

大規模レシートデータの分析と活用

2011SE178 村上史也 2011SE251 鈴木政樹

指導教員：鈴木敦夫

1 はじめに

現在、小売業ではPOS (Point Of Sales) システムを導入し、個人情報を含んだ顧客IDが付与されたレシートデータを大規模かつ継続的に収集している。このID付きレシートデータを利用し顧客動向や販売動向などを分析できると期待されている。[1]

しかし、大規模に収集されたID付きレシートデータはデータ量が膨大であり、その利用法が確立されていない。

今回委託研究を受けたホームセンターもID付きレシートデータを収集しており、以前にもID付きレシートデータの提供を受けている。しかし、分析範囲は数週間程に限られており、大規模なID付きレシートデータの活用法を模索している状態である。[2]

本研究ではそのID付きレシートデータの抽出、分析を行い実際の企業活動での分析や店舗の販売促進策に活用できる手法の確立を目指していく。

販売促進策活用への手法として、提供されたID付きレシートデータをソフトウェアを利用し抽出し対応表や散布図の作成を行った。また、主に主成分分析の結果を基に販売促進策を提案することにした。

2 使用データについて

2.1 提供されたID付きレシートデータ

本研究では提供されたID付きレシートデータから売上日付、部門、JAN、商品漢字名、規格漢字名、数量、販売売価、売上金額、会員番号、年齢、性別、特売コード、特売本体売価を適宜使用した。JANは商品固有の値であり単独で識別可能であるが、商品漢字名、規格漢字名も利用して確認する。数量と売上金額、部門は商品の購買パターン分析に利用する。売上日付、会員番号、年齢、性別は顧客のパターン分析に利用する。特売コード、特売本体売価はホームセンターの行う広告配布や値下げの分析に利用される。

2.2 特売商品の販売促進策

広告掲載商品である特売商品は現在販売促進策の要と考えており、本研究でも優先して研究を行った。ホームセンターは特売を行う日が通常の営業日に対し、多いと考えており特売日の来店日数より会員顧客の売上に注目した。図1は会員顧客1人の売上と特売商品の売上の散布図である。この図では一定の売上まで特売商品の売上も増加していることが分かる。

会員顧客個人の売上が会員顧客個人の総売上に対する割合を特売反応率とし、その対応表を作成した。表1は特売商品の販売促進策を行う際の人数把握に活用できる。特売反応率が高いほど広告に興味があると考えられる。会員顧客グループを想定し効果的な広告を行うことができる。

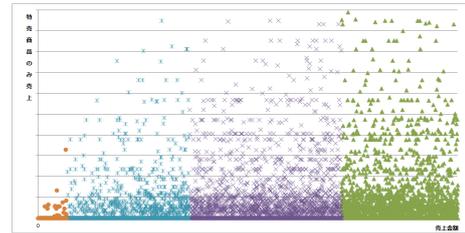


図1 売上特売散布図

表1 特売反応率会員数表

特売反応率 / 会員数	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
0%	25177	4	37	84	20	14	10	7	5	3	1
10%	19866	24	60	140	32	24	21	9	5	2	1
20%	16237	84	176	446	103	55	4	6	2	1	1
30%	14269	183	239	423	389	55	1	6	2	1	1
40%	11539	283	307	431	362	10	1	1	1	1	1
50%	9799	447	289	385	31	10	1	1	1	1	1
60%	8566	453	349	355	46	2	1	1	1	1	1
70%	6866	457	236	321	32	6	2	3	2	1	1
80%	4566	457	174	274	30	6	2	1	1	1	1
90%	4461	453	157	233	26	7	1	1	1	1	1
100%	19111	936	380	1340	75	11	1	1	1	1	1

3 主成分分析によるレシートデータ分析

3.1 分析方法

会員番号ごとに顧客がどのような種類の商品を購入するのかを分析し、顧客をグループ化した。分類する種類はA,B,C,D,E,Fに設定した。Aは土や植物、レンガなど園芸、農業に関するもの、Bは軍手、木材、やすりなどの工具用品に関するもの、Cはペット用品やスポーツなど娯楽に関するもの、Dは洗剤や食料品、医薬品などの雑貨に関するもの、Eはカーテンや机など家具に関するもの、Fはテレビ、洗濯機などの家電用品に関するものである。

3.2 分析結果

以下が主成分分析の結果である。

表2 主成分分析の結果1(標準偏差, 寄与率, 累積寄与率)

	PC1	PC2	PC3
Standard deviation	1.3544	1.0155	0.9441
Proportion of Variance	0.3057	0.1719	0.1486
Cumulative Proportion	0.3057	0.4776	0.6261
	PC4	PC5	PC6
Standard deviation	0.9140	0.8646	0.8126
Proportion of Variance	0.1393	0.1246	0.1100
Cumulative Proportion	0.7654	0.8900	1.0000

分析結果から標準偏差, 寄与率, 累積寄与率は表2の結果となった。この結果から第1主成分(PC1)の寄与率0.3057と第2主成分0.1719, 第3主成分までで主成分分析を行う。

3.3 考察

主成分得点の結果としてPC1(第1主成分)はすべての係数が正の値をとる。そのため第1主成分は顧客の購入金額を表している。係数が正の値が大きいほど購入金額が大きいことを表している。

表 3 主成分分析の結果 2(主成分得点)

	PC1	PC2	PC3	PC4	PC5	PC6
A	0.343	0.351	-0.792	0.168	0.262	0.186
B	0.381	0.548	0.080	-0.375	-0.602	-0.211
C	0.336	-0.622	-0.216	-0.535	-0.177	0.369
D	0.500	-0.330	-0.015	0.106	0.215	-0.764
E	0.436	-0.141	0.228	0.698	-0.374	0.333
F	0.430	0.247	0.517	-0.214	0.593	0.301

PC2(第2主成分)は係数が正の値で大きいものはA, B, 係数が負の値で大きいものはC, Dとなった。そのため第2主成分はホームセンターに目的があって来店してくるグループか、スーパーと同じ目的で来店するグループに分けることができる。係数が正の値が大きいほど園芸や日曜大工の用具など、係数が負の値が大きいほどペット用品や日用消耗品などを購入する傾向があると考えられる。

PC3(第3主成分)は係数が正の値で大きいものはA, E, 係数が負の値で大きいものはA, Cとなった。そのためこの値が大きいほど家電やインテリアなどを目的としたインドアなグループと考えられる。また負の値が大きいほど園芸やペット用品、スポーツなどを目的としたアウトドアなグループであると考えられる。

3.4 散布図

第1主成分と第2主成分, 第3主成分から視覚化をするために散布図で表した。図2が会員を点として横軸第1主成分, 縦軸第2主成分, 図3が横軸第1主成分, 縦軸第3主成分, 図4が横軸第2主成分, 縦軸第3主成分として表示したものである。横軸では右に行くほど第1主成分の正の特徴が, 左に行けば負の特徴が強くなる。縦軸では上に行くほど第2主成分の正の値の特徴が強くなり, 下に行くほど負の値の特徴が強くなる。

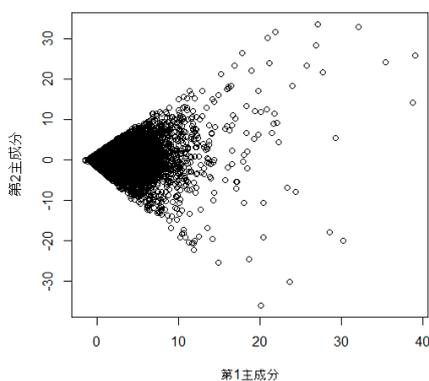


図 2 第1主成分 第2主成分

3.5 散布図の考察

図2より(0.0)から扇形に広がっていることが分かる。これにより第2主成分の値が正負ともに大きいとき第1主成分の正の値も大きくなるという傾向が強いということが分かる。つまり購入する商品がある一定の分野に限

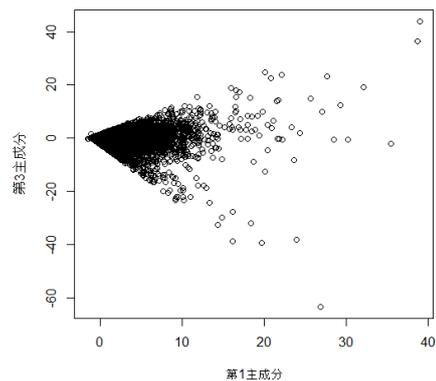


図 3 第1主成分 第3主成分

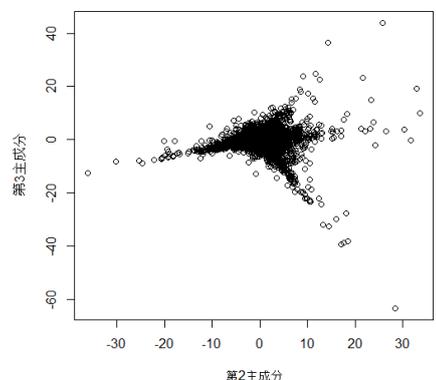


図 4 第2主成分 第3主成分

られている会員ほど購入金額は多くなるということである。よって顧客の購入金額を増やすためには, 購入している商品の種類を限定し, 販売促進策を行うことにより, 左上のグループを右上のグループへ, 左下のグループを右下のグループへ移動させ購入金額を増やすことが可能になると考える。

図3は第1主成分, 第2主成分の図と同様に第3主成分の傾向が強いほど第1主成分の値も大きくなっていること考えられる。

図4より日用品が目的である人にはインドア, アウトドアの影響は出にくいと考えられる。また第2主成分が正の値が大きくなるにつれば第3主成分も増加する傾向にあるため, ホームセンターの商品が目的である傾向が強くなるたびにインドア, アウトドアの傾向も強くなると考えられる。

4 クラスタ分析

商品分類, 顧客の関係から似た顧客をグループ化し分類するためにクラスタ分析を用いた。このクラスタ分析よりデンドログラムを作成し視覚化した。

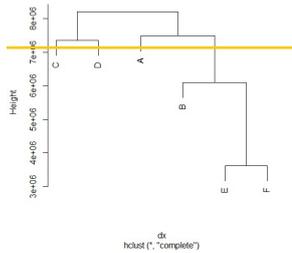


図 5 デンドログラム

4.1 考察

分類した関係を 4 つのグループにまとめる。C 単独のグループ、D 単独のグループ、A 単独のグループ、E と F、B のグループに分けることができる。

C, D, A の単独のグループでは他のラインとの関係性が薄いと考えられる。C はペット用品のみの購入、D は消耗品のみの購入、A は園芸用品のみの購入する客層が多いためと考えられる。E と F, B のグループは家具が関連するものであると考えられる。E と F が特に関係が近くインテリア関係であり、B も木材や工具など家具などを作成する際に必要な物だと考えられる。

5 顧客の一般化

顧客を優良顧客、一般顧客に分類する。グループごとの購入金額が多い顧客を優良顧客、購入金額が少ないものを一般顧客とする。初めに用いた顧客情報は会員全体のものであるため一般顧客に対する結果としては不十分などところがある。そのため会員情報から外れ値である特殊な顧客を除き、できるだけ一般会員のみデータの近づく主成分分析を行うことにした。特殊な顧客を除く条件として、グループごとに平均値、標準偏差をとり、平均値から標準偏差を二倍以上の場合は、分析対象から特殊な顧客として除く。これにより一般顧客の情報とし顧客を一般化させる。

5.1 一般顧客の分析結果

表 4 一般顧客、主成分分析の結果 1(標準偏差, 寄与率, 累積寄与率)

	PC1	PC2	PC3
Standard deviation	1.4634	0.9865	0.9118
Proportion of Variance	0.3569	0.1622	0.1386
Cumulative Proportion	0.3569	0.5191	0.6577
	PC4	PC5	PC6
Standard deviation	0.8766	0.8373	0.76454
Proportion of Variance	0.1281	0.1169	0.09742
Cumulative Proportion	0.7857	0.9026	1.00000

第 1 主成分は顧客の購入金額を表している。係数が負の値が大きいほど購入金額が大きいことを表している。

第 2 主成分はホームセンターとして目的として来店してくるグループか、一般商品を目的として来店するグループに分けることができる。係数が正の値が大きいほど園芸や日曜大工の用具など、係数が負の値が大きいほどペッ

表 5 一般顧客の主成分分析の結果 2(主成分得点)

	PC1	PC2	PC3	PC4	PC5	PC6
A	0.392	-0.482	0.222	-0.556	-0.380	0.334
B	0.389	-0.564	0.179	0.241	0.574	-0.334
C	0.335	0.547	0.692	0.031	0.223	0.242
D	0.483	0.281	-0.071	-0.079	-0.410	-0.713
E	0.397	0.264	-0.607	-0.382	0.463	0.210
F	0.438	-0.037	-0.258	0.693	-0.306	0.407

ト用品や日用消耗品などを購入する傾向があると考えられる。

第 3 主成分は正の値が大きいほど家電やインテリアなどを目的としたインドアなグループと考えられる。また負の値が大きいほど園芸やペット用品、スポーツなどを目的としたアウトドアなグループであると考えられる。

第 4 主成分は正の値が大きいほど家電を目的としたグループと考えられる。また負の値が大きいほど、園芸を目的としたグループであると考えられる。

5.2 顧客のグループ化

第 4 主成分までを利用して、顧客購買傾向を 16 のグループに分類する。

		アウトドア		インドア	
		家電	園芸	家電	園芸
額多	一般	1	2	3	4
	専門	5	6	7	8
額少	一般	9	10	11	12
	専門	13	14	15	16

- 購入金額が多く、一般商品を購入し、アウトドア商品を多く購入し、家電を目的としたグループ。
- 購入金額が多く、一般商品を購入し、アウトドア商品を多く購入し、園芸を目的としたグループ。
- 購入金額が多く、一般商品を購入し、インドア商品を多く購入し、家電を目的としたグループ。
- 購入金額が多く、一般商品を購入し、インドア商品を多く購入し、園芸を目的としたグループ。
- 購入金額が多く、専門商品を購入し、アウトドア商品を多く購入し、家電を目的としたグループ。
- 購入金額が多く、専門商品を購入し、アウトドア商品を多く購入し、園芸を目的としたグループ。
- 購入金額が多く、専門商品を購入し、インドア商品を多く購入し、家電を目的としたグループ。
- 購入金額が多く、専門商品を購入し、インドア商品を多く購入し、園芸を目的としたグループ。
- 購入金額が少なく、一般商品を購入し、アウトドア商品を多く購入し、家電を目的としたグループ。
- 購入金額が少なく、一般商品を購入し、アウトドア商品を多く購入し、園芸を目的としたグループ。
- 購入金額が少なく、一般商品を購入し、インドア商品を多く購入し、家電を目的としたグループ。
- 購入金額が少なく、一般商品を購入し、インドア商品を多く購入し、園芸を目的としたグループ。

を多く購入し、園芸を目的としたグループ。

13. 購入金額が少なく、専門商品を購入し、アウトドア商品を多く購入し、家電を目的としたグループ。
14. 購入金額が少なく、専門商品を購入し、アウトドア商品を多く購入し、園芸を目的としたグループ。
15. 購入金額が少なく、専門商品を購入し、インドア商品を多く購入し、家電を目的としたグループ。
16. 購入金額が少なく、専門商品を購入し、インドア商品を多く購入し、園芸を目的としたグループ。

6 購買分類の細分化

顧客の購買傾向のグループをより細かく知るために27種類に分類した。

表6 購買分類

記号	分類	記号	分類
a	園芸	o	消耗品
b	園芸2	p	文具
c	園芸3	q	キッチン
d	園芸4	r	バス用品
e	園芸5	s	薬品
f	作業用品	t	食品
g	金物	u	インテリア1
h	工具	v	インテリア2
i	塗装	w	インテリア3
j	木材	x	インテリア4
k	カー用品	y	インテリア5
l	玩具	z	インテリア6
m	レジャー	aa	インテリア7
n	ペット用品		

6.1 分類の購入傾向

前章での16の顧客グループに会員を分ける。分類方法はRを使用し顧客の第1主成分、第2主成分、第3主成分、第4主成分がそれぞれ正か負か判別し分類する。

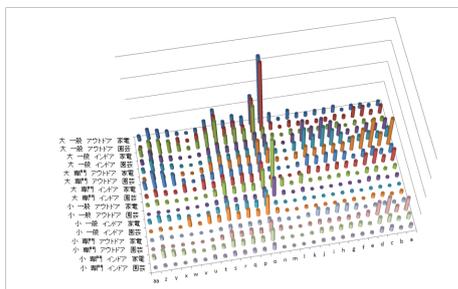


図6 分類ごとの購入金額を表した図

6.2 考察

グラフの結果から専門商品を購入する顧客は専門商品だけではなく、一般商品も購入することが分かる。一般商品を購入する顧客は専門商品はほとんど買わず、一般商品

のみを購入していることが分かる。インドア、アウトドアは細分化された項目であると、インドアはペット用品、アウトドアは消耗品の傾向があると考えられる。

		ペット用品		消耗品	
		家電	園芸	家電	園芸
額多	一般	1	2	3	4
	専門	5	6	7	8
額少	一般	9	10	11	12
	専門	13	14	15	16

7 主成分分析を基にした購買傾向による販売促進策

今後の販売販促としては購入金額が少ない会員顧客を購入金額が多い会員顧客グループに誘導させるため、専門商品を購入する客層には専門商品のみを販促するのではなく、一般商品も販促していく方がよいと考える。一般商品を購入する客層には専門商品を販促させずに一般商品のみを販促させた方がよいと考える。ペット用品商品を購入する客層はペット用品を強く販促しながらも消耗品や薬品も販促していく方がよいと考える。消耗品を購入する客層には消耗品を強く販促しながらも日用品を販促していく方がよいと考える。園芸を目的とした客層には園芸類の販促を行い、家具を目的とした客層には日用品を販促していく方がよいと考える。そして、本研究で主成分分析を基に効果があると思われた特定の会員顧客グループに対し、販売促進となるサービスクーポンのダイレクトメールを送付する予定である。

8 おわりに

本研究ではホームセンターより提供されたID付きレシートデータを分析し販売促進策につなげることを目的とした。そのためにID付きレシートデータを抽出、分析し表化することでそのまま販売促進策となる方法や主成分分析を基にした販売促進策を考え実施予定まで到達した。

展望として、従来より限られた期間、人数で評価されてきた研究も更に大規模なデータを利用することで新しい発見ができると思う。本研究でも期間や対象会員の増加によってより有意義な結果を示せるのではないだろうか。また、研究課題として、爾後の研究は蓄積されたノウハウを残し、利用可能にすること。ホームセンターの会員数増加、対象期間長期化で更なる大規模なデータ利用が将来予想され、PCのスペック上昇やデータ分割法を更に考慮することをあげる。

参考文献

- [1] 久松俊道, 朝日弓未, 山口俊和: ドラッグストアのID付きPOSデータを用いた日用品購買パターンの比較分析, オペレーションズ・リサーチ Vol.56, No.2, pp63-69, 2012.
- [2] 崔康幸, 岩瀬爽, 岡村彩音: 広告掲載商品の最適選定問題, 2012年度南山大学情報理工学部情報システム数理学科卒業論文, 2013.