

百貨店における店舗配置からの知識発見 —GAを用いた最適店舗配置の提案—

北口 大輔, 羽室 行信, 加藤 直樹, 加藤 玲

1. はじめに

近年の百貨店業界は、大手百貨店の倒産に見られるように、不況にあえいでいる。多くの百貨店では、テナントとしての外部企業に店舗の運営を委託しており、テナント店舗の売上比率は、百貨店全売上の90%にまで達しているといわれている[5]。近年の百貨店業界の不振の一因は、多くの百貨店が「売場貸業」になってしまったためであるとも言われている。このような状況からの脱却を図ろうと、いくつかの百貨店では、委託による売場の比率を下げ（自主運営売場の比率を上げ）、本来の小売業の形態に立ち戻ろうとする動きもある。しかしながら一方では、売場貸業からの脱却は容易ではなく、むしろ売場貸業を前提として、店舗の出退店管理や効果的なフロア内での店舗配置を積極的に進め、経営の合理化を図るほうが現実的であるという意見もある[8]。

一方で、急速なコンピュータ技術の向上に伴い、多くの百貨店では会員管理とPOSシステムを連動させ、顧客購買履歴データ（いわゆるID付きPOSデータ）を蓄積しはじめている。このデータを利用することによって、よりきめ細かなデータ解析が可能となり、ビジネスに有用な知識を獲得できることが期待されている。しかしID付きPOSデータは、従来の部門管理による集計データに比べ、飛躍的にデータ量が多くなる。そこでKDD（大規模データベースからの知識発見）やデータマイニングといった技術が注目を浴びて

おり[9]、特にここ数年、これらの技術を導入しようとする動きが活発化しはじめている。

以上のことを背景に、本稿では、「売場貸業」としての百貨店であっても実施可能である店舗管理に焦点を当て、顧客にとって魅力的な売場を構成するための手法を提案する。この手法は、ID付きPOSデータにデータマイニング技術を適用することによって可能となる。

本稿の構成であるが、節2において、今回の事例研究で利用したデータについての簡単な基礎分析結果を示す。節3では、魅力的な店舗配置を実現する手法について説明する。そして節4にて、提案する手法を実際のデータに適用して得られた結果についての考察を行う。最後に節5にて、まとめと課題について述べる。

2. 分析対象データ

分析対象のデータは、ある百貨店1店舗における2001年1年間のID付きPOSデータ（顧客購買履歴データ）である¹。この百貨店では、会員カードを配布しており、今回利用した購買データに含まれている顧客人数は約27万人である。

売場の区分として、食料品、婦人服、紳士服、子供用品、家具、家庭用品、スポーツ、食堂など多様な部門が存在するが、今回の分析では婦人関連部門（婦人服、婦人服飾、婦人洋品の3部門）を対象とした。婦人関連部門で1回でも買い物をしたことのある顧客は、27万人のうち10万人で37%を占めている。また、婦人関連部門の総売上は、全体の約37%で、食品部門に次いで高い売上を示しており、経営的に見て重要な部門であると言える。

図1に婦人関連部門に属する店舗のフロア²構成が

きたぐち だいすけ

大阪産業大学 大学院経営・流通学研究科

〒574-8530 大東市中垣内3-1-1

はむろ ゆきのぶ, かとう あきら

大阪産業大学 経営学部

〒574-8530 大東市中垣内3-1-1

かとう なおき

京都大学 大学院工学研究科建築学専攻

〒606-8501 京都市左京区吉田本町

受付 03.7.25 採択 03.10.25

¹ 日本OR学会MDA研究部会平成14年度データ解析コンペティションより提供されたデータを用いている。

² この百貨店は三つの館で構成されており、ここでフロアとは、同一階の同一館の売場のことを指す。

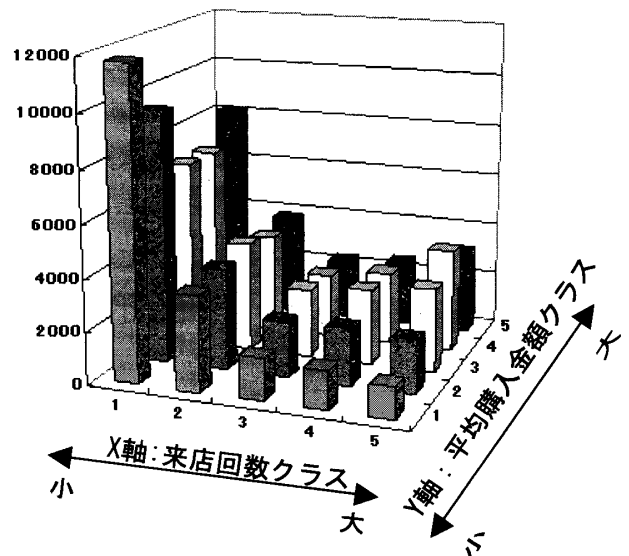
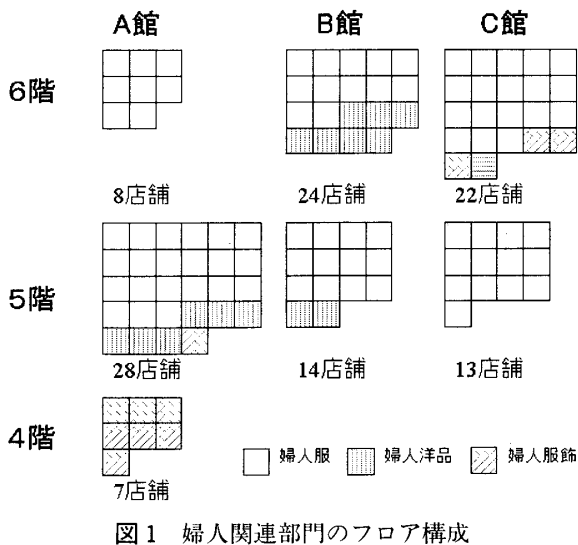
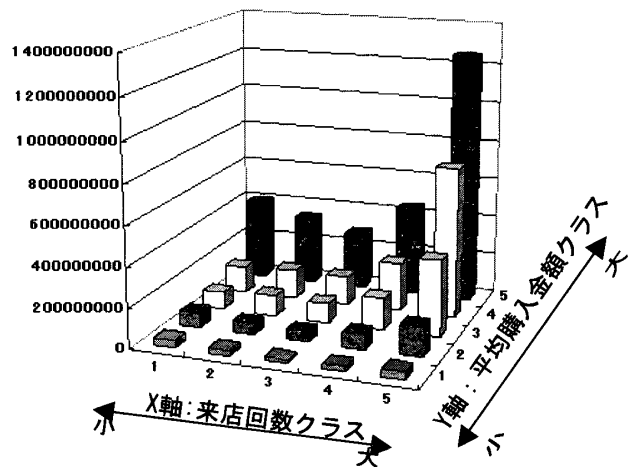


表1 来店回数と購買金額の5分類

クラス	来店回数	購買金額/訪問
5	6回以上	¥25,850以上
4	4,5回	~¥25,850
3	3回	~¥15,750
2	2回	~¥10,050
1	1回	~¥5,950

図2 顧客分類に基づいた人数分布



示されている。一つのマス目は一つの店舗を表している。ただし、マス目の位置は、実際の店舗の位置とは無関係である。この百貨店では、4階から6階のフロアに多くの婦人関連部門の店舗が集中している。各フロアにおける店舗数は図中のとおりである。この図を見てわかるように、店舗の構成は、部門をフロアで輪切りにする（同じ部門の店舗を同一フロアに配置する）傾向にあり、顧客にとって魅力ある売場を作るという観点から見れば、店舗のフロア配置に工夫の余地があると思われる。

次に、各顧客の来店回数と訪問当たりの購買金額について、全10万人の顧客を五つのクラスに分類した（表1）。分類は、各クラスの人数が可能な限り均等になるように行った。

これら二つのクラス分類に基づき、顧客の人数分布と総売上分布を見てみると、それぞれ図2、3のようになる。

図2に示されるように、1年間での来店回数が1回もしくは2回の顧客が圧倒的に多いことがわかる。図3は、それぞれのセルに分類された顧客が、合計でどれくらいの金額を支払っているかを示したものである。

		来店回数				
		1回	2回	3回	4,5回	6回以上
平均購入金額 一来店あたりの	¥25,850以上					
	~¥25,850					
	~¥15,750					
	~¥10,050					
	~¥5,950					

図4 優良顧客の定義

当然ながら、来店回数、売上金額ともに高い顧客が全体の売上の多くを占めていることがわかる。

図4は、今回採用した二つの基準によるマトリックスであるが、灰色で示した六つのセルに分類される顧客は、来店回数、購買金額ともに高く、百貨店に対する貢献度が非常に高い顧客であり、これらの顧客を、

「優良顧客」と定義する。

優良顧客に属する顧客人数は、全体の約17%である。一方で、優良顧客の購買金額合計は全体の約60%に達している。

百貨店としては、顧客にとって魅力あるフロアを構成するに当たっては、売上貢献度の高い優良顧客の購買行動をより重要視する必要があるであろう。そこで本分析では優良顧客の購買データのみを利用することにした。

以上で見てきたように、本分析は、4, 5, 6階に位置する婦人関連部門を対象とし、優良顧客に分類される約1万8千人の顧客購買履歴データを用い、116店舗の7フロアへの配置問題について考えていく。

3. 分析枠組

本節で対象とする問題は、分析対象となる116店舗を七つのフロアにいかにして配置するかというものである。配置の組み合わせ方は、どのような特徴をもったフロアを構成するかによって異なってくるが、以下に示すような方針が考えられるであろう。

- 1) 関連の高い店どうしをできるだけ同一フロアに配置する。
 - 2) 訪問顧客の年齢層が同じ店を同一フロアに配置する。
 - 3) 扱ひ商品単価が同程度の店は同一フロアに配置する（高級商品を販売する店を集めて高級フロアを構成する）。
 - 4) 同一部門に属する店舗は同一フロアに配置する。
- どの方針を選ぶかは、百貨店の考え方に依存して考えると考えられる。本研究では、紙面の都合および説明の簡潔さを理由に、方針1)と2)のみに絞って説明を行うが、本研究で我々が開発したシステムでは、様々な目的の設定も可能である。

以下では、節3.1, 3.2にて方針1)と2)についての目的関数を定式化し、これら目的関数による多目的最適化問題を考える。そして節3.3にてその妥当性について検討する。本研究ではこの問題について遺伝的アルゴリズムを適用するが、節3.4では、その方法について簡単に説明していく。

3.1 店舗間関連度

関連の高い店どうしを同一フロアに配置するに当たって、店舗間の関連性の強さをいかにして定義するかを考える。

百貨店内での店舗を小売店における商品にたとえる

ことにより、店舗間関連度にマーケットバスケット分析[1]で利用される指標を用いることができる。マーケットバスケット分析では、商品間の関連性の強さについて、確信度 (*Confidence*)、支持度 (*Support*)、興味値 (*Interest*)、上昇値 (*Lift*) など様々な指標で表すことが提案されている[10]。どの指標を利用するかは、分析領域に依存するといわれているが、ここでは *Interest*[3]を利用することにした。

その理由は、*Interest* はマーケットバスケット分析で一般的に広く利用されている *Lift* 値[2, 4]と同等であり、また *Confidence* をはじめとするその他の多くの指標においては、商品間の方向性を考慮する必要があり、分析が複雑になるためである。

Interest について説明するために、図5を例として用いる。ある期間に百貨店に訪れた顧客1000人について、店舗*i*を訪れたことのある顧客が200人、店舗*j*を訪れたことのある顧客が100人、そしてそれらの顧客のうち、*i, j*両店舗とも訪れたことのある顧客が40人いたとする。

このとき、*Interest* とは式(1)で表される。

$$Interest = \frac{P(i \wedge j)}{P(i) \cdot P(j)} \quad (1)$$

ここで $P(i)$, $P(j)$ は、顧客が店舗*i*および店舗*j*をそれぞれ訪れる確率を表し、図5に示された人数を当てはめると、 $P(i)=200/1000$ で0.2、 $P(j)=100/1000$ で0.1と計算される。分子の $P(i \wedge j)$ は、*i, j*両店舗とも訪れる確率を表しており、 $P(i \wedge j)=40/1000$ で0.04と計算される。そして *Interest* の値は、 $0.04/(0.2 \cdot 0.1)$ で2と計算される。

式(1)は、顧客が店舗*i, j*をそれぞれ独立に訪れる確率から期待される同時訪問確率(分母)に対する、実際に観測された同時訪問確率(分子)の比を表しており、その値が1より大きければ2店舗間の関連性が

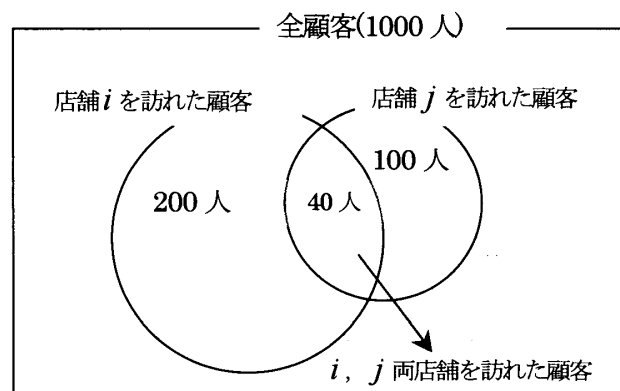


図5 2店舗の来店顧客人数分布

高く（正の関連）、1より小さければ関連性が低い（負の関連）ことを意味する。図5の例では、店舗*i*と店舗*j*の同時訪問確率を見ると、期待される確率に比べ、実際に観察された確率が2倍高く（*Interest*が2）、両店舗の関連度は強いと解釈される。

*Interest*の最小値は0で、*i*、*j*両店舗を同時に訪問した顧客がいない場合、すなわち式(1)の分子が0のときに最小値をとる。一方で式(1)の分母が非常に小さい数の場合、*Interest*の値が極端に大きくなる。例えば、図5において、*i*、*j*店それぞれの訪問人数が1人で、かつ、両店を訪れた人数も1人であったとすると、*Interest*の値は $1/1000 / \{(1/1000) \cdot (1/1000)\} = 1000$ という意味のない非常に大きな値となる。この問題を回避するために、 $P(i) \cdot P(j)$ がある閾値 α より小さいときは、*Interest*の値を1とする。これにより、次に説明する目的関数の最大化には一切貢献しないことになる。*Interest*を式(2)に示されるように定義しなおす。

$$Interest(i, j) = \begin{cases} \frac{P(i \wedge j)}{P(i) \cdot P(j)} & \text{if } P(i) \cdot P(j) \geq \alpha \\ 1 & \text{if } P(i) \cdot P(j) < \alpha \end{cases} \quad (2)$$

この定義により、顧客が店舗*i*、*j*をそれぞれ独立に訪れる確率から期待される同時訪問確率 $P(i) \cdot P(j)$ が低い店舗間の関連性は無視されることになる。本分析では、閾値 α を0.005としている。

図6に全店舗ペアについての $P(i) \cdot P(j)$ （縦軸）と*Interest*（横軸）との関係を散布図で示している。図中の各点は、各店舗ペアに対応している。 $P(i) \cdot P(j)$ の値が0.005を下回る（図中の横線より下側）と、*Interest*の値が極端に大きくなるケースが多くなっていることが確認できる。これらの店舗ペアの店舗間関連度は1となり、これらの関連は無視されることになる。

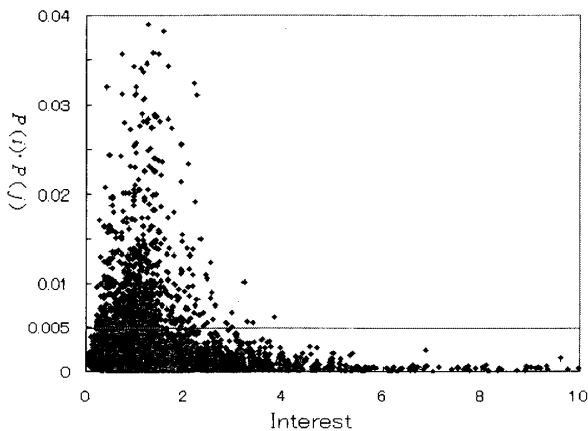


図6 $P(i) \cdot P(j)$ と*Interest*との関係

次に、店舗*i*をフロア*k*に、店舗*j*をフロア*l*に配置するように、店舗間関連度（*Interest*）を用いて、店舗の最適配置のための目的関数 f_1 を以下のように定式化する。

$$f_1 = \text{maximize } \sum_i \sum_j \sum_k \sum_l I(i, j) \cdot d(k, l) \cdot x_{ik} \cdot x_{jl} \quad (3)$$

$$\text{subject to } \sum_k x_{ik} = 1 \text{ for all } i \quad (4)$$

$$\sum_i x_{ik} = c_k \text{ for all } k \quad (5)$$

$$x_{ik} = 0, 1 \text{ for all } i, k \quad (6)$$

ここで x_{ik} および x_{jl} は店舗*i*をフロア*k*（店舗*j*をフロア*l*）に配置すれば1をとり、そうでなければ0をとる（式(6)）。一つの店舗は一つのフロアにのみ配置されるものとする（式(4)）。また、フロア*k*には c_k で表される数の店舗を配置するものとする（式(5)）。 c_k の実際の値は、図1に示されている。

$d(k, l)$ は、フロア*k*とフロア*l*の近さを表している。フロア間の近さは、距離が近いほど大きな数値を与える。本分析では、同一フロア間は1、それ以外は0と設定した。フロア間の近さは、ユーザにより変更することが可能で、例えば隣接フロア間の近さを0.5と設定するなどの工夫も可能である（表2）。

目的関数 f_1 を最大化するような店舗配置を求めることによって、正の関連の高い店舗は同一フロアに、そして負の相関の高い店舗は異なるフロアに配置されるようになる。

3.2 フロア年齢適合

次に、二つ目の目的関数 f_2 としてフロア年齢適合について考える。ユーザは、フロアごとに、ターゲットとする顧客年齢（ターゲット年齢と呼ぶ）を決める。この設定にできるだけ適した店舗を配置することを目的とする。この目的関数を用いることによって、若者向けの店舗と年配者向けの店舗が混在することを防止し、売場に統一性を持たせることができる。フロア年

表2 フロア間の近さマトリックス

	4A	5A	5B	5C	6A	6B	6C
4A	1						
5A	0	1					
5B	0	0	1				
5C	0	0	0	1			
6A	0	0	0	0	1		
6B	0	0	0	0	0	1	
6C	0	0	0	0	0	0	1

年齢適合についての目的関数を以下のように定式化する。

$$f_2 = \text{minimize} \sum_k \left(\frac{\sum_i (a_i - t_k)^2 \cdot x_{ik}}{c_k} \right) \quad (7)$$

$$\text{subject to} \sum_k x_{ik} = 1 \text{ for all } i \quad (8)$$

$$\sum_i x_{ik} = c_k \text{ for all } k \quad (9)$$

$$x_{ik} = 0, 1 \text{ for all } i, k \quad (10)$$

ここで、 a_i は店舗 i に来店したことがある顧客の平均年齢で、 t_k はフロア k のターゲット年齢を表し、 c_k は、フロア k の店舗数を表している。

この目的関数は、各フロアに配置された店舗の顧客平均年齢とそのフロアのターゲット年齢との差の二乗値の平均をフロアごとに計算し（平均年齢差）、各フロアの平均年齢差の合計値を最小化するというものである。

各フロアのターゲット年齢については、ユーザの考えに応じて調整可能であるが、今回の実験では、図7に示される値を用いた。

3.3 多目的最適化問題適用の妥当性

ここで、現状の配置における店舗関連度および各店舗の対象顧客年齢について確認することによって f_1 , f_2 による多目的最適化問題を考えることに意味があることを示す。

	A館	B館	C館
6階	60歳	55歳	60歳
5階	40歳	55歳	40歳
4階	40歳		

図7 各フロアのターゲット年齢

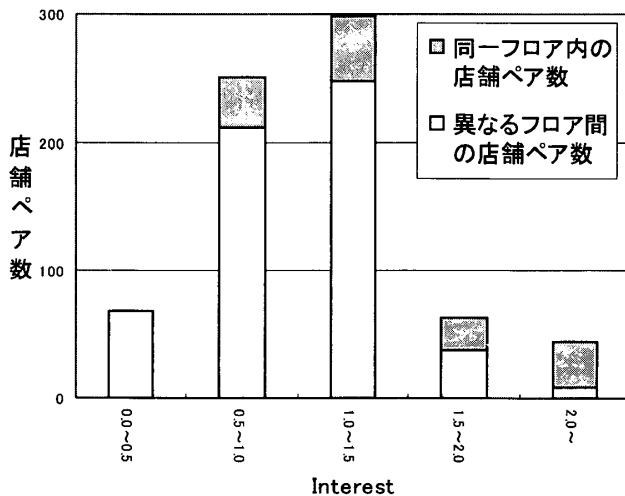


図8 Interest と店舗ペア件数の関係

図8は、式(2)で定義された Interest 別に店舗ペア数を示したものである。横軸に Interest の範囲が 0.5 刻みで示されており、縦軸は、各 Interest の範囲にある店舗ペアの数である。式(2)で定義したように、 $P(i) \cdot P(j)$ が閾値 α (0.005) より低い店舗ペアについては省いている。また、棒グラフの上部（灰色模様）は、同一フロアに配置されている店舗ペア数を表しており、下部（無地模様）は、異なるフロアに配置されている店舗ペア数を表している。

目的関数 f_1 では、店舗間関連度の低い ($Interest < 1$) 店舗ペアは異なるフロアに配置し、店舗関連度の高い ($Interest > 1$) 店舗は同一フロアに配置することを目的とするが、現状の配置を見ると、店舗間関連度が低いにも関わらず同一フロアに配置されていたり、店舗間関連度が高いにも関わらず異なるフロアに配置されている店舗ペアも多く確認できる。このことから、店舗間関連度に関する目的関数を考えることは意味あることと言える。

一方で、図9は、各フロアに配置されている各店舗に来店した顧客の平均年齢および年齢の標準偏差についてのグラフである。グラフは各フロアごとに表示されており、図中の一つの点は一店舗に対応している。各グラフのスケールはすべて同一で、横軸は平均年齢（原点 40, 最大値 70）を示しており、縦軸は年齢の標準偏差（原点 7, 最大値 15）を示している。また各グラフの右下の数値は、各フロアにおける店舗の平均年齢の幅（最大値から最小値を引いた値）を示している。

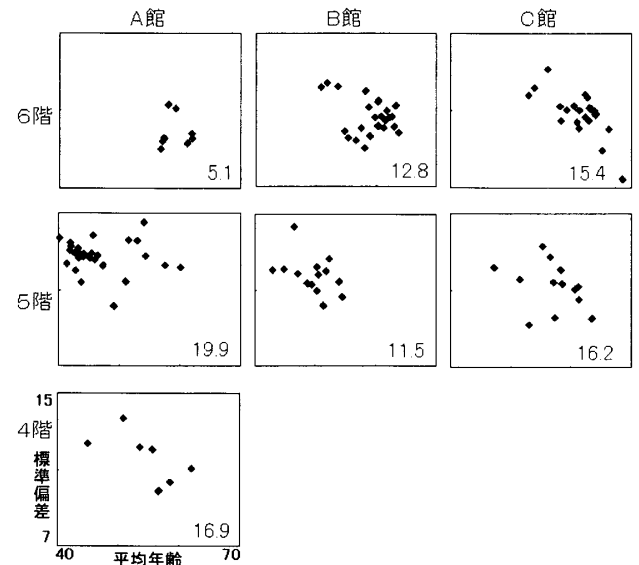


図9 各店舗が対象とする顧客の平均年齢および標準偏差

6Aのように、比較的、顧客の年齢層に偏りが見られる（平均年齢の幅が比較的小さい）フロアもある一方で、5A, 4A, 5Cのように年齢層のばらつきが大きいフロアも見られる。また同じ年齢層の店が異なるフロアに配置されていることも確認できる。このように、平均年齢だけを見ると、年齢適合度に関する目的関数 f_2 を考えることは意味あることと言える。

しかしながら一方で、年齢の標準偏差について見ると、ばらつきの比較的高い店舗も多く見られる。これは、顧客の年齢分布について、二山分布や横長の分布になっているのではないかと考えられる。このようなケースが多い場合は、顧客年齢の平均値を、店舗が対象とする顧客年齢の代表値とすることに問題となる。

そこで全店舗の年代別人数分布を確認することにした。まず、15歳代から95歳代までの5歳きざみの年代別人数割合から作られる17次元ベクトルデータを各店舗別に作成し、そのデータに対してk-means法でクラスタリングを行い、全店舗を四つのグループに分類した。それぞれのグループに属する店舗の年代別人数分布を図10に示している。図中の各折線は各店舗に対応している。横軸は年代（最左15~20歳、最右95~100歳）を示しており、縦軸は人数割合（原点0、最大値0.35）を示している。縦線は、各グループにおける平均年齢を表しており、その数値が横に示されている。またグラフの右下には各グループに属する店舗数が示されている。

グループ(b), (c), (d)は、おおむね単峰の分布を示しており、平均値によって各店舗が対象とする顧客の年齢を代表することに問題はないと言えるであろう。

グループ(a)は二山分布（25~35歳と50~60歳にピークがある）を示しているが、本ケースでは、以下に示す理由により、グループ(a)の店舗についても顧客の平均年齢によって、店舗の対象顧客年齢とすることに

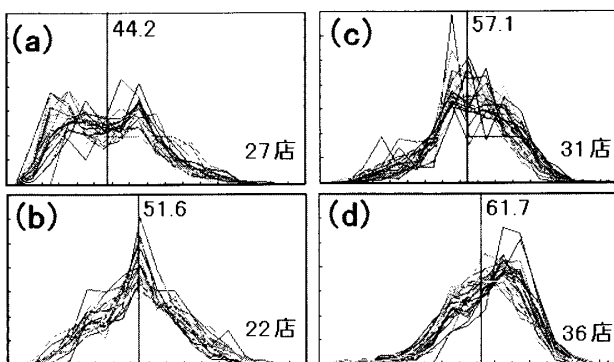


図10 年代別顧客人数分布

した。

グループ(a)に属する店舗のみ若者層の割合が他のグループに比べ多く、明らかに他のグループとは異なる年齢層を対象とした店舗であると考えられる。また、平均年齢も他のグループに比べ低く、例え平均年齢を店舗の対象顧客年齢としても、このグループに属する店舗は、ターゲット年齢の低いフロアに配置され、他のグループに属する店舗とは異なるフロアに配置されるようになる。

以上に示したように店舗関連度およびフロア年齢適合度についての目的関数 f_1, f_2 に関する多目的最適化問題を考えることは十分に意味があると言える。

3.4 遺伝的アルゴリズム

f_1, f_2 で示される二つの目的関数で表される多目的最適化問題について遺伝的アルゴリズム（多目的GA）を用いてパレート最適解（いずれの目的関数の改善も他の目的関数の劣化なしには実現できない解）を求める。

遺伝的アルゴリズムとは、生物進化の原理に着想を得たアルゴリズムであり、確率的最適化の一手法である[6]。本分析で用いた多目的GAは、文献[7]における方法に従い、ランダムに作成した異なる遺伝子を持つ個体数500からなる個体群について100世代まで交配を行った。個体選択ではランキング選択法を採用し、交叉確率0.6、突然変異確率0.05で計算した。

次に、本分析で用いた遺伝子型へのコーディング方法、致死遺伝子への対応方法、および突然変異の方法について簡単に示す。

3.4.1 コーディング

GAでは解を遺伝子として表現する。本稿における問題では、どの店舗をどのフロアに配置するかについての解を求めるので、図11に例示されるようなコーディングを採用した。

この図で各セル（遺伝子座）は各店舗を示しており、本分析では116店の店舗を扱うので、遺伝子の長さは116である。各遺伝子座には、遺伝子としてフロア番号が入る。今回の分析で対象としたフロアは七つ（4A, 5A, 5B, 5C, 6A, 6B, 6C）で、それぞれ1から7の番号に対応している。図11で見ると、左のセル（店舗）から、4A, 5B, 5B, 5A, …に配置

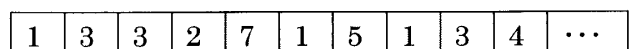


図11 解の遺伝子型へのコーディング例

することを意味する。

解の探索空間は、単純化して言えば、 7^{116} 存在し、全探索によって最適解を求めることは不可能である。そこでGAでは、ランダムに生成された複数の個体(解)から、選択、交叉、突然変異のプロセスを繰り返すことにより、最適解の探索を試みる[6]。

3.4.2 致死遺伝子

式(5), (9)で示されているように、各フロアに配置可能な店舗数には制限がある。そのため上記のような遺伝子型による表現を採用すると、個体を単純に交叉すると実行不可能解、すなわちフロア内の店舗数の制約を破る解(致死遺伝子)が生じる可能性が非常に高くなる。そこで本稿では次のような方法をとることにした。

個体 p_1, p_2 から個体 s_1, s_2 を得る交叉を考える。 s_1 は p_1 の前半の遺伝子と p_2 の後半の遺伝子を受け継ぐ (s_2 は p_2 の前半の遺伝子と p_1 の後半の遺伝子を受け継ぐ)。後半の遺伝子を受け継ぐとき、既に受け継がれた前半の遺伝子によってフロア配置可能店舗数 (c_k) を超えている場合は、その遺伝子座を無視し、フロア配置可能店舗数の制約を破らない遺伝子座のみ p_2 の後半の遺伝子 (s_2 については p_1 の後半の遺伝子) から受け継ぐ。その後、まだフロアを設定していない遺伝子座については、配置可能フロアをランダムに設定する。以上の方法により、できるかぎり親の遺伝子 p_1, p_2 の特徴を保ちながら、かつ致死遺伝子を持たない個体 s_1, s_2 を生成させている。

3.4.3 突然変異

各フロアに設定された配置可能店舗数の制約を満たしながら突然変異を行うために、本ケースでは、次のような方法を採用した。

交叉の結果得られた新たな個体について、0.05 の確率で突然変異を行う個体が選ばれる。次に、その個体において、ランダムに二つの遺伝子座を決定する。そして、それら二つの遺伝子座の遺伝子を交換する。

4. 分析結果

以上に示した分析手法を、百貨店の婦人関連部門の顧客購買データに当てはめ、その結果得られたパレート最適解が図12に示されている。

横軸は目的関数 f_1 (店舗間関連度: 値が高いほど最適化されている) を表し、縦軸は f_2 (年齢適合度: 低いほど最適化されている) を表している。また左上隅に離れて示された解 a (■印) は、現在の店舗配置に

よるものである。いずれのパレート解も、現在の店舗配置より、二つの目的関数の値について優れていることがわかる。

ユーザは、目的に応じて、二つの目的関数の比重を検討し、パレート最適解から適切な解を選ぶことができる。ここでは、 f_1, f_2 ともある程度優れた解 b (▲印) および解 a について考察を進めていく。

図13, 14は、各フロアに配置された各店舗が、二つの目的関数の最適化にどの程度貢献しているかを、解 a, 解 b のそれぞれについて視覚化したもので、各フロアに配置された店舗がプロットされている。すべてのグラフのスケールは同一である。

横軸 (原点 0, 最大値 50) は各店舗が目的関数 f_1 (店舗間関連度) の最大化にどの程度貢献したかを示しており、店舗 i の店舗間関連貢献度 IC_i は式(11)で示される値である。これは、目的関数 f_1 (式(3)) における店舗 i を固定したときの項を表している。この値が原点から遠いほど貢献度は高い。

$$IC_i = \sum_j \sum_k \sum_l I(i, j) \cdot d(k, l) \cdot x_{ik} \cdot x_{jl} \quad (11)$$

また、各グラフ内に示された縦線は、あるフロアに配置された任意の店舗 i について、式(11)においてすべての j に対して $I(i, j) = 1$ としたとき (つまり、すべての店舗間は、正の関連も負の関連もない) の値 IC_i^* を示している。例えば、6階 A 館では $IC_i^* = 7$ である。グラフ上にプロットされた店舗 i について、 $IC_i < IC_i^*$ ならば、店舗 i の f_1 への貢献度は低く、逆に $IC_i > IC_i^*$ ならば、店舗 i の f_1 への貢献度は高いことになる。

一方で、縦軸 (原点 0, 最大値 600) は目的関数 f_2 (年齢適合度) の最小化にどの程度貢献したかを示しており、店舗 i の年齢適合貢献度 AC_i は式(12)で示さ

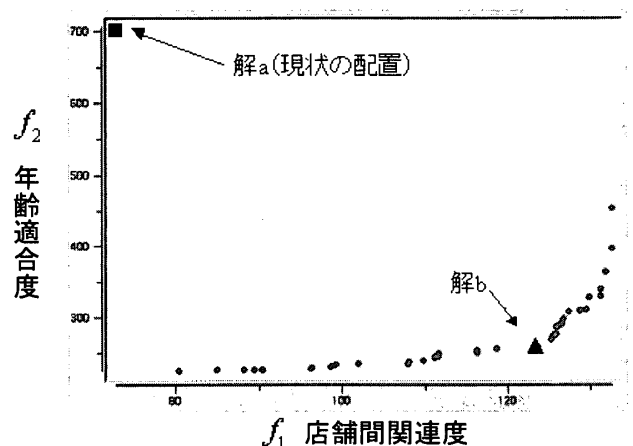


図12 パレート最適解

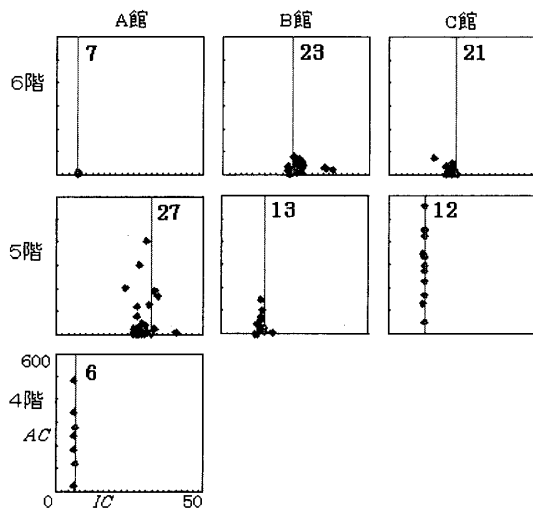


図13 解aのフロア別貢献度

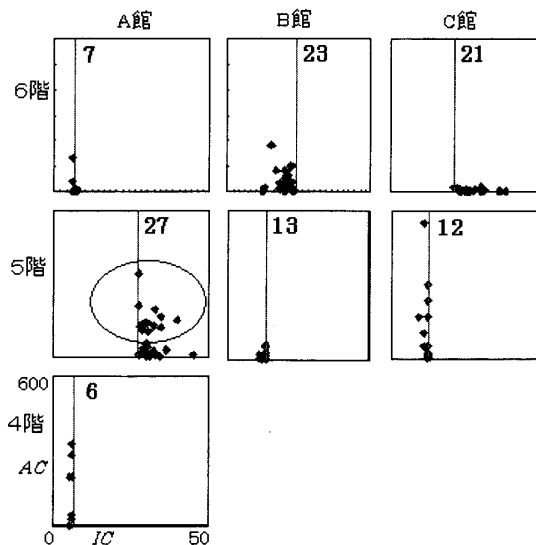


図14 解bのフロア別貢献度

れる値である。これは、店舗*i*の顧客平均年齢と配置フロアのターゲット年齢との乖離度を表している。この値が原点に近いほど貢献度は高い。

$$AC_i = \sum_k (a_i - t_k)^2 \cdot x_{ik} \quad (12)$$

解aと解bの図では、店舗の配置が全く異なってくるため、各フロア間の直接の比較はできないが、両目的関数に貢献している店舗が多くなっていることは確認できる。例えば、解aには6B、5Aに店舗間関連貢献度の高い店舗がいくつか見られるだけだが、解bには6C、5Aに店舗間関連貢献度の高い店舗が多く配置されている。また年齢適合貢献度についても、解aでは、5C、4Aに年齢適合貢献度の低い店舗が多く見られるが、解bにおいては、比較的その数が少なくなっている。

解bにおける6Cのフロアに属するすべての店舗の年齢適合度への貢献は高く、また店舗間関連度への貢献度が高い店も多く、理想的な構成と言える。

解bの5Aのフロアは、店舗間関連度の高い店が多いが、一方で年齢適合度が低い店も多い（丸枠で囲われた店舗群）。このような店舗群は、同一フロアに配置すべきかどうかは、その他の情報を加味して検討することが必要となるであろう。

その他のフロアはすべて、店舗間関連度に対する貢献度の高い店がほとんど見受けられない。これは次のことが原因していると考えられる。節3.1で述べたように、*Interest*を式(2)で定義することにより、同時訪問確率 $P(i) \cdot P(j)$ が低い店舗間の関連度は無視されているため、そのような店舗は、年齢適合度でのみ配置が決まる傾向にある。このようなケースでは、節3の最初に述べたような、その他の目的関数を導入することを考えるべきであろう。

5. むすび

本稿では、百貨店において魅力ある店舗配置を決めるための手法について、バスケット分析や遺伝的アルゴリズムといった技術を利用することによって示してきた。そして、その手法を、実際の百貨店における顧客購買履歴データに適用し、その妥当性について見てきた。

本手法を利用することで、定式化された目的関数の枠内でのみパレート最適解を導出可能であるが、実際に現場で利用する際には、最適解としてではなく、むしろ参考情報としてとらえるべきで、魅力あるフロア構成を考えるための一つのツールとして考えるべきであろう。

また、本稿では、紙面の都合上、店舗間関連度とフロア年齢適合の2目的のみを取り上げたが、我々が開発したシステムでは二つ以上の目的が設定されていても、パレート解を導出できるようになっている。

本稿で提案した手法を、実際の企業で利用するに当たっては、今回利用したデータには含まれていなかった店舗位置や店舗面積の情報も重要となるであろう。またコストの面から見れば、店舗の移動に伴う引越コストも無視できない。さらに、今回は婦人関連部門のみを分析対象としたが、百貨店全体としての魅力を考えると、全部門を対象とする必要が出てくる。今後は、今回の研究をベースに、これら様々な問題点を考慮に入れ、より実用性の高い手法を構築していきたいと考

えている。

参考文献

- [1] Agrawal, R., et al.: Mining association rules between sets of items in large databases. Proc. of 1993 ACM-SIGMOD Int. Conf. on Management of Data, pp. 207-216, 1993.
- [2] M. J. A. Berry, et al.: *Data Mining Techniques: From Marketing, Sales, and Customer Support*, John Wiley & Sons, 1997.
- [3] S. Brin, et al.: *Beyond Market Baskets: Generalizing Association Rules to Correlations*, Proc. of the ACM-SIGMOD, pp. 265-276, 1997.
- [4] P. Cabena, et al.: *Discovering Data Mining*, Prentice Hall, 1998.
- [5] 江尻弘:『百貨店のID付きPOSデータのマーケティング・マーチャングアイジングへの活用に関する提案』, 日本マーケティング・サイエンス学会第2次ID-POS部会資料, 2002/9.
- [6] 北野宏明:『遺伝的アルゴリズム1』, 産業図書, 1995.
- [7] 玉置久:『遺伝的アルゴリズムと多目的最適化』, 遺伝的アルゴリズム2(北野宏明編), 第3章, 産業図書, pp. 71-87, 1995.
- [8] 松岡真宏:『百貨店が復活する日』, 日経BP, pp. 188-202, 2000.
- [9] G. Piatetsky-Shapiro (ed.), *Knowledge Discovery in Databases*, AAAI Press, 1991.
- [10] P.-N. Tan, V. Kumar, and J. Srivastava: *Selecting the Right Interestingness Measure for Association Patterns*, Proc. of the Eighth International Conference on Knowledge Discovery and Data Mining, pp. 32-41, 2002.