1　はじめに

マーケティングについて，アメリカ・マーケティング協会 (AMA) によれば「マーケティングとは，顧客に向けて価値を創造，伝達，提供して，組織と組織の利害関係者に利するように顧客との関係性を管理する組織の機能及びその一連のプロセスである」と定義されている (American Marketing Association ウェブサイト, 2011). Kotler and Keller (2008) は，このマーケティングの目的を達成するためには，マーケティング・マネジメントが発生すると述べている．彼らはマーケティング・マネジメントを「ターゲットとなる市場を選択し，優れた顧客価値を創造・提供・伝達することにより，顧客を獲得・維持・育成していく技術と科学」と定義している．市場や消費者の多様化については，近年様々な場面で言及されてきている．マーケティング・マネジメントにおいて，こういった多様化対応するためには，消費者の行動や嗜好を理解し，究極的には個々の消費者のニーズや問題解決に適切な財やサービスを提供しなければならない．ただし，一部を除けば個々の消費者・顧客のニーズに完全に満たすためには多大なコストを必要とするため，多くの場合，同質と考えられる相応のサイズのグループをマーケティング・プログラムの対象とする．つまりセグメント・マーケティングが現実的である．

STP4P (Segmentation, Targeting, Positioning; Product, Price, Place and Promotion) に代表されるマーケティング・プロセスは，マーケティング活動を円滑にかつ効率的に進めるためのフレームワークであり，特にセグメンテーションはその第一歩である．

つまり，企業がマーケティングの内部統制要因である 4P の戦略を決定する前段階として，STP つまりまず適切なセグメンテーションを通して，自社のターゲットを明らかにし，自社の戦略ポジショニングを決定することが重要となる（図 1）．

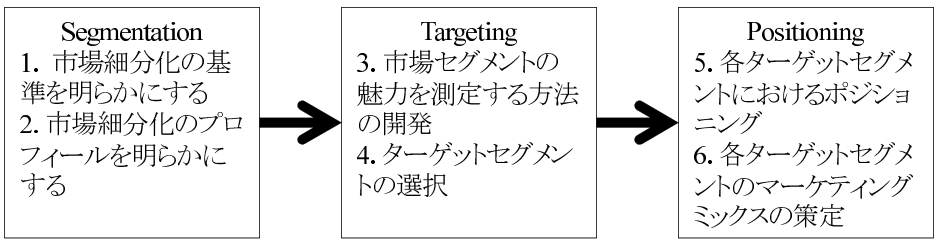

図 1　STP の段階（Kotler and Armstrong, 2009 より）

適切なセグメンテーションとターゲティングは，そのセグメントに属する消費者

に対して効果的に経営資源を投下することができ，そのことで，ターゲット消費者との長期にわたる良好な関係を築いて行くことが期待できる．

2　セグメンテーションの基準

セグメンテーションは企業がそのビジョンやコンセプトを明確に伝えたいターゲットとなる消費者を特定するために，顧客を適切な基準で同質を考えられる有効な大きさのグループに分ける手法の総称である．

図 2 は市場の細分化のレベルを示したものである．Kotler and Armstrong (2009) によれば，マス・マーケティングはその名の通り，市場全体を一つとみてセグメンテーションを行わない状態である．対極のミクロ・マーケティングは究極のセグメントであり，各顧客にまで分割することであり，完全な細分化である[1]．しかし，実質的なマーケティング戦略を考えると，一部の例外を除き一つのグループの大きさがあまりに小さいセグメントは機能しない．

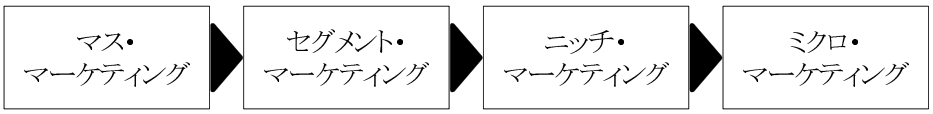

図 2　市場セグメントのレベル（Kotler and Armstrong, 2009 より）

セグメント・マーケティングはセグメントを作るのが目的ではなく，セグメントを特定し，ターゲットを決定するために行われる．マス・マーケティングで対象とする市場よりサイズは小さくなるが，同質と考えられる集団を形成できるため，企業はより効果的な商品やサービスを，適正な価格でターゲット・セグメントに提供していくことができ，さらに競合他社の動きに応じたマーケティング活動の調整も可能である．

ニッチ・マーケティングは，さらに狭いセグメントであり，明確なベネフィットの組合せを望むグループである．消費者の望むベネフィットは多くの場合は潜在的な意識により決定するものであり，観測することは一般には困難を伴う．

Kotler and Keller は有効なセグメントが満たす要件として「測定可能性」「実質性」「到達可能性」「実行可能性」の 4 つを挙げている．このうちデータ分析にあたって中心的課題になるのは，前半の 2 つである．

[1] ワン・トゥ・ワン・マーケティングとも呼ばれる．

「測定可能性」はターゲットとするか否かの決定を実際に行うことができるかどうかを意味する．後述する変数のうちたとえば地理的変数であれば，顧客の居住地や居住地の属性などにより，各消費者がターゲットに入るかどうかを明確に判定することができる．一方で，店頭で見た目で年齢を判断する場合には多少の錯誤の可能性は否めない．いずれにしても，ある基準により，消費者が自身のターゲットに入るかどうかが判定できなければならない．

一方,「実質性」は，マーケティング・プログラムを実行できるだけの大きさがなければならないことを示唆している．伝統的な百貨店の外商部のようにコンシェル・ジュサービスを顧客全体に展開していくことは，実質的には不可能であり，現実的には，共通のマーケティング・プログラムを相当の大きさを持つセグメントに効果的に投下していく．セグメントの数があまりに多くなりすぎると，セグメントごとにマーケティング・プログラムを個別に用意することは困難である．分析上は，同質なセグメントをどのように同定するかという問題に帰着されるが，多くの場合は情報量基準などを利用した最適なセグメント数決定問題についても言及しなければならない．ただし現実的には，セグメントの特徴をみながら，セグメントを決定するということも一般的である．

消費者のセグメントの基準は様々であるが，現在広く使われている変数を表 1 にまとめる．

表 1 のうち，地理的変数や，人口統計的変数は，いわば個人の外的変数であり，データはすでに存在もしくは入手は容易なものが多い．しかし，行動変数は測定の仕組みが必要であり，心理的変数にいたっては，消費者自身でもはっきりと理解していない場合もある．したがって，観測可能性の観点からいえば，地理的変数,，人口統計的変数は把握しやすく，心理的変数，行動変数はそれらに比べて観測が困難である．しかし，消費者の消費行動という観点からいえば，測定困難な要素ほど消費者心理に近づくことができる．

さらに，Wedel and Kamakura (2000) によると，観測の可能性と範囲に言及したマーケット・セグメンテーションのための区分変数を表 2 のように整理している．

表 2 はマーケティングの対象と変数の関係を示しているが，この表からもセグメンテーションのを必要とする対象や目的により，利用される変数が異なることがわかる．

3　セグメンテーションのための分析手法

これまで述べてきたように，セグメンテーションそのものは，マーケティング分野においては重要な位置を占めているが，基準や分析手法については統一的な見解というのは存在せず，それぞれの主体や目的，消費者特性に合わせた手法が選択され

表 1　セグメントの基準変数（Kotler and Keller, 2008 より）

地理的変数	
地域	太平洋沿岸，山岳部，北西中部，北東中部，南東中部，南部大西洋沿岸，中部大西洋沿岸，ニューイングランド
人口規模	5000 人未満，2 万人未満，5 万人未満，10 万人未満，25 万人未満，50 万人未満，100 万人未満，400 万人未満，400 万人以上
人口密度	都市，郊外，地方
気候帯	北部，南部
人口統計的変数	
年齢	6 歳未満，12 歳未満，20 歳未満，35 歳未満，50 歳未満，65 歳未満，65 歳以上
世帯規模	1〜2 人，3〜4 人，5 人以上
ライフサイクル	若い独身，若い既婚子供なし，若い既婚末子 6 歳未満，若い既婚末子 6 歳以上，高年既婚子供あり，高年既婚 18 歳未満こどもなし，高年独身
性別	男性，女性
所得	1 万ドル未満，1 万 5 千ドル未満，2 万ドル未満，3 万ドル未満，5 万ドル未満，10 万ドル未満，10 万ドル以上
職業	専門職，技術職，経営者,，事務員，販売員，工員，熟練工，農業従事者，退職者，学生，主婦，無職
教育水準	中卒，高校中退，高校卒，大学卒
宗教	カトリック，プロテスタント，ユダヤ教，イスラム教，ヒンズー教，そのほか
人種	白人，黒人，アジア系，ヒスパニック系
国籍	北アメリカ，南アメリカ，イギリス，フランス，ドイツ，イタリア，日本
社会階層	最下層，下の上，労働者階級，駐留，中の上，上の下，最上流
心理的変数	
ライフスタイル	文化志向，スポーツ志向，アウトドア志向
パーソナリティ	神経質，社交的，権威主義的，野心的
行動変数	
オケージョン	日常，特別
ベネフィット	品質，サービス，経済性，迅速性
ユーザーの状態	非ユーザー，元ユーザー，潜在的ユーザー，初回ユーザー，敵的ユーザー
使用頻度	ライト，ミドル，ヘビー
ロイヤルティ	なし，中程度，強い，絶対的
購買準備段階	認知なし，認知，情報あり，関心あり，購入希望，購入意図
製品に対する態度	非常に肯定的，肯定的，無関心，否定的，非常に否定的

表 2　マーケット・セグメンテーションのための区分変数

	一般的変数	製品・店舗固有の変数
観測可能	文化，地理変数， デモグラフィクス変数 社会・経済変数	使用頻度，ブランド・ロイヤルティ 店舗ロイヤルティ，採用時期， 消費場面
観測不能	パーソナリティ，生活価値， ライフスタイル	心理的属性，便益，知覚，弾力性， 購買意図

る．また，具体的な分析手法については，学術的な成書はあまり多くなく，Wedel and Kamakura (2000) や中村 (2008) など少数に限られている．そのほかはマーケティングにおけるデータ分析について述べられた著書の中の一部で紹介されているに過ぎない．Wedel and Kamakura (2000)，中村 (2008) で実際に紹介されている手法は表 3 の通りである．

以下で，いくつかの代表的な方法について論じる．

3.1　教師なしデータによるセグメンテーション手法

教師なしデータとは，外的基準がない，つまり共変量で説明される結果のデータがないデータを意味する．これに対して，たとえば購買の有無が事前に測定できており，購買の有無に影響のある変量を特定してその変量によりセグメントを作成するというような場合には，購買の有無が教師データとなる．

3.1.1　集計によるセグメント

従来より，顧客が企業にもたらす利益は顧客ごとに異なり，したがって，マーケティング・プログラムを顧客に合わせて変更するべきであると考えられてきた．そのために RFM 分析や ABC 分析に代表される過去の履歴を集計して顧客の価値によるセグメントを作る方法が利用されてきた．これらの方法は特に，優良顧客を選別するために使われてきた．ABC 分析はパレート分析とも言われ，ある一定期間の顧客ごとの購買金額をや購買回数など，ある基準に従って顧客の購買行動を集計し，その値の高い順に並び替えをする．そして，上位を優良顧客と特定することでマーケティング・プログラムを効果的に実施し，これら優良顧客の囲い込みと離反防止を行おうとする手法である．ただし，どういった購買顧客が優良であるかは，様々な見方が考えられる．

RFM 分析はある時点から一定期間さかのぼった顧客の購買履歴を多面的に集計することで，顧客をスコアリングする手法である．RFM 分析では，次の 3 つの基準で顧客の購買行動を集計する．

表 3　文献で紹介されているセグメンテーション手法

Wedel and Kamakura (2000)	
Clustring Methods	Hierarchical Clustering Methods
	Non Hierarchical Clustering Methods
	Overlapping and Fuzzy Methods
Mixture Model	
Mixture Regression Model	
Mixture Unfolding Model	
Profiling	
Dynamic Segmetation	Markov Model
	Latent Change
中村 (2008)	
階層型クラスター分析	
多次元尺度構成法	INDSCAL，ALSCAL，PREFMAP
自己組織化マップ	
判別分析	
非階層型クラスター分析	k-means
決定木分析	C4.5　CHAID
ロジット分析	
潜在クラス分析	Finite Mixture Model
	Poisson Mixture Model
RFM 分析	
ネットワーク分析	

- Recency：直近購買日
- Frequency：累積購買回数
- Monetary Value：累積購買金額

つまり，RFM 分析においては優良顧客を「より最近来店し」,「より回数を重ね」,「より多額の購買」をした客として定義していることになる．集計そのものは，システムとしても困難なものではないため，小売店での導入事例もいくつかあり，また，たとえばジェリコ・コンサルティングはスコアリングからクラス化をおこなう「RFM セルコード ®」を商標登録しており，各セルの顧客の特徴とマーケティング対応について言及している (Hugees, 2006).

3.1.2　クラスター分析

クラスター分析は，消費者を説明する変数の近接度をもとにグループにまとめていくと手法で，大きくは階層型クラスター分析と非階層型クラスター分析に二分される．前者は各サンプルおよび構成されたグループ間の距離を測り，近い順にグループに含めて大きなグループを段階的に作っていく[2]．この操作を繰り返してセグメントに分ける手法である (田中・垂水, 1995).

一方後者は，あらかじめ複数与えられたグループの中心値に対して各サンプルからの距離を測り，所属するグループを決定する．代表的な方法として k-means 法がある (Bradley and Fayyad, 1998)．k-means 法では，所属サンプルをもとに平均値を新たな平均値を求め，グループの中心値を更新する．さらに更新された中心地に対して再度所属するサンプルを決める．これを収束するまで繰り返す.

前者は毎回サンプルおよびグループの全組み合わせについて計算しなければならないのに対して，後者は各中心値までの距離のみを計算すればよく，また一般には収束が速い．計算量としてはサンプル数を n，求めるクラスター数を k とすると，階層型クラスター分析が $O(n^2)$ であり，k-means 法は $O(kn)$ である．したがって，サンプル数が多くなると，階層型クラスター分析は計算時間の面で k-means 法に遠く及ばない.

クラスター分析に共通する問題としては，最適なクラスター数の決定が挙げられる．この問題に対してはいくつかの指標が提案されているたとえば，Calinski and Harabasz (1974) では，グループ内の変動とグループ間の変動の比が最大となる場合が望ましいとしている．また，Scott and Symons (1971) は，データ全体の共分散行列とクラスタごとの共分散行列の行列式の比（の対数）により，結果の良さを評価している．他にもいくかの指標が提案されているものの，標準的な方法は存在しない.

3.1.3　固有値問題に帰着される手法

従来から複数の変量を持つデータ分析に利用されている多変量解析の分析手法の中で，セグメンテーションに比較的よく使われてきたものは，主成分分析，因子分析，コレスポンデンス分析，数量化理論 II 類，数量化理論 III 類などである．これらはすべて固有値問題に帰着される．上記のうち因子分析を除いた分析手法は，観測変数の相関関係をもとに，変量の次元縮約をしようというものである[3]．共通の考え方としては，観測変数間の共分散行列もしくは相関係数行列を出発点とし，固

[2]逆に，最初すべてのサンプルを一つのグループとし，距離の度合によって順に分割する方法もある.

[3]形式的には MDS（Multidimensional Analsysis：多次元尺度構成法）もこれらと同様の使われ方をするが，MDS では対象間の距離を少数次元でできるだけ再現しようとする方法であり，アルゴリズムや変数間の差異の考え方は全く異なる.

有問題に帰着することで，説明力の高い[4] 合成変数をうまく得ることにより，元のデータに対して説明力の高い少数の合成変数を得る．これら少数の合成変数により，変数間，サンプル間の差異をはっきりさせることができる．

古くから広く用いられてきた因子分析は観測変数の裏側にある潜在的因子を仮定し，潜在因子と観測変数との関係を論じるもので，一般的な方法としては反復主因子法のように収束するまで反復して解くことで安定した解を求めようとする．さらに，因子を回転させることで，単純構造を持つより解釈しやすい因子を得るということも一般にはなされている．これらの方法を用いて低次元の空間に布置された各サンプルは，しばしばクラスター分析などを通じて少数のセグメントに分割される．

3.1.4　自己組織化マップ

データマイニングは大量データから有用な情報を引き出す分析手法の総称である (Wu and Kumar, 2009, Han and Kamber, 2001)．一つの特徴としては，集計によらないサンプルを直接利用する点にあるといえる．前述のクラスター分析もこの特徴によればマイニングの範疇に入るといえる．データマイニングが注目されるようになった背景には，計算機環境の劇的な進化もさることながら，より粒度の細かいデータが大量に蓄積されたこと，そしてなによりもマーケティング分野においては，消費者の多様化から，個々人に目を向けていかなくてはならなくなったことが挙げられる．

自己組織化マップはニューラルネットワークを応用した多次元のデータを 2 次元平面上に射影する方法である (Kohonen, 2000)．2 次元地図上に連結された複数のセルを用意し，同じ特徴を持つサンプルを同じセルに当てはめるように関数を定める．この時さらに隣接するセルが近い性質を持つように割り当てられるのが自己組織化マップの特徴である．つまり同じセルに入るサンプルが同質と考えられる以外に，より特徴が近いと考えられるグループがなるべく近い位置に求められる．したがって，変数間の相関関係だけでなくより詳細な購買の関係をセル単位並びに近接セルの比較により行うことができる（図 3）．

3.2　潜在クラス分析

これまでに論じた分析手法は，結果として各サンプルの所属セグメントが排他的に決定する．これは各セグメントに所属するサンプルは，ほかのセグメントに所属するサンプルと比較すると明らかにお互いが類似していることを仮定している．しかし，消費者の異質性は本来は個別のものであり，少数の特徴を持つ集団に明らかに

[4]共分散行列もしくは相関係数の対角項の和がデータ全体の分散を示す．対称な非負定値行列であるので，固有値分解することで，分散の大きい合成変数を求めることができる．

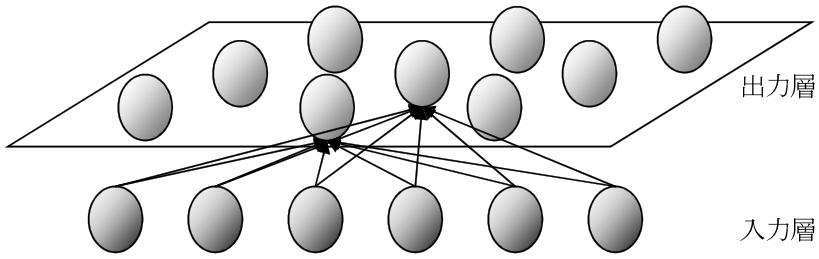

図 3　自己組織化マップの概要

分割できるという仮定は，この個人の異質性を十分に反映できない．そこで，複数のセグメントを個人と特徴の代表値として求め，個人はその複数の特徴を合わせて持つ，つまり重複してセグメントに所属することを許すことで個人の異質性を考慮することを考える．

このような方法の考え方は因子分析である．因子分析は，サンプル全体で共通する少数の潜在的な因子を特定し，各サンプルの観測値はその共通因子から個人ごとに異なる影響を受けて得られると仮定される．ただし，因子分析も主成分分析と同様に変数間の共分散行列もしくは相関係数行列をもとしてに求められるものであり，実質的には個人差は結果として後から推定される．

これに対して各サンプルで観測されるデータから複数の特徴の異なるセグメント（クラス）を求め，それらのセグメントに所属する確率を求めることで異質性を表現しようとする手法が潜在クラス分析である．ある意味では，因子分析に近い手法とも言えなくはないが，サンプルを直接利用しようという点が異なることと，因子分析が一般には量的変数のみを対象にするのに対して，潜在クラス分析では，質的変数も容易に扱うことができる[5]．

サンプル x が y に対して関係がある確率 $p(x,y)$ の潜在クラスモデルは一般には以下のように表される．

$$p(x,y) = \sum_s \pi_s(x) p_s(y) \tag{1}$$

$\pi_s(x)$ は x がセグメント s に所属する確率であり，$\pi_s(y)$ はセグメント s が y と関係する確率である．各 x について $\sum_s \pi_x(s) = 1$ であり，各 x はセグメントに対して各々異なる確率で所属する．共変量を用いる場合は，これらの確率それぞれに確率選択モデルを当てはめることも行われる．

[5]質的変数についてもカテゴリ化すれば

このほかにも，k-means 法をファジィ測度に拡張した，Fuzzy c-means も潜在クラス分析と同様の方法と言える (Bezdek, 1981).

3.3 教師付データによるセグメンテーション手法

以下では，教師付データをもとにしたセグメンテーション手法について概観する．これらの手法の特徴としては，あらかじめ答え（外的基準）がわかっているデータについて，属するクラスを左右する説明変数を選択し，その変数の影響を分析しようというものである.

3.3.1 判別分析

教師データが所属クラスであるようなデータの関数関係を求める分析を総称して判別分析というが，ここでは線形判別分析に限定する．クラスが 2 つである時を二群判別分析といい，それ以上の場合を多群の判別分析という．二群判別分析は回帰分析の枠組みで分析することが可能であり，多群判別分析の場合は正準相関分析により分析可能である (田中・垂水, 1995)．そのほかにもマハラノビスの汎距離による判別分析などもあるが，いずれにしても，説明変数と教師データとの間の関係は線形式であることを仮定している[6] .

3.3.2 決定木分析

決定木分析は，教師データのクラスを分割するもっとも妥当な変数を一つ選択し，その閾値を同時に決めることで，クラスの異なるサブグループを作る手法である．分割されたサブグループでも十分に判別ができない場合は，そのサブグループについて再度決定木アルゴリズムにより分割変数と閾値を決めることで，再度分割する．これを繰り返すとトーナメント表を逆にたどるイメージで，根から木が生えていくように分割が進められるため決定木と分析呼ばれる.

アルゴリズムとしては，χ^2 値をもとに分割基準を決める CHAID (Kass,1980) や，情報量基準をもとにした C4.5/C5.0/See5 などがある (Quinlan, 1993, 1996)．すべての変数と変数の値すべてについて分割基準としてふさわしいかを判定していく必要があり，すべての場合を考慮するためには相当数の組合せを試行しなければならない.

[6]ただし，二次判別分析なども提案されている.

3.3.3 Support Vector Machine

Support Vector Machine (SVM) は，数理計画法による判別分析である．最大の特徴としては，観測データの計量空間で直接判別のための関数を求めるのではなく，特徴空間という元の値から見ると非線形の空間上にデータを射影して判別が行われる．別の言い方をすれば，最もうまく分割できるような非線形空間上でのウェイトを求めることになる．その意味でも，SVM は非線形の判別分析である．もっとも一般的な線形カーネルは

$$f(x, y) = x \cdot y \tag{2}$$

として表される．ただし x は説明変数であり，y は外的変数，つまり従属変数である[7]．このカーネル関数を用いて，次の問題を解く．ただし，スラック変数 ξ_i を導入してある程度のご判別を許容するソフト制約が入ったモデルである．

$$\begin{aligned} \min \quad & \frac{1}{2}|\boldsymbol{w}|^2 + C\left(\sum_i \xi_i\right)^2 && (3)\\ \text{s.t.} \quad & y_i(\boldsymbol{w} \cdot \boldsymbol{x}_i + b) > 1 - \xi_i,\ \forall i && (4)\\ & \xi_i \geq 0,\ \forall i && (5) \end{aligned}$$

この問題では最小マージンは次の式で与えられる．

$$\min_i \frac{|\boldsymbol{w} \cdot \boldsymbol{x}_i + b|}{|\boldsymbol{w}|^2} \tag{6}$$

したがって，(6) 式を最大化することは，(6) 式の分母を最小化することになる．各サンプルの所属クラスは上記の最適化問題を解くことで得られる（図 4）．

3.4 それぞれの手法の特徴

以上で紹介した分析手法の特徴について以下にまとめる．

RFM 分析は簡便な手法であり，購買実績を直接集計するという意味では，最も理解しやすくまた顧客のセグメントの特徴もはっきりしている．しかし，顧客が何を購買しているのか，また，なぜ購買するのかといった背後の情報は使われない．顧客の購買行動について背後の因子の抽出がしづらい場合や，購買パターンが比較的はっきりしている場合などは，RFM が有効に働くこともある．

固有値問題に帰着される方法は解は一意に定まり，かつ解釈もしやすく，もっとも広く用いられている手法である．その反面，これらの方法のもっとも大きな問題

[7] 外的変数は 2 クラスであることが基本であるが，多クラスに拡張されたモデルもある．

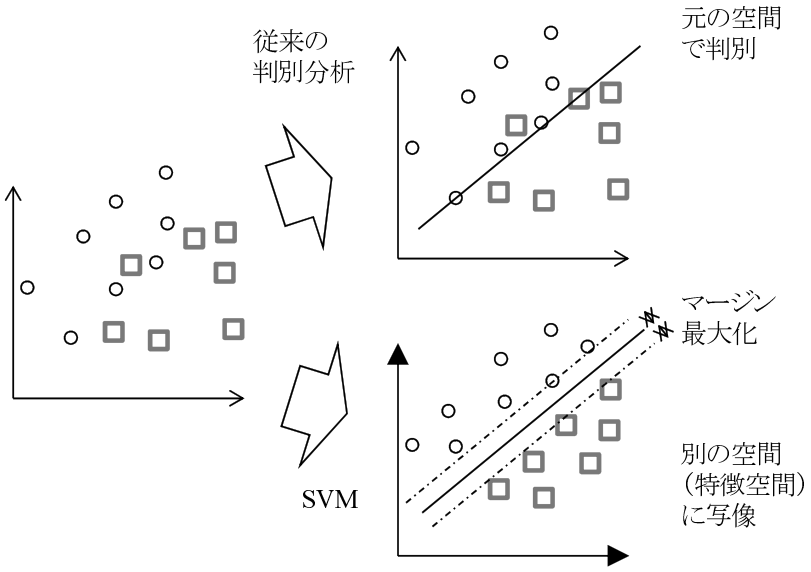

図 4　Support Vector Machine の概念

は，たとえば二軸で結果を表現する場合，固有値問題から得られる第一軸と第二軸の間には本来なんら計量的な関係がないのにもかかわらず，あたかも二次元のユークリッド空間であるがごとく扱う点である．実際には，布置された座標についてサンプルをクラスター分析によってセグメントに分けるということが行われることが多いが，本来は任意の二点間の距離は測定できない．こうした問題に対しては，MDS などを利用することもできるが，分析パッケージが充実しているとはいえず，また，MDS でも対象間の距離は与えなければならないので，性質の異なる複数の変数を用いて，どのように一元的に距離を与えるかについては，統一的な方法があるわけではない．ただし，現実的には「理解しやすい」従来の多変量解析が用いられる．

データマイニングの範疇に入る自己組織化マップや k-means 法といったマイニング，潜在クラスモデルなどは，個々のサンプルを扱おうという手法であり，多変量解析手法と異なり集計値によらない方法なので，各サンプルの解釈が可能である．しかし，計算プロセスは複雑で，最適解を求めるのではなく，EM アルゴリズムなどに代表される反復計算により妥当な解を求める方法が一般的である．したがって，初期値やアルゴリズムに解が依存してしまい，常に同じ解が求まるわけではない．また，解がなぜ得られたかについて求解過程の背後が明確でない場合も多く，客観的

な解釈が難しいことも少なくない．

教師付データを用いた分析では，変数が結果にどのように影響を与えるかを求めることができる点が特徴である．しかし，SVM などの手法ではどうしてそのような解が求まるのかという背後の意味づけは簡単ではないため，場合によっては使いづらい．また，マイニングの諸手法は一般に非線形問題をを扱っているため，パラメータ推定に用いる学習データに過剰にフィッティングさせるために，検証用データではまったく異なる傾向が得られる場合もある．こういった問題を克服するために，交差検証をおこなったり，繰り返しパラメータを求めて過学習を抑えるアルゴリズムなどが用いられることが一般的である．

4　実データによる分析例

これまでに述べたセグメンテーション手法のいくつかについて，小売店の実際の ID 付 POS データを用いて分析し，セグメントの特徴を述べる．本稿では，特に顧客のセグメンテーションを中心に行う．

4.1　データ概要

本稿で用いるデータは，日本の地方都市にある食品スーパーマーケットの ID 付 POS データである．セグメンテーションのために用いる期間は 2005 年 7 月 1 日〜31 日の 1 カ月である．分析にはもっとも大きな購買カテゴリ分割（部門）を用いた．このデータでは 24 部門に分かれる．データの概要は以下の通りである．

- 顧客数　10,448 人
- 総顧客数　56,640 人
- 購買金額合計　130,819,838 円
- 一回当たり客単価　2,310 円
- 平均購買回数 5.4 回

部門と金額シェアを表 4 にまとめる．

4.2　RFM 分析

まず，顧客行動について RFM 分析のための集計を行った．セグメントを作るために，R, F, M の値それぞれについて，独立におおよそ 1/3 ずつになるように 3 分割した．そして R, F, M のクラスの組み合わせでセグメントに分割した．したがって，$3 \times 3 \times 3 = 27$ のセグメントが作成できる．各クラスの閾値は表 5 の通りである．

表 4　部門と販売金額シェア

部門名	野菜	果物	水産	酒
金額シェア	8.4%	4.5%	7.6%	5.3%
部門名	和日配（漬物）	農産乾物	食肉	加工肉
金額シェア	4.1%	2.6%	3.0%	0.0%
部門名	雑貨	惣菜	和日配（水物）	和日配（冷食）
金額シェア	7.3%	0.9%	8.2%	11.1%
部門名	和日配（練製品）	洋日配（乳製品）	洋日配（パン）	調味料
金額シェア	1.8%	0.9%	2.4%	5.4%
部門名	嗜好品	菓子	塩干	米
金額シェア	1.2%	1.3%	0.7%	6.7%
部門名	たばこ	衣料	洋日配（バター・チーズ）	牛肉
金額シェア	3.2%	3.2%	5.0%	5.4%

それぞれのセグメントについて，次の一カ月間での購買状況を集計した結果が表 6 である．表 6 の 1 列目の値は左から R, F, M のそれぞれのクラスを表しており，数字が大きいほど好ましいと思われるクラスである．したがって (333) のセルが最も好ましく，(111) が最も好ましくないと考えられる．次月平均購買購買金額 A はセグメント全員の平均購買金額であり，平均購買金額 B は継続購買客のみを分母にした値である．離脱率は (111) のセグメントは (333) のセグメントの 4 倍であり，次月平均購買金額は半額以下である．Recency は小売業においてもっとも効果的な変数ともいわれるが，この結果を見ると，Frequency，Moneytary Value の両者も考慮しなければ，継続優良顧客の判定がうまくいっていないことがわかる．つまり，Recency のみが高くとも F, M が低いセグメント（たとえば (311) のセグメント）は継続購買率も低く，平均購買金額も決して高くない．むしろ，本データでは F, M の両者が高いセグメントがよい結果を示している．

表 5　RFM の閾値

グループ	R		F		M	
	最小値	最大値	最小値	最大値	最小値	最大値
1	9	30	1	1	50	3785
2	2	8	2	4	3788	12064
3	0	1	5	86	12965	365388

表 6　RFM 分析の評価

RFM 値	所属人数	次月離脱客	反復購買確率	次月平均購買金額 A	次月平均購買金額 B
111	1629	1286	21.1%	555	2,634
112	284	213	25.0%	1,078	4,311
113	10	10	0.0%	0	0
121	549	346	37.0%	753	2,036
122	724	438	39.5%	1,496	3,787
123	97	50	48.5%	2,666	5,502
131	21	7	66.7%	1,002	1,503
132	115	42	63.5%	1,894	2,984
133	100	34	66.0%	3,495	5,295
211	479	390	18.6%	452	2,430
212	89	65	27.0%	1,150	4,265
213	3	3	0.0%	0	0
221	423	259	38.8%	698	1,799
222	856	454	47.0%	1,607	3,422
223	232	93	59.9%	3,718	6,205
231	42	12	71.4%	929	1,300
232	502	137	72.7%	1,977	2,720
233	919	159	82.7%	5,383	6,509
311	159	137	13.8%	354	2,560
312	73	60	17.8%	776	4,356
313	5	5	0.0%	0	0
321	152	95	37.5%	809	2,157
322	444	254	42.8%	1,665	3,891
323	165	90	45.5%	2,390	5,257
331	27	7	74.1%	829	1,119
332	397	96	75.8%	1,779	2,346
333	1952	220	88.7%	6,630	7,472

4.3　主成分分析

顧客別部門別購買点数を集計し，そのデータを元に主成分分析を行った．分析に際しては相関係数をもとに行った．第 1 主成分を横軸，第 2 主成分をたて軸として，主成分負荷量ならびに主成分得点を同時プロットしたものを図 5 に示す．

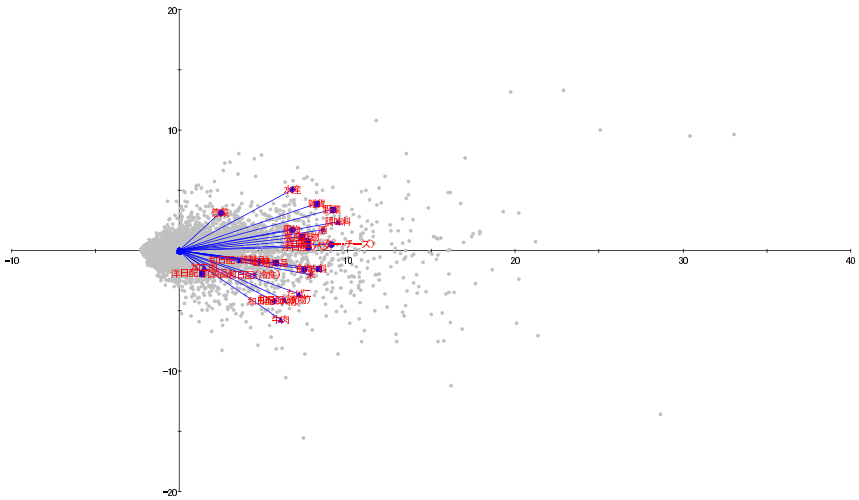

図 5　主成分分析の結果

主成分分析では，主成分負荷量を元に軸を解釈し，各サンプルの差異をもとにグループ化をしていくが，この分析例の場合，第一軸はすべての項目の主成分負荷量が正の値となっており，いわば総合的な購買量を示しているに過ぎない．逆にこの結果の限界としては，あくまで集計レベルで購買が多かった顧客を抽出しているにとどまってしまう点である．

また，第 1 主成分から第 4 主成分までの寄与率および累積寄与率を 7 にまとめる．この結果を見ると，第 1 主成分が示す総合購買量で 30%を超える寄与率を示しているものの，第 2 次主成分以降は高々5%程度であり，本データが 24 部門からなることを考えれば，一つの変数が持つ情報量の平均は $1/24 = 4.1\%$ であり，第 2 主成分以降ではデータ縮約にはあまり寄与していないことが見て取れる．したがって，第 1 主成分における主成分得点については，差異に十分な意味があるとはいえるものの，第 2 主成分以降については情報量は少ない．

表 7　主成分の情報量

	寄与率	累積寄与率
第 1 主成分	34.0%	34.0%
第 2 主成分	5.4%	39.4%
第 3 主成分	4.6%	44.0%
第 4 主成分	4.2%	48.2%

4.4　クラスター分析

前節の主成分分析に用いたデータを用いて，k-means クラスタ分析を行う[8]．なお，対象間の距離としては，以下に示す余弦を利用した．この定義を利用することで，変量の大きさよりむしろ購買部門の比率に着目することができる．

$$d(i,j) = \frac{(\boldsymbol{x}_i \cdot \boldsymbol{x}_j)}{||\boldsymbol{x}_i||\ ||\boldsymbol{x}_j||} \tag{7}$$

セグメント数は 7 と設定して分析した．各セグメントにおける各部門の値の平均とを表 8 及び図 6 にまとめる．なお，図 6 の軸は各項の値が大きく異なるため，基数を 10 とした対数変換をおこなったものである．

表 8　k-means クラスタ分析の結果概要

クラスタ	所属人数	代表的部門
1	1317	和日配（漬物），洋日配（乳製品），牛肉
2	1622	米，たばこ
3	805	水産，雑貨，惣菜
4	2351	酒，洋日配（パン），嗜好品，菓子
5	2157	野菜，果物，加工肉，和日配（練製品）
6	1470	和日配（水物）
7	726	和日配（冷食）

この結果を見ると，購買の差異についてはクラスタによって傾向の差を見ることができるが，クラスタ 4 が全体的にレーダーチャートの外側に位置することもあり，

[8]なお，階層型クラスター分析を試みたところ，1 時間以上っかっても終了しなかった．これは，サンプル数が 1 万を超えているため，全組み合わせの距離に時間がかかっているためと推測できる．

やはり全体の購買量によってセグメントが作られている印象はぬぐいきれない[9]．

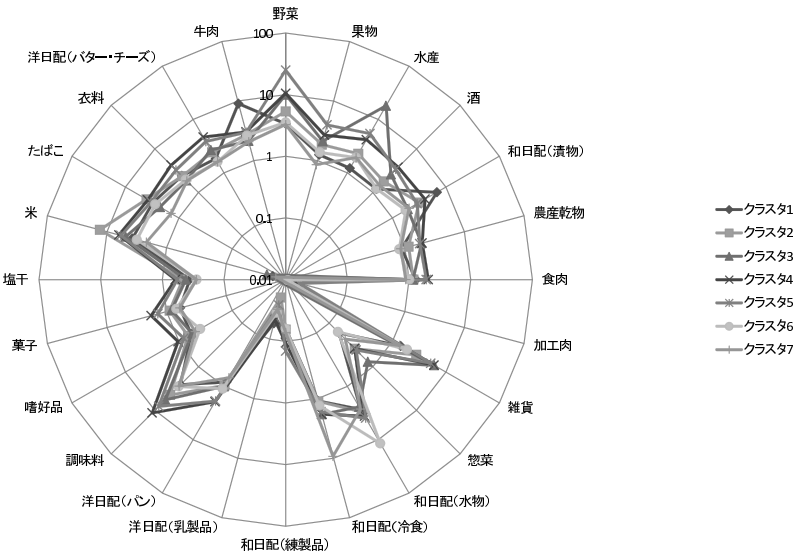

図 6　k-means 法によるセグメントの各部門購買平均（対数）

4.5　自己組織化マップ

対象データについて，顧客別部門別購買点数を集計し，自己組織化マップで分析した．図 7 から図 10 は結果の一部である．これらの図に示すように，4×10 の六角格子を用いた[10]．

それぞれのセルに所属する人数を表 9 に示す．

所属する顧客分布はセル (5,1) を中心としており，このセルは平均的な購買をしているセグメントであるが，実施には購買金額のあまり多くないセグメントである．各図の右上のセグメントが購買の多い顧客のセグメントであるが，部門を比較すると，購買の広さが部門によって異なることがわかる．しかし，たとえば，(4,10) の

[9]なお，対象間の距離を市街地距離もしくは，ユークリッド距離として分析すると，ある一つのセルに全体の 60%のサンプルが集まる結果となった．これはセグメント数を変えて分析しても同様の傾向であった．

[10]色の濃いところが購買が多く，薄いところは購買が少ない．

ように他の食品部門の購買の高いセグメントを見つけることができ，隣接するセルを追うことで購買動向の変化の推移を把握することができる．

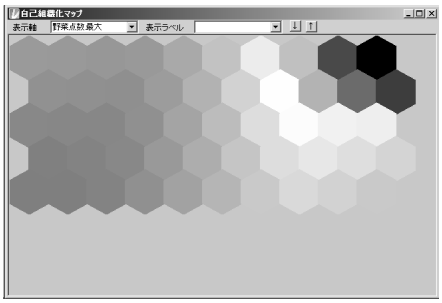

図 7　自己組織化マップ（野菜）

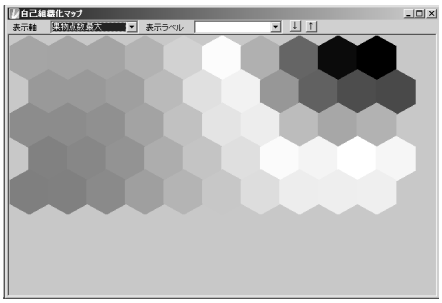

図 8　自己組織化マップ（果物）

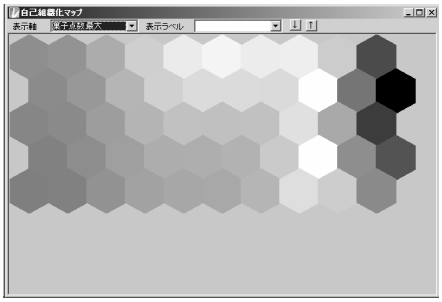

図 9　自己組織化マップ（菓子）

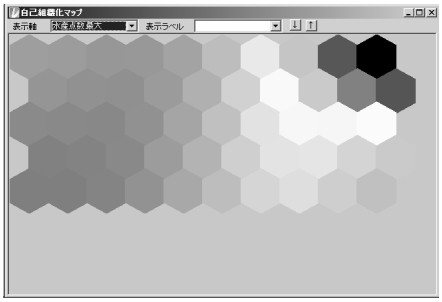

図 10　自己組織化マップ（水産）

4.6　潜在クラス分析

4.6.1　集計データによる潜在クラス分析

サンプルごとに部門別の購買の有無（購買がある場合は 1，購買がない場合は 0）とした 10488 × 24 の行列を用いて，潜在クラス分析を行う．モデル選択は情報量規準の AIC や BIC がしばしば利用される[11] 潜在構造を持つ場合は BIC がしばしば

[11] AIC（Akaike Information Criterion：赤池情報量規準）と BIC (Bayesian Information Criterion：ベイズ情報量規準) はモデル選択のための情報量基準のであり，それぞれ，$\mathrm{AIC} = -2 \times \ln\{L\} + 2 \times k$ および $\mathrm{BIC} = -2 \times \ln\{L\} + k \times \ln\{n\}$ で求められる．ただし，L は尤度，k は独立変数の数，n はサンプルの数であり，これらの値が小さいものが望ましい．

表 9　SOM のセルに所属する人数

格子セル	1	2	3	4	5	6	7	8	9	10	合計
1	232	147	68	90	59	53	41	33	25	15	767
2	325	116	86	54	60	34	24	23	15	14	757
3	598	210	193	119	116	74	64	38	27	17	1461
4	871	424	167	143	92	76	48	34	15	17	1892
5	4028	550	283	188	193	144	67	46	36	30	5569
合計	6056	1449	800	596	521	384	246	176	119	95	10448

用いられるため，本分析でも BIC をもとにクラス数を決定した．表 10 に示すように，7 クラスモデルが最良とされた．

表 10　潜在クラス分析のモデル選択規準の比較

クラス数	AIC	BIC
5	238765	239664.
6	238379	239460
7	238100	239362
8	237920	239364
9	237802	239427

各クラスの購買動向の相違について，クラスごと，部門ごとの生起確率で比較する．この結果を表 11 および図 11 に示す．

また，このときの各クラスの構成比率を表 12 に示す．クラス 4 は広い部門にわたって購買しており，クラス 1 は特に生鮮品，クラス 4 は酒類などの嗜好品の購買が顕著なクラスである．

4.6.2　非集計データによる潜在クラス分析

前節では，顧客ごとに購買状況を集計したデータからの潜在クラス分析を行った．ここで紹介する二項ソフトクラスタリングは，トランザクションごとに二つの変数間の関係を分析しようというものである (（株）数理システム, 2011)．二組の変数 $(X, Y) = [\boldsymbol{x}, \boldsymbol{y}]$ について x_i と y_j（たとえば，顧客 i と購買部門 j）の共通起確率 $p(x_i, y_j)$ は (8) 式で表される．

表 11　各クラスの部門別平均値

クラス	1	2	3	4	5	6	7
野菜	90.2%	99.0%	62.3%	100.0%	6.7%	45.1%	94.5%
果物	53.8%	81.2%	28.0%	89.8%	11.2%	29.1%	56.5%
水産	73.3%	95.3%	46.3%	95.6%	5.3%	22.7%	42.7%
酒	37.8%	51.6%	20.3%	74.4%	26.5%	38.8%	44.9%
和日配（漬物）	37.8%	68.8%	19.3%	91.5%	2.1%	20.7%	52.3%
農産乾物	45.7%	81.7%	18.0%	97.9%	17.7%	53.7%	86.6%
食肉	62.7%	85.2%	31.3%	98.5%	3.0%	25.0%	79.9%
加工肉	43.0%	70.4%	22.4%	96.5%	2.9%	22.0%	71.9%
雑貨	33.2%	64.3%	12.5%	84.9%	14.7%	33.4%	55.1%
惣菜	62.0%	80.9%	34.0%	89.5%	40.0%	58.6%	69.5%
和日配（水物）	87.3%	99.9%	51.1%	100.0%	8.2%	52.3%	92.8%
和日配（冷食）	15.0%	27.3%	6.7%	63.8%	2.3%	14.7%	44.9%
和日配（練製品）	29.9%	62.0%	10.8%	88.9%	0.7%	11.1%	50.2%
洋日配（乳製品）	76.2%	92.4%	40.6%	99.6%	35.1%	75.7%	93.5%
洋日配（パン）	60.6%	83.3%	29.3%	95.2%	25.6%	63.4%	86.2%
調味料	60.0%	93.0%	27.0%	98.8%	15.3%	47.1%	83.2%
嗜好品	46.0%	75.8%	17.9%	96.3%	36.4%	75.9%	89.6%
菓子	46.1%	76.0%	17.6%	93.7%	27.8%	68.4%	86.5%
塩干	79.7%	96.6%	43.9%	98.4%	3.1%	20.3%	55.8%
米	3.9%	11.5%	1.5%	20.8%	1.9%	1.6%	6.9%
たばこ	1.1%	3.5%	0.2%	4.3%	0.6%	1.5%	2.0%
衣料	0.4%	0.8%	0.1%	2.7%	0.3%	0.7%	0.7%
洋日配（バター・チーズ）	16.0%	28.0%	4.9%	64.3%	2.2%	15.4%	41.3%
牛肉	8.9%	19.5%	5.4%	40.2%	1.0%	4.4%	14.8%

表 12　各クラスの構成比率

クラス	1	2	3	4	5	6	7
サイズ	22.5%	18.4%	15.1%	13.3%	13.2%	10.2%	7.4%

$$p(x_i, y_i) = \sum_k p(z_k)p(x_i|z_k)p(y_j|z_k) \tag{8}$$

ただし，$\sum_i p(x_i|z_k) = 1$, $\sum_j p(y_j|z_k) = 1$ である．二項ソフトクラスタリングにおける共起確率の概念を図 12 に示す．このように，潜在クラス k を橋渡しとし，そのの生起確率 $p(z)$ とクラス z に関する x_i と y_j の条件付き生起確率を求め，この和がこのペアの生起確率となる．(8) 式により，各サンプルに関する尤度が求められるので，同様に AIC, BIC などを求めることができる．

分析データについて，各顧客 ID を X の変数に，また各部門の購買金額を Y として，二項ソフトクラスタリングを行った．クラス数を変化させて AIC と BIC を求めたところ，表 13 のようになった．最良のクラス数は 4 であるが，従来の潜在クラ

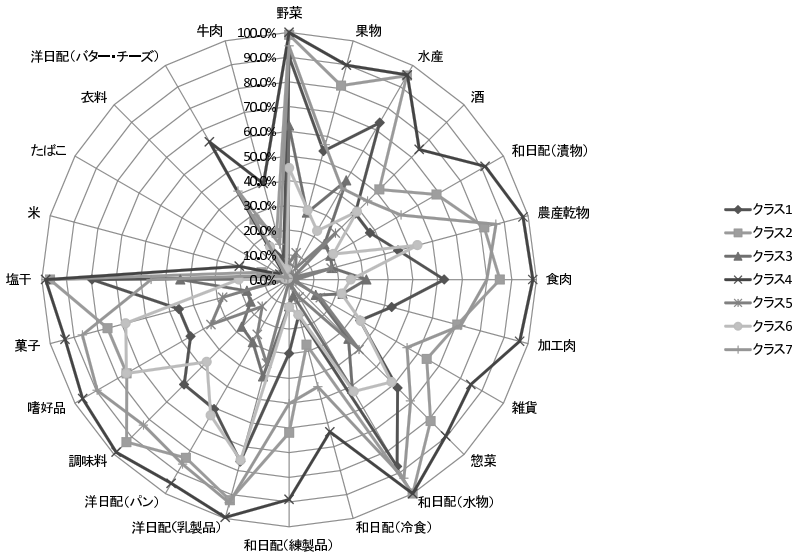

図 11　各部門の生起確率のレーダーチャート

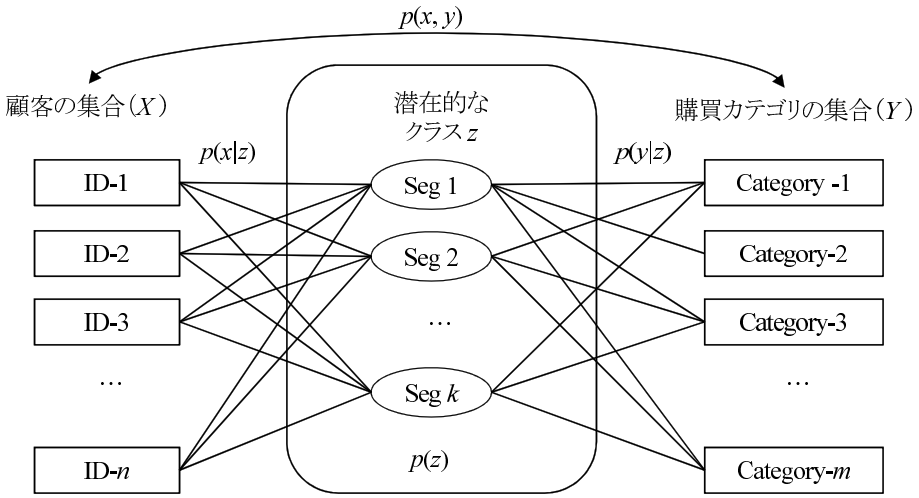

図 12　二項ソフトクラスタリングの概念

ス分析と比較してクラス数の少ないモデルが支持された理由としては，本分析手法では，各サンプルに対して生起確率を独立に[12] 付与するため，求めるべきパラメータ数が多くなることが理由の一つである．

表 13　二項ソフトクラスタリングのモデル選択規準の比較

クラス数	AIC	BIC
2	12294617	12622059
3	12257829	12545290
4	12241325	12446599
5	12242104	12485802
6	12248575	12704522

4 クラスモデルについて各クラスに対する各部門の所属確率を表 14 に，またそのレーダーチャートを図 13 に示す．

各クラスの構成比率は表 15 の通りである．

このような枠組みで共起確率を求めると，x_i を条件として y_j が生起する確率を x_i から z_k を通じて y_j に向かうパスを考え，(9) 式のように求めることができる．(9) 式で得られる確率の高い順序で，x_i に関係の深い，つまり推薦できる対象であるといえる．

$$p(y_j|x_i) = \sum_k p(y_j|z_k)p(z_k|x_i) \tag{9}$$

従来の潜在クラス分析では，各潜在クラスへの所属確率は求められるが，この分析のように，X と Y の要素ごとの共起確率を考えるわけではなく，共変量ベクトルの所属確率を一意に求めるため，(9) 式のように各変量に対する確率を求めることはできない．このことも本分析の特徴といえる．実際前節の潜在クラスモデルと比較しても購買の違いがはっきりとしている．

各顧客について上位 3 位までの推奨部門を抽出し，クラスごと[13] の推薦割合を表 16 にまとめる．

表 16 から，クラスごとに推薦の傾向がかなり異なることがわかる．また，推薦順位をさらに広げると，ここにでてきた以外の部門も現れてくる．

[12] 正規性のみが条件となる．
[13] 所属クラスは確率最大のクラスに割り当てた．

表 14　各部門の所属確率

部門名	クラス 1	クラス 2	クラス 3	クラス 4
野菜	0.0%	12.6%	0.0%	87.4%
果物	0.0%	45.7%	0.0%	54.3%
水産	0.0%	0.0%	1.9%	98.1%
酒	0.0%	10.8%	0.0%	89.1%
和日配（漬物）	0.0%	98.5%	1.5%	0.0%
農産乾物	0.1%	29.2%	0.1%	70.6%
食肉	1.8%	71.0%	2.9%	24.3%
加工肉	0.0%	82.0%	17.9%	0.1%
雑貨	0.0%	0.0%	0.0%	100.0%
惣菜	0.6%	0.0%	3.6%	95.8%
和日配（水物）	100.0%	0.0%	0.0%	0.0%
和日配（冷食）	0.0%	0.0%	100.0%	0.0%
和日配（練製品）	0.0%	0.0%	100.0%	0.0%
洋日配（乳製品）	100.0%	0.0%	0.0%	0.0%
洋日配（パン）	8.6%	24.2%	0.4%	66.8%
調味料	0.1%	23.1%	0.5%	76.2%
嗜好品	0.2%	60.5%	4.0%	35.4%
菓子	0.0%	28.3%	0.0%	71.7%
塩干	0.0%	57.5%	5.1%	37.4%
米	1.5%	72.8%	0.0%	25.8%
たばこ	15.0%	67.2%	0.0%	17.8%
衣料	2.7%	56.2%	1.8%	39.3%
洋日配（バター・チーズ）	0.0%	36.7%	2.1%	61.2%
牛肉	3.2%	95.3%	1.4%	0.0%

表 15　各クラスの構成比率

クラス	1	2	3	4
構成比率	10.2%	29.7%	13.6%	46.5%

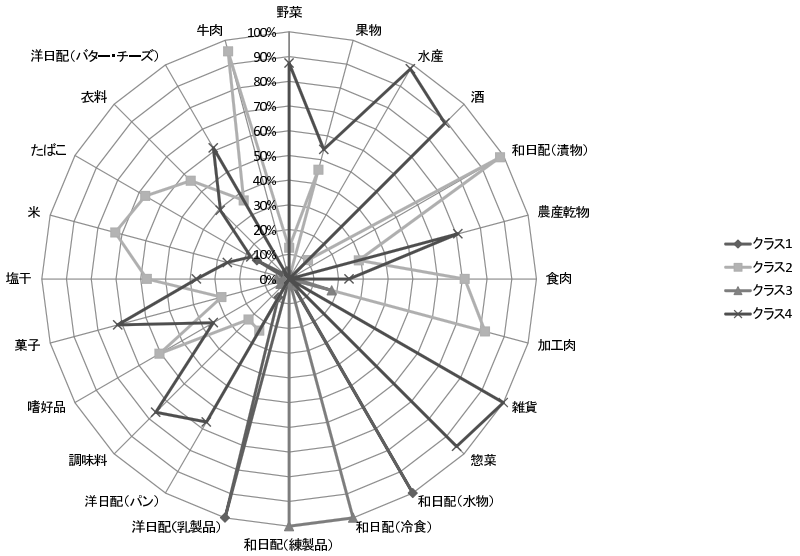

図 13　各部門の所属確率のレーダーチャート

表 16　クラスごとの部門別推薦割合

部門　クラス	クラス 1	クラス 2	クラス 3	クラス 4
野菜	15%	13%	13%	91%
水産	16%	0%	37%	88%
和日配（漬物）	0%	44%	0%	0%
雑貨	5%	0%	1%	54%
和日配（水物）	100%	30%	34%	27%
和日配（冷食）	27%	20%	100%	27%
和日配（練製品）	3%	0%	77%	0%
洋日配（乳製品）	46%	0%	0%	0%
米	28%	97%	13%	11%
たばこ	26%	0%	0%	0%
牛肉	34%	95%	24%	2%

表 17　SVM と線形判別分析の比較 (1)

	線形判別分析			
		平均未満	平均以上	合計
学習用データ	平均未満	5617	27	5644
	平均以上	1869	845	2714
	合計	7486	872	8358

	Support Vector Machine		
	平均未満	平均以上	合計
平均未満	5334	310	5644
平均以上	1006	1708	2714
合計	6340	2018	8358

		平均未満	平均以上	合計
検証用データ	平均未満	1368	15	1383
	平均以上	469	238	707
	合計	1837	253	2090

	平均未満	平均以上	合計
平均未満	1302	81	1383
平均以上	257	450	707
合計	1559	531	2090

表 18　SVM と線形判別分析の比較 (2)

	線形判別分析			
		平均未満	平均以上	合計
学習用データ	平均未満	99.5%	0.5%	100.0%
	平均以上	68.9%	31.1%	100.0%
	合計	89.6%	10.4%	100.0%

	Support Vector Machine		
	平均未満	平均以上	合計
平均未満	94.5%	5.5%	100.0%
平均以上	37.1%	62.9%	100.0%
合計	75.9%	24.1%	100.0%

		平均未満	平均以上	合計
検証用データ	平均未満	98.9%	1.1%	100.0%
	平均以上	66.3%	33.7%	100.0%
	合計	87.9%	12.1%	100.0%

	平均未満	平均以上	合計
平均未満	94.1%	5.9%	100.0%
平均以上	36.4%	63.6%	100.0%
合計	74.6%	25.4%	100.0%

4.7　Support Vector Machine

教師付データによるセグメンテーションとして，Support Vector Machine による例を挙げる．ここでは，対象となる 1 か月間の購買データについて，顧客別部門別に購買金額を集計した 10488 × 24 のデータを作成した．目的変数は，次月の購買が平均金額以上であったかどうかのバイナリ変数である．なお，閾値は約 13,500 円である．顧客の購買金額は裾が広く分布しているため，平均以上の購買があった顧客は 3421 人であり，次月の購買実績がなかった顧客も含み，平均未満の購買顧客は 7027 人であった．評価のために，このデータを学習用データ 80%，検証用データ 20%と分割し，前者でモデル構築を行い，その結果を用いて後者を評価した．なお，カーネル関数には線形カーネルを用いた[14] ．

SVM の評価のために，線形判別分析の結果と比較する．表 17 は判別結果を，表

[14]他に多項式カーネル，Radial Basis Function，Sigmoid カーネルなども利用されている

18 は行方向の比率を求めたものである．

この結果を見ると，線形判別分析での学習用データでの判別率は 77.4%であり，検証用データの判別率は 76.8%であったのに対して，SVV は，学習用データでは 84.3%でり，検証用データでは，83..8%であった．表 18 をもう少し詳しく見ると，もともと判別対象（0-1 データ）に偏りがあったが，サンプル数の少ない平均購買以上の顧客の判別率が，線形判別分析ではサンプル数に応じてウェイト付しているのにもかかわらず 30%程度と低いのに対して，SVM では 60%以上適切に判別できている．目的にもよると思われるが，少数のサンプルを判別したいという場合には，線形モデルはうまく抽出できない場合があることが示唆される．

5　おわりに

本稿では，マーケティングにおけるセグメンテーションの重要性について述べ，実際にどのようにセグメンテーションを行うかについて，その分析モデルを概観した．さらに，紹介したいくつかのセグメンテーション手法について，実データを用いた分析結果を示し，その特徴について考察した．

最初に述べたように，セグメンテーションは，マーケティング・プロセスの第一段階であり，実際には作成されたセグメントからマーケティング活動実施へ向けて計画を進め，その効果についての検証が必要である．ただし，これを実現するためには，実務の方々とのタイムリーなコラボレーションが必要である．したがって，得られた結果からどのようにマーケティング活動を進めていくかのステップまでを含めてセグメンテーションの効果について述べるべきであったが，本稿ではデータの都合上そこまで踏み込めなかった．

EC サイトなどでは，よりワン・トゥ・ワン・マーケティングへの志向が顕著であるが，これは個別の顧客を認識するためのシステムが整っているためである．EC サイトでは主に，顧客へのレコメンデーションや情報処理の補助・代行などが行われており，アイテム間の協調フィルタリング技術なども用いられている．本稿で紹介した，二項ソフトクラスタリングは，これに代わる方法の一つとして期待されるが，これについては今後の展開を待ちたい．

本稿は，平成 19 年度専修大学研究助成共同研究「小売業における POS データの多面的分析とその活用」の研究成果の一部である．

参考文献

[1] Americal Marketing Association ウェブサイト
http://www.marketingpower.com/（2011/02/28 アクセス）．

[2] Bezdek, J.C. (1981) *Pattern Recognition with Fuzzy Objective Function Algorithms*, Springer.

[3] Bradley, P.S. and U.M. Fayyad (1998) "Refining Initial Points for K-means Clustering," *Proceedings of the 15th International Conference on Machine Learning*, 91–99.

[4] Calinski, R.B. and J. Harabasz (1974) "A Dendrite Method for Clustering Analysis," *Communications in Statistics*, 3, 1–27.

[5] Han, J. and M. Kamber (2001) *Data Mining Concept and Techniques*, Academic Press.

[6] Hughes, A. (2006) *Strategic Database Marketing*, 3rd Edition, McGraw-Hill.

[7] Kass, G.V. (1980) "An Exploratory Technique for Investigating Large Quantities of Categorical Data," *Applied Statistics*, 29, 110-127.

[8] Kohonen, T. (2000) *Self-Organizing Maps*, Springer.

[9] Kotler, P. and K. Keller (2008) *Marketing Management: International Edition*, Prentice Hall.

[10] Kotler, P. and G. Armstrong: (2009) *Principles of Marketing*, 13th Edition, Prentice Hall.

[11] （株）数理システム (2011)『Visual Mining Studio マニュアル バージョン 7.0』，数理システム.

[12] Murtagh, F. (1983) "A Survey of Recent Advances in Hierarchical Clustering Algorithms," *The Computer Journal*, 26, 354–359.

[13] 中村博（編著） (2008)『マーケット・セグメンテーション』，白桃書房.

[14] Okada 岡太彬訓，木島正明，守口剛（編著） (2001)『マーケティングの数理モデル』，朝倉書店.

[15] Quinlan, J.R. (1993) *C4.5: Programs for Machine Learning*, Morgan Kaufmann.

[16] Quinlan, J.R. (1996) "Improved Use of Continuous Attributes in C4.5," *Journal of Artificial Intelligence Research*, 4, 77–90.

[17] Scott, A.J. and M.J. Symons (1971) "Clustering Methon based on Likelihood Ratio Criteria," *Biometrics*, 27, 387–397.

[18] 田中豊，垂水共之 (1995)『Windows 版　統計解析ハンドブック　多変量解析』共立出版.

[19] Wu, X. and V. Kumar (eds.) (2009) *The Top Ten Algorithms in Data Mining*, Chapman & Hall.

[20] Wedel, M. and W. Kamakura (2000) *Market Segmentation, Conceptual and Methodological Fundations*, 2nd Edition, Kluwer.