析

小分類の中でも、売上構成比が高い、店舗別の120小分類の売上データを用いて主成分分析を行う。

第1主成分で十分説明が可能であるが、第3主成分まで意味付けを行う。

第1主成分(寄与率:0.8571)は、「120小分類の売上の総合的な大小」を示す主成分、第2主成分(寄与率:0.0262 累積寄与率:0.8833)は、「企業需要が高いか、一般家庭需要が高い」かを示す主成分、第3主成分(寄与率:0.0206 累積寄与率:0.9039)は、「マンションでの需要が高く季節のない商品がよく売れるか、一軒家での需要が高く季節商品がよく売れる」かを示す主成分である。

次に、分散1、平均0に標準化を行い、ward法を用いてクラスター分析を行った。その結果、5つのグループに分けることができた。

そして、各主成分をプロットしたものをグループに分けた(図中に円で示す)。また、図2は、横軸が第1主成分、縦軸が第2主成分のプロット図である。

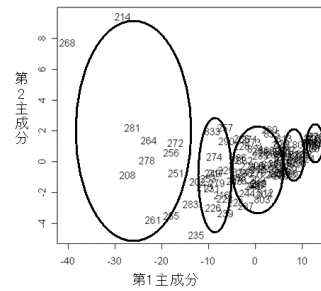


図2 第1主成分と第2主成分のプロット図

以上の結果より、以下のように意味付けをすることができる。第1群は、「全体的に売上が小さい群」、第2群は、「全体的に売上がやや小さい群」、第3群は、「売上の大きさと各需要の高さが平均的な群」、第4群は、「全体的に売上が大きい群」、第5群は、「売上が大きめで、一般家庭需要が高い群」となる。

第1群には店舗番号700番代の特別に小さな売り場形態で営業している店舗が多く、実際売上も小さくなっている。また、第2群には北陸地方の店舗が多い。

6.2 48小分類の主成分分析・クラスター分析

次に、売上構成比の高い48小分類に絞って解析を行う。120分類と同じように、解析を行うと以下のような結果になった。主成分分析を行った結果、第1主成分で十分説明が可能であるが、第3主成分まで意味付けを行う。

よって、第1主成分(寄与率:0.8596)は、「総合効果により全体的に売り上げが低い」主成分となる。第2主成分(寄与率:0.0238 累積寄与率:0.8835)は、「企業需要が高いか、一般家庭需要が高い」という主成分となる。第3主成分(寄与率:0.0218 累積寄与率:0.9053)は、「一軒家の需要が高く季節商品がよく売れるか、マンションでの需要が高く季節性のない商品がよく売れる」という主成分となる。

次に、クラスター分析を行う。その結果、第1群は「全体的に売上が小さい群」、第2群は「全体的に売上がやや小さく、やや一軒家の需要が高い群」、第3群は「売上がやや大きく、一般家庭需要が高い群」、第4群は「売上の大きさ、各需要の高さが平均的な群」、第5群は「全体的に売上が高い群」となる。

120小分類で分析を行った場合と同じく、第1群には700番代の小型店が多く、実際売上も小さくなっている。また、第2群には東北地方の店舗が多い。この傾向は、120小分類で分析を行った場合よりも強く表れている。

7 商圈情報による店舗グループ化

7.1 主成分分析・クラスター分析による店舗グループ化

7.1.1 データについて

売上予測モデルを作成する際の商圈情報には、田舎都会フラグ1、田舎都会フラグ2、道路面数、各店舗からの5分以内・10分以内・15分以内の人口総数、各店舗の商圈内の15歳未満・15~64歳・65歳以上の各人口割合、第

1次・第2次・第3次の各産業の就業者割合の12変数があった。しかしここでは、質的なデータである田舎都会フラグと道路面数を抜いた10変数を用いて主成分分析・クラスター分析を行う。

7.1.2 分析結果

累積寄与率が0.8を超える第3主成分までを考察する。第1主成分(寄与率:0.4766)は、「65歳以上の人口割合が高く第1次・第2次産業が盛んか、店舗のまわりの人口が多くて15~64歳の人口割合が高く、第3次産業が盛んかの主成分」。第2主成分(寄与率:0.2735, 累積寄与率:0.7501)は、「65歳以上の人口割合が高いか、15歳未満・15~64歳の人口割合が高く第2次産業が盛んかの主成分」。第3主成分は(寄与率:0.1060, 累積寄与率:0.8560)「第1次産業が盛んか、そうではないかの主成分」である。

データを標準化をし、ward法でクラスター分析を行った。図3は、第1主成分と第2主成分のプロット図である。クラスター分析で分けたグループを、図中に円で示す。

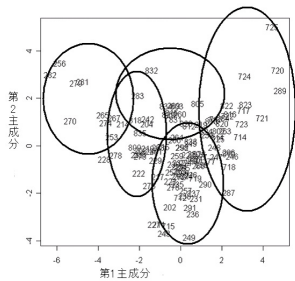


図3 第1主成分と第2主成分のプロット図

主成分分析・クラスター分析による各群の意味付けをした。第1群は、「人口総数10分・15分が少なく、人口15~64歳割合が低くて65歳以上割合が高く、第1次産業就業者割合が高い群」。第2群は、「人口65歳以上割合が低く、人口15歳未満・15~64歳割合が高く、第2次産業就業者割合が高い群」。第3群は、「人口総数5分・10分・15分が多く、人口15~64歳割合・第3次産業就業者割合が高く、第1次産業就業者割合が低い群」。第4群は、「人口65歳以上割合が高く、人口15歳未満・15~64歳割合・第2次産業就業者割合が低い群」。第5群は、「人口総数5分・10分・15分が多めで、人口15~64歳割合が高く、第3次産業就業者割合が高めの群」となる。

7.2 数量化III類による店舗グループ化

7.2.1 データについて

数量化III類では主成分分析、クラスター分析とは異なり、質的なデータを取り扱うことができる。量的なデータに関しては、少・中・多の3段階に分けて使用した。それぞれ当てはまる店舗には1が、当てはまらない店舗には0が格納されている。34変数、122店舗のデータを使用した。

7.2.2 分析結果

累積寄与率が0.8を超えるのは第10固有値まで見た場合である。しかしそれでは解釈がしづらい。そのため、寄

与率が第2固有値と第3固有値の間で大きく下がっていると見て、第2固有値までについて考えるとする。第1, 2軸の得点の散布図をプロットすると、各群円でくくった図4となる。

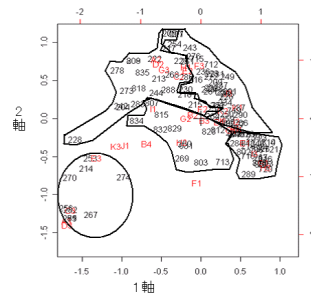


図4 第1・2軸の得点の散布図

クラスター分析で分けた店舗グループと同様のグループで意味付けを行った。第1群は、「65歳以上の人口割合が高く、15~64歳の人口割合が低く、第1次産業就業者が多く、道路面数が少ない群」。第2群は、「15歳未満の人口割合が高く、第2次産業就業者が多く、第3次産業就業者が少ない群」。第3群は、「店舗のまわりの人口が多く、都会である群」。第4群は「15歳未満の人口割合が低く、65歳以上の人口割合が平均的である群」。第5群は「店舗から5分から10分圏内の人口総数は平均的であり、15~64歳の割合が高く、田舎でも都会でもない地域である群」である。

このように、クラスター分析と主成分分析による意味付けと類似した意味付けを行うことができた。第1群には700番台の小型店が多い。第2群は平均的な店舗が多い。第3群は名古屋市内の店舗が多い。第4群は北陸の店舗が多い。

8 おわりに

本研究では、売上予測モデルの作成と店舗のグループ化を行った。

全体売上予測モデルの予測精度は $R^2 = 0.7405$ であり、昨年度の研究の結果より0.0318ポイント上昇した。

また、予測精度を比較することにより、部門別の売上を予測するには、全体の売上予測モデルで可能であると確認された。

店舗のグループ化に関しては、売上予測モデルから算出した影響指標を用いる方法、120小分類と48小分類の小分類別の売上データを用いる方法、商圈情報のデータを用いる方法の3種類の方法で結果を得ることができ、今後のホームセンターの必要に応じて状況に見合ったグループ分けを提案できる可能性を示すことができたと思われる。

参考文献

- [1] 田中豊・垂水共之：Windows版統計解析ハンドブック 多変量解析，共立出版株式会社，1995
- [2] 大井俊二：あるホームセンターの売上分析，南山大学数理情報学部数理科学科卒業論文，2008