

SOM によるファン層の解析に基づく CD 購買予測モデルの作成

五反田 剛, 石井 良和, 原 健一郎, 関 庸一

1. はじめに

音楽 CD 販売店舗やレーベルにとっては、音楽 CD の個々の顧客がどのように CD を購買するか把握できれば、プロモーションを行う上で有益な情報と成り得る。この購買を説明する最大の要因は顧客の音楽嗜好と考えられる。

顧客の音楽嗜好は、購買した CD、特にそのアーティストから理解できると期待される。しかし、個々の顧客の購買履歴が少なくデータが疎である場合、購買されたアーティスト情報から嗜好を具体的に特定することは、マイナーなアーティストを特徴づけることが難しいため困難である。そこで、本研究では、購買アーティストのロングテール部分についてはジャンルにまとめて購買アーティストの嗜好の特徴を把握し、これが似た顧客を SOM (Self-Organizing Maps, 自己組織化マップ) [1][3]によって類型とし、顧客を層別する方法を提案する。これにより、各顧客の購買履歴が少ないデータの場合でも顧客の音楽嗜好を把握することが可能となると考えられる。

また、新譜購買行動に結びつく要因は顧客ごとに異なると考えられる。しかし、音楽嗜好を共有する顧客には類似した行動傾向が想定できる。そこで、購買に結びつく要因が共通するセグメントを、前述の類型化の結果に基づき構成する方法を与える。具体的には、一つの購買予測モデルで説明できる類型群を統合してセグメントとするものである。これにより、共通して効いている変量からそのセグメントの特徴が理解できると考えられる。

ごたんだ つよし, いしい よしかず, はら けんいちろう

群馬大学 工学研究科情報工学専攻

せき よういち

群馬大学 工学部情報工学科

〒376-8515 桐生市天神町 1-5-1

受付 06.7.21 採択 06.11.15

以上のアプローチにより、音楽 CD 販売店における ID 付き POS データから各顧客の音楽嗜好と顧客の購買行動を理解する方法論を示す。

2. データ概要

対象としたデータは、「平成 17 年度データ解析コンペティション」で提供された ID 付き POS データであり、ある音楽 CD 販売店系列の十店舗において 2003 年 9 月から 2005 年 8 月までの 731 日間に収集された 324,727 人の約 125 万回、約 28 億円分の購買データである (約 90 Mbyte)。対象期間中、1 人当たり 3.86 枚、8,527 円の購買が行われている。

2.1 商品構成

対象音楽 CD は、25 項目からなる大ジャンル (図 1) に分類され、例えば J-POP は「歌謡曲」、「オムニバス」などに細分類され、合計 120 項目からなる中ジャンルに分類されていた。

十店舗の大ジャンルごとの売上構成としては、J-POP (58.0%)、ロック & ポップス (17.0%) が売上枚数の主要な部分 (合計 75.0%) を成している (図 1)。さらに各大ジャンルの中ジャンルごとの売上構成を見ると偏りが見られ、中ジャンルの「その他」の売上枚数が、属する大ジャンルの売上枚数の七割以上を

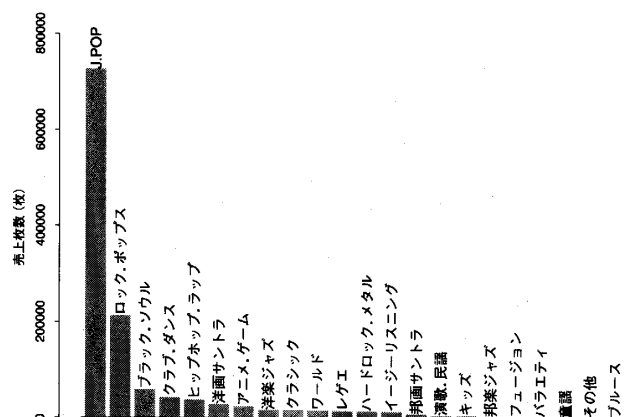


図 1 全店舗の大ジャンルごとの売上構成

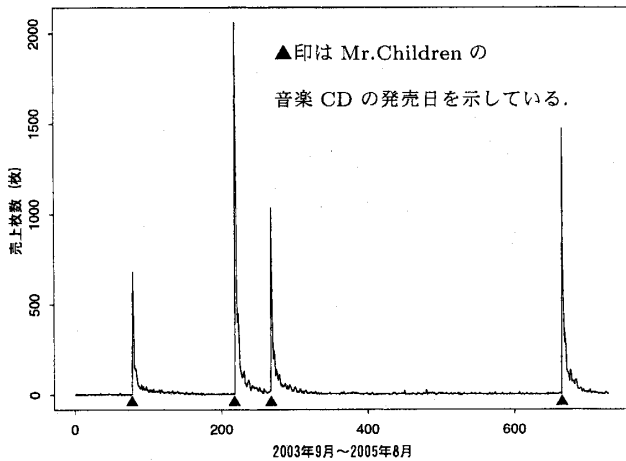


図2 Mr. Childrenの売上枚数の変化

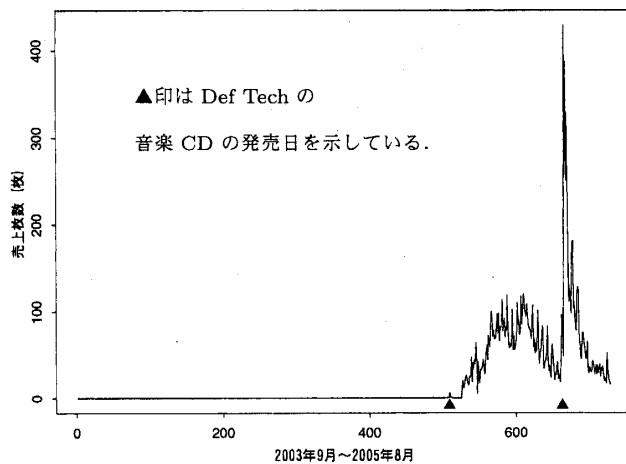


図3 Def Techの売上枚数の変化

占める大ジャンルが13あり、中ジャンルはバランスのとれた商品分類とはなっていない。

また、各CDは存在した11,852グループのいずれのアーティストによる音楽CDなのか識別できるようになっていた。そのうち売上枚数が多い2,371グループのアーティスト(20.0%)によって、売上枚数の主要な部分(合計83.1%)が成されており、ロングテールの状況が見られた。

2.2 CD発売後の売れ方の特徴

アーティストごとのCDの日々の売上枚数をみると、J-POPなどの多くの場合、発売日に最大の売上枚数を示し、一ヶ月ほどの時間が経つにつれて売上枚数が減少していくことがわかる(図2)。ただし、図3に示すように新規デビューアーティストなどの場合、1枚目の音楽CDは発売後から緩やかに売上枚数が増加した後に減少するものもあった。

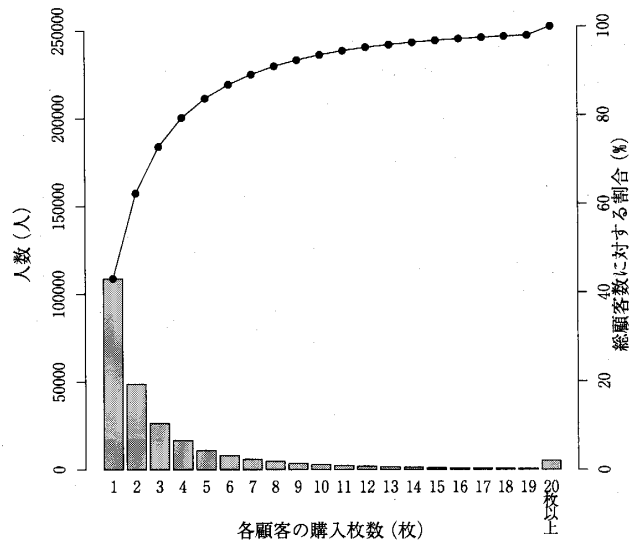


図4 一年半での顧客のCD購買枚数のパレート図

3. SOMによるファン層の理解

本節では、顧客をその音楽嗜好からタイプ分けする。タイプ分けの方法として、顧客がどのようなアーティストのCDを購入したかという、顧客とアーティストの共起行列を購買履歴から作成し、これを入力とするSOMによって顧客の層別を行う。

3.1 共起行列の作成

二年間の観測期間のうち最後の半年を予測対象として残し、はじめの一年半(2003年9月1日~2005年2月28日)の期間をタイプ分けに用いる。この、一年半の顧客の総人数は約25万人、一年半のCDの総売上金額は約21億円となった。

図4に示すようにCDを1枚のみ購買している顧客が約43%、2枚購買している顧客が約19%と一人当たりの購買頻度が低かった。後述のように分析において顧客内アーティストシェアを利用するので、一年半の間にCDを5枚以上購買している顧客(52,411人・全体の約20%)を対象顧客を限定した。

一方、分析対象となるアーティストについても限定を行う必要がある。大ジャンル単位で購買CDを集計し、顧客×大ジャンルの共起行列を考えた場合、図1から分かるようにJ-POPとロック&ポップスのボリュームが大きく、この二つの大ジャンルの内部の構造が見えなくなり内容の乏しいマップとなってしまう。他方、アーティスト単位で購買CDを集計して顧客×アーティストの共起行列を考えた場合、マイナーなアーティストまでを区別して扱くと巨大な行列となり分析が困難になる。そこで、売上上位アーティスト(売

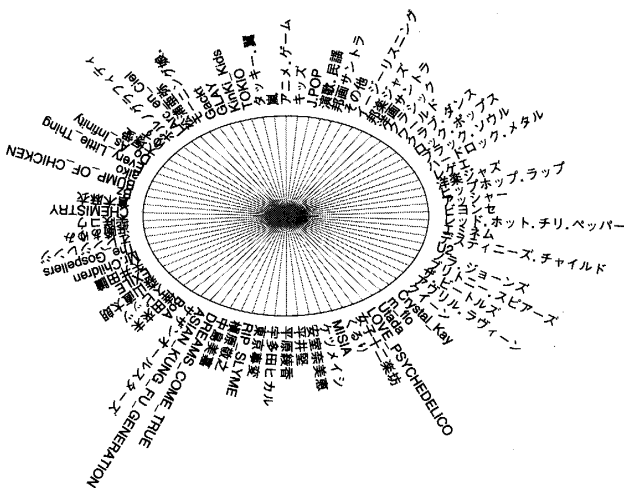


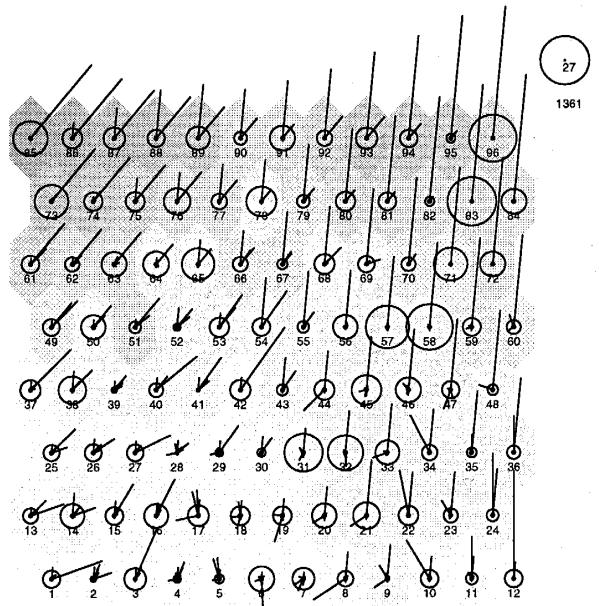
図5 購買CD比率を表すレーダーチャート

上枚数が1,500枚以上のアーティスト)は区別するが残りのロングテールについては大ジャンルごとに集計し、この分類を基準とした。このような集計方法を行うことで大ジャンルと売上上位アーティストとの関連性を見る事が可能になる。

以上のような分析対象の限定により、52,411×78(顧客×(アーティスト+大ジャンル))の共起行列を作成した。さらに、この行列を各顧客毎に総購買枚数で割ることで個人内のシェアへと基準化を行った。この変換により購買ボリュームの情報が捨てられ、音楽嗜好の情報のみが残ることとなる。この行列を入力としてSOMによって顧客のタイプ分けをする。なお、個人内シェアは図5のレーダーチャートで表現する。線の方向がCDを表し、線の長さが対応する項目の購買比率を示す。

3.2 SOMによる顧客のマップ

SOMは、予備的な知識なしで、多次元の属性をもつサンプルを、2次元上に規則的に配置した格子点(ノード)に、サンプル間の類似性に応じて層別する手法である。SOMによる層別では、まず初期値として各ノードにランダムに決められた参照ベクトルを与える。この参照ベクトルを次の手続きにより反復更新する。まず、ランダムに選んだ顧客について、その78次元個人内シェアを各ノードの参照ベクトルと比較し、最もユークリッド距離の近い参照ベクトルをもつノードに層別する。また、同時にそのノードの近傍ノードで、参照ベクトルをその顧客のデータに類似するように更新する。以上の手続きを反復し、顧客の所属ノードが変わらなくなるまで実施することによって、類似した音楽嗜好をもつ顧客は相互に近隣のノードに



右上の凡例は、所属人数が最小のノード(27人)と所属人数が最大のノード(1361人)を意味している。また、各ノードの下の数字はノード番号を意味する。

図6 SOMによる顧客のタイプ分け

所属するように層別されることになる。

図6は、対象顧客52,411人をSOMによって2次元六角格子上の12×8=96個のノードに層別した結果である。各ノードにおいて、それぞれのレーダーチャートが参照ベクトルを表し、円の大きさが、そのノードに所属する顧客人数を表す。また各ノードに1から96のノード番号を与えた。今回の対象データからは多くの個性的なノードが得られると予想され、[3]よりは大きめの12×8のノード配置を採用している。収束に当たっては、初期値として近傍半径を10ノード分、学習率係数を0.04とし、半径と学習率係数を等差的に減少させながら100万回の参照ベクトルの反復更新を行っている。このように近傍半径初期値を広くとり、多く反復を行うことで、マップ全体が大域的に整理されることが期待される。

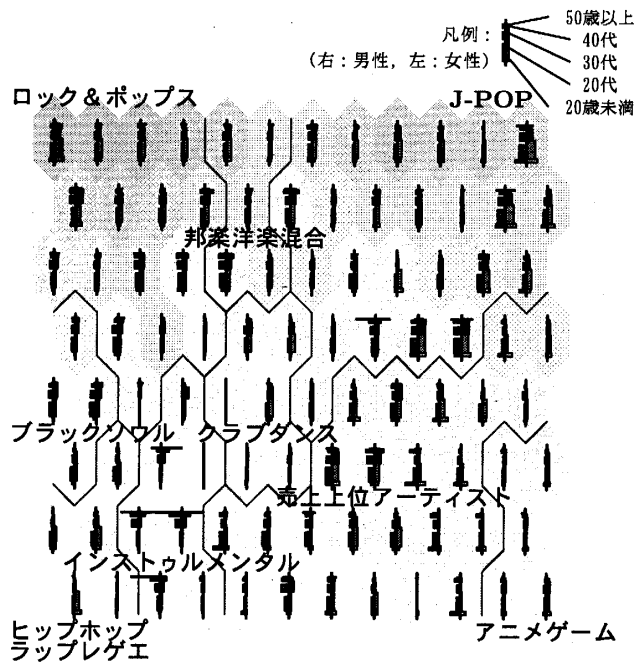
SOMによって作成されたマップ上では、類似した音楽嗜好をもつ顧客が近隣のノードに所属するので、各ノードの近隣関係に意味がある。これを利用し、各ノードに所属する顧客について属性値を集約し、マップのノードの位置に描くことによってマップ上の各領域の傾向を見ることが出来る。図7は、SOMの12×8格子の配置位置に、各ノードに所属する顧客の人口ピラミッドを示している。

図7には96のノードを図6の参照ベクトル(購買シェア)からタイプ分けした結果も示す。タイプ分けを行うに当たっては、原則として各ノードにおいてシェ

アが最も高いジャンル/アーティストをそのノードのタイプと定義した。ただし単独アーティストについて、シェアが5%以上存在するノードはその売上上位アーティストタイプと定義した。また35,36番ノードについてはJ-POPが最も高いシェアを持つが、次に高いシェアのある項目がアニメゲームであるので、隣接するアニメゲームが最大であるノード24,11,12とあわせアニメゲームタイプと定義した。さらにJ-POPまたはロック&ポップスが最大シェアであるノードについては、両者の購買比率の差が10%以下のノードについて邦楽洋楽混合タイプと定義した。なお洋画サントラ、クラシック、ワールド、洋楽ジャズ、ハードロックメタルが最も高いシェアを持つノードはまとめてインストゥルメンタルタイプとした。以上のタイプ

分けは、境界の設定などに任意性が残っているが、マップ上の各領域の特徴づけとしては有効と考える。

以上の各タイプの特徴量の全平均からの差を表1にまとめた。利用した変数のうち*の変数の詳細については表2に示す。表1から次のようなことが分かる。邦楽洋楽混合タイプは、一人当りの購買が多い。邦楽・洋楽問わず様々なCDを購入しているからだと考えられる。インストゥルメンタルタイプは平均年齢が高く、男性率が低い。また一人当りの購買、火水曜利用率が低い。クラシックやジャズといった音楽は性別を問わず、比較的年齢が高い顧客に好まれ、発売日を気にせず購買していると考えられる。ヒップホップラップレゲエタイプは、平均年齢が低く男性率が高い。また朝昼利用率が低い。夜に購買を行う若い男性の顧客の集合であると考えられる。売上上位アーティストタイプは、男性率が低く朝昼利用頻度が高い。また、早期購買・火水曜利用・一途度が高い。性別を問わず、特定のアーティストに関心を持つ顧客の集合であると考えられる。



各ピラミッドの各棒グラフの面積がその性年代層の人数を意味する。各長方形の高さを全顧客中のその性年代層の存在比に比例させ、すべての層で均等なら全体として、ほぼ長方形の形状になるようにしている。女性側の各棒グラフ間に隙間が存在するのは、男性側の棒グラフの始点と統一したためである。

図7 各ノードに属す顧客の人口ピラミッド

表2 作成した変数

変数名	内容
アルバム率	CDのパッケージ種別にはアルバム、マキシシングル、シングルの三種類があり、顧客の購買したCDのうち、アルバムを購買した割合を示す指標。
朝・昼・夜・深夜利用率	午前10時から6時を昼として以降、6時間刻みで夜・深夜・朝とした。
早期購買回数	発売日から3日以内に購買した回数。データに発売日がないため、二年間で200枚以上売れたCDについてのみ考え、そのCDが初めて売れた日を発売日として扱う。
火水曜利用率	CDは原則として火曜日に発売される。発売日を計る作成した指標。
一途度	特定アーティストを購買するかを計る指標。購買CD種類数、購買アーティスト数と定義した。
重複なし総購買数	同一CDを複数枚購買した場合は1枚とみなす。総購買種類数。
購買遅延日数平均	発売日から購買した日までの日数遅延の平均。初めてCDが売れた日から7日間で7枚以上売れているCDについては、初めて売れた日を発売日とする。また、そうでない場合は日数遅延を180日に固定した。
一来店平均購買数	一回の来店当りに購買したCD枚数平均。

表1 タイプ毎の特徴量 (全平均からの差分。*の変数の詳細については表2参照)

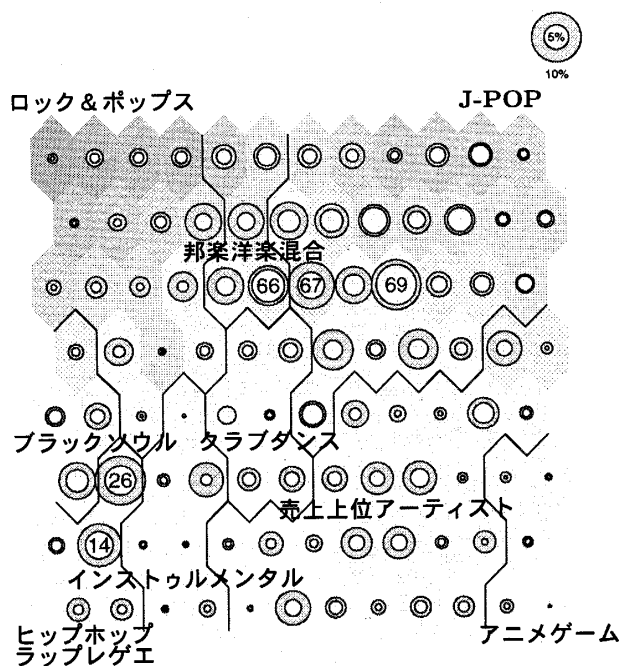
タイプ名	所属人数	平均年齢	男性率	一人当り購買数	アルバム率*	朝昼利用率*	早期購買回数*	火水曜利用率*	一途度*
J-POP	15571	-0.4	2.5	1.0	-7.3	-0.2	-0.2	1.2	0.0
洋楽邦楽混合	2774	0.7	2.4	2.1	9.4	-1.6	-0.2	-3.7	-0.2
アニメゲーム	1822	-0.4	8.3	-0.3	-12.4	6.4	-0.2	0.1	0.1
インストゥルメンタル	2913	6.8	-3.7	-1.6	16.4	3.7	-0.3	-6.4	0.0
クラブダンス	2103	-0.3	10.5	-0.8	17.6	-5.9	-0.2	-5.5	0.2
ヒップホップラップレゲエ	2237	-3.0	6.4	-0.7	18.2	-6.2	-0.2	-6.2	-0.2
ブラックソウル	2276	1.0	-8.7	-0.6	18.2	-5.7	-0.2	-7.3	-0.2
ロック&ポップス	9085	0.6	5.6	0.2	17.7	-2.3	-0.2	-5.4	-0.2
売上上位アーティスト	13630	-1.1	-8.6	-1.0	-16.2	3.1	0.5	7.4	0.2
全平均		29.1	57.3	11.0	76.0	24.7	0.3	34.0	1.4

3.3 新規アーティストの浸透

本節では、マップ作成時には存在しなかった新規アーティストが、作成したマップ上でどのように顧客を獲得していったかを検討してみる。対象アーティストとしては、2005年1月21日にデビューし、集計期間二年間で最も売上のあったCDをリリースしたDef Techとする。

図8には、CD「Def Tech」のノード毎のノード所属人数に対する購買者割合を前期購買者（累計売上枚数が総売上枚数の50%に達する2005年5月6日までに購買した顧客；内側の円）と、期間中の総購買者（外側の円）に分けて示してある。また、ノード番号が与えてあるノードは前期購買者割合が高い上位5つのノードである。

図8の前期購買者割合の上位5ノードを見ると、14, 26, 69番ノードはJ-POPとヒップホップラップレゲエのシェアが共に10%以上である唯一のノード群となっている。Def Techの音楽ジャンルはJ-POPとヒップホップラップやレゲエの中間に属するので、Def Techは自らの音楽ジャンルと音楽嗜好が一致している顧客層にまず支持されたといえる。また、66,67番ノードでは一人当たり購買枚数がそれぞれ13.8枚と13.1枚と高く（全平均11.0枚）、一途度が共に



内側の円は各ノードに所属する顧客の総人数に対する前期購買者割合を表す。外側の円は各ノードに所属する顧客の全購買者割合を表す。右上の凡例は内側の円の面積が5%を、外側の円の面積が10%の割合を意味する。また、前期購買者割合の上位5つのノードにノード番号を与えてある。

図8 CD「Def Tech」のノード毎の購買者割合

1.2と低い（全平均1.4）。このことから、様々なアーティストの音楽CDを購入しているような顧客層にも、Def Techは早くから支持されていたといえる。一方その周辺の音楽嗜好が類似したノード、特に66,67,69番ノード周辺から売上上位アーティストのノードの一部では、後期購買者の割合が高いノードが存在する。以上より、Def Techはその音楽ジャンルと音楽嗜好が一致している顧客層や様々な音楽CDを購入しているような顧客層にまず支持され、その後それらと音楽嗜好の類似した顧客層に浸透していったといえる。このように、新規アーティストの情報なしで作成したマップ上で、そのアーティストがどのようなファン層を震源とし、以後どのようなファン層に浸透していったかを把握することができた。

4. 新譜購買モデルの構築

各顧客の属性やSOMによる結果を用いて、対象としたアーティストの購買状況を説明するモデルを構築する。具体的には、集計期間二年間のうち、最初の一年半の購買状況を用い、残り半年の購買状況を説明するモデルを、モデル統合分析法という手法を用いて構築する。

モデル統合分析法とは、SOMにより層別されたサンプル群毎に対し、MDL基準に基づき一般化線形モデルを推定し、隣接した層が同一のモデルで説明可能ならば二つの層を統合することで、モデルが適用できる層（顧客セグメント）を与える方法である[2]。

4.1 対象アーティストの選択

対象とするアーティストとしては、最後の半年のうちの最後のヶ月を除く五ヶ月に少なくとも1枚の予測対象CDを発売している必要がある。最後のヶ月に発売したCDを対象外とするのは、図2から分かるように、売上が落ち着くのに発売後およそ一ヶ月必要なためである。

この条件を満たし、マップ上で確立したファン層を持つアーティストとしてMr. Childrenを選択した。

表3 対象アーティストと発売したCD

Mr.Children			
タイトル	パッケージ種別	売上枚数	発売日
掌	シングル	2953	2003/11/18
シフクノオト	アルバム	7864	2004/04/06
Sign	シングル	4150	2004/05/25
* 四次元 FourDimensions	シングル	5269	2005/06/28
Def Tech			
タイトル	パッケージ種別	売上枚数	発売日
* Def Tech	アルバム	10762	2005/01/21
* Lokahi Lani	アルバム	3020	2005/06/28

Mr. Childrenは、二年間と最初の一年半それぞれでの総売上枚数が共に1位である。また、新規アーティストとして、3.3節で取り扱ったDef Techを選択した。対象とするアーティストとその発売CD・パッケージ種別・二年間での総売上枚数・発売日を表3に示す。

4.2 モデル統合結果

対象アーティストについて、表3の*印のCDを購入した枚数(ただし、同一タイトルを複数購入しても1枚とした)を各アーティスト毎に目的変数とした。つまりMr. Childrenならば{0,1}であり、Def Techならば{0,1,2}となる。この目的変数が二項分布に従うと仮定して予測モデルを構築した。説明変数としては、最初の一年半データにおける顧客一人ひとりについての、年齢、性別、アルバム購買数、マキシシングル購買数、火水曜利用率の他、表2の変数を作成して用いた。なお、重複なし総購買数、アルバム購買数、購買遅延日数平均、一来店平均購買数は対数変換処理を施した。

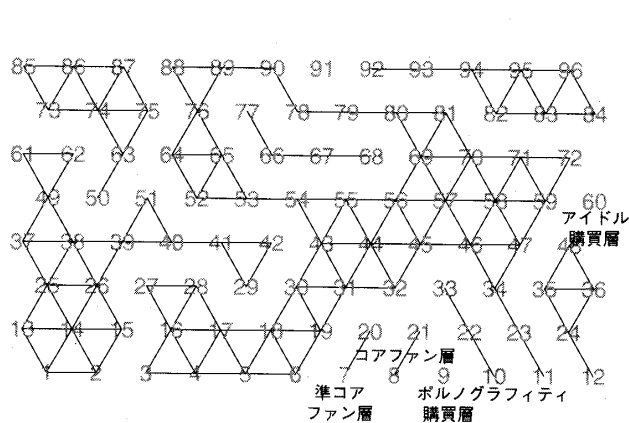


図9 Mr. Childrenのノード統合図

図9は、Mr. Childrenを対象に各ノードの統合状況を視覚的に示した図であり、線でつながれたノード群が顧客セグメントを形成している。これにより12の顧客セグメントが得られた。表4に、各セグメントの特徴を一人当たりの購買枚数の上位4セグメントまで示している。例えば、アイドル購買層ではマキシシングル購買枚数が少ない顧客ほど、Mr. Childrenの新譜を購入すると言える。これは、アイドルの音楽CDはマキシシングルが多いため、マキシシングルの購買枚数が少ない顧客は、アイドルだけに興味があるのではなくMr. Childrenも購入する余裕があると考えられる。

また図10は、Def Techのノード統合図である。表5と併せて見ると、採用された変数にはMr. Childrenほど顕著な特徴は見られず、全体としての数が少ない結果となった。これは、購買する顧客がSOMマップ上に散っており、全体の平均値を定数で説明するとい

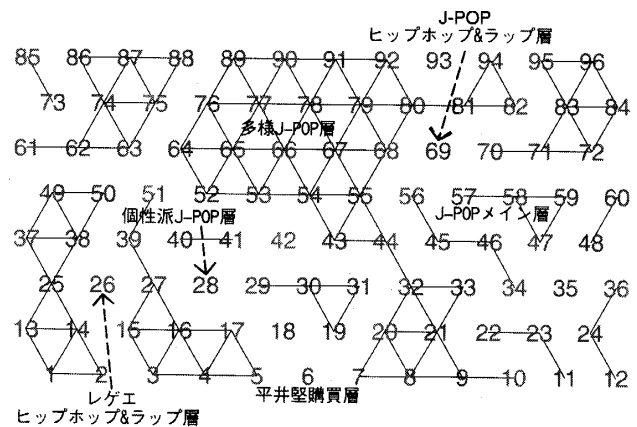


図10 Def Techのノード統合図

表4 Mr. Children・モデル毎の係数表

ノード番号	ノード群 特徴	説明変数						対象平均購買数(枚)	人数(人)
		重複なし総購買数	マキシシングル購買数	一途度	火水曜利用率	購買遅延日数平均	一来店平均購買数		
8,21	コアファン層	1.162	-	0.538	0.806	-0.111	-3.043	0.229	1190
9	ホルノグラフィティ購買層	-	-	2.735	-	-	-	0.177	119
7,20	準コアファン層	1.828	-	-	-	-0.241	-2.054	0.119	1301
60	アイドル購買層	-	-0.211	-	-	-	-	0.081	395

表5 Def Tech・モデル毎の係数表

ノード番号	ノード群 特徴	説明変数					対象平均購買数(枚)	人数(人)
		重複なし総購入数	アルバム購入数	早期購入回数	夜利用率	一来店平均購入数		
69	J-POP ヒップホップ&ラップ層	-	1.514	-	-	-	0.132	471
26	レゲエヒップホップ&ラップ層	-	-	-	-	-	0.127	479
28	個性派 J-POP 層	-	-	-	0.526	-	0.116	43
6	平井堅購買層	-	-	0.267	-	-	0.104	705
47,57,58~	J-POP メイン層	0.599	-	-	-	-	0.087	3497
7,8,9~	多様 J-POP 層	0.758	-	-	-	-0.684	0.080	16477

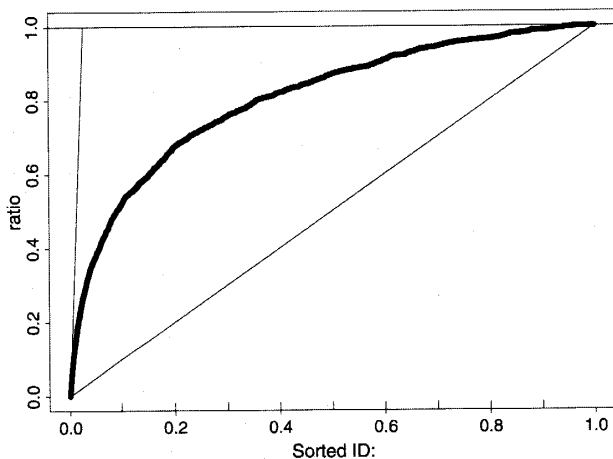


図 11 Mr. Children リフト図

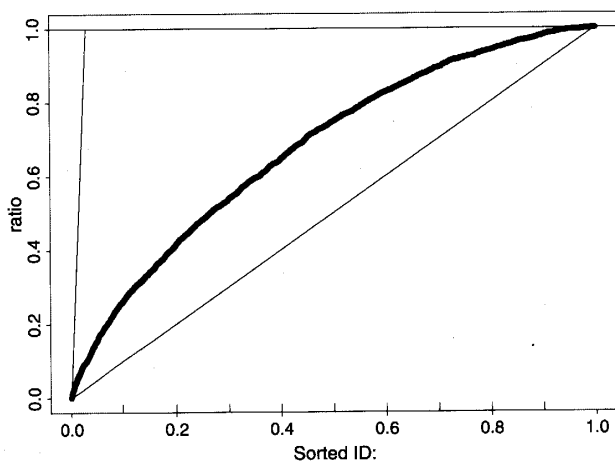


図 12 Def Tech リフト図

う状況が多くあったためであると考えられる。

図 11 と図 12 は二組の対象アーティストについて構築されたモデルのリフト図である。購買履歴のある Mr. Children については上位一割の顧客で約五割の売上を説明できており、モデルが良く適合していると考えられる。一方、購買履歴のない Def Tech については適合度が低い結果となった。

5. おわりに

本研究では、購買履歴が疎なデータにおいて、多様

な顧客の音楽嗜好を層別するために、アーティストのロングテール部分を大ジャンルに集約した上で顧客×アーティストの共起行列を作成し、SOM を適用することを提案した。その結果、顧客の音楽嗜好における大ジャンルとアーティストの親近性が視覚的に把握でき、得られた類型からタイプ分けが可能になった。さらに、このように音楽嗜好の近さで顧客を分類したマップを利用することで新規アーティストの顧客集団への浸透の状況を把握することが可能となることを示した。

また、このようなマップ上でモデル統合分析法を適用することで共通のモデルで説明できる顧客セグメントを得る方法を示した。このモデルの構造を理解することで購買要因が明らかになり、各セグメント毎に、異なるアプローチを行うことで効率的なプロモーションが可能になると考えられる。なお今回、十分な規模のデータがなかったため、取りおいたテストデータによる予測モデルの評価はできていない。今後、採用された一般化線形モデル一つひとつに対する評価を含め、モデル統合分析法の検証法について検討を進めたい。

以上のアプローチより、音楽 CD 販売店舗にとっては自店舗が抱える顧客の傾向を把握することができ、また、レーベルにとっては所属アーティストの位置付けが把握でき、新たなファン層の獲得や新規アーティストの方向性を決める指標につながると期待される。

参考文献

- [1] T. コホネン：自己組織化マップ，シュプリンガー東京（1996）。（T. Kohonen: *Self-Organizing Maps, 3rd ed.*, Springer, Berlin (2001).）
- [2] 関庸一，長井歩，石原淳一郎，渡邊亮：“自己組織化マップによる行動履歴の類型化—クレジットカード利用履歴を利用したキャッシング移行予測—”，日本経営工学会誌，Vol. 57, No. 5, pp. 404-412 (2006).
- [3] 渡辺亮，北村裕人，星野直人，関庸一：“買回りタイプによる顧客購買行動の理解”，オペレーションズ・リサーチ，Vol. 50, No. 9, pp. 42-51 (2005).