

グラフ分割手法を応用した CD 販売店の顧客購買分析

中原 孝信, 森田 裕之

1. はじめに

近年、インターネットやモバイル端末の普及により、CDのインターネットによる通信販売や、音楽の有料(無料)配信など、消費者が音楽を入手する手段が多様化しており、CD販売店を取り巻く環境は急激に変化している。日本レコード協会の報告[6]では、インターネット通販の利用率は、2004年度に比べ2005年度は拡大傾向にある。また、有料音楽配信の認知率は60%以上になり、2005年の秋より有料音楽配信の利用者が急増している。このことは、今後CD販売店にとって、脅威となることが予想される。

一方、消費者がCDを購入する主な理由には、「自分のものにしたい」、「歌詞カード・ジャケットが欲しい」、「コレクションにしたい」などの購買に対して積極的な理由と、「好きな曲だけを購入したい」、「レンタルは面倒である」などの消極的な理由が存在している[2]。前者はCDを物理体として入手することに価値を見出すような顧客であり、後者は、音楽配信などのより便利な代替手段に移行しやすい顧客と考えることができる。以上のように、従来のCD販売店での流通チャネルを維持していく観点に立てば、音楽CDの購買を望む顧客と長期的に良好な取引関係を築いていくことが重要になると考えられる。

本稿では、優良顧客と考える「コレクター顧客」と、必要なものしか購買しない「一時的購買顧客」を購買データから定義し、コレクター顧客の維持拡大を目的に分析を行う。分析では、グラフ分割手法を応用して、購買者が併買するCDの状況から、関係の深いアー

ティスト同士を分ける新たな分類を提案する。そして、この分類を利用して、両者の購買行動の違いを明らかにする。

2. 分析対象データと基礎分析

分析に用いるデータは、某CD販売店10店舗における2年間(2003年9月~2005年8月)のID付きPOSデータである¹。データのレコード数は約120万件で、顧客人数は約32万4千人である。CDの情報としては、11,853人のアーティストと50,483の作品タイトルが利用でき、各作品タイトルは、22種類の大ジャンル、43種類の中ジャンルに分類されている。そして、アルバム、シングル、マキシシングルなどの区分属性が利用できる。また、顧客属性は、年齢、性別が利用可能である。

データの基礎集計から2年間の総売上は、約27億6千万円であり、年代別に見ると、総売上の50.1%が20代、そして28.0%が30代とこれらの年代で大部分を占めている。ジャンル別に見ると、大ジャンルのJ-POPが総売上の56.7%を占めており、次いでロック&ポップスが17.5%、ブラック/ソウルが4.5%であり、大ジャンルにも売上に大きな偏りがある。さらに大ジャンルJ-POPの内訳をみると、その中分類の売上構成は、同名の中分類であるJ-POP(中)が47.1%を占めており、大ジャンル、中ジャンル共にJ-POPという分類に極端な偏りが確認できる。

表1は、顧客との取引状況を表している。新規顧客は2年目のみ、また離脱顧客は1年目のみ購買履歴のある顧客集合であり、継続顧客は、2年間継続して購買が確認された顧客集合を表している。

継続顧客に比べ、新規顧客、離脱顧客の人数は多いが、継続顧客は、全体の約20%の人数で売上の54%

なかはら たかのぶ
大阪府立大学 大学院経済学研究科
もりた ひろゆき
大阪府立大学 経済学部
〒599-8531 堺市中区学園町1-1
受付06.7.19 採択06.11.15

¹「平成17年度データ解析コンペティション」で提供して頂いたデータを用いている。

表1 顧客との取引状況

	人数	金額
新規顧客	147,272人 (45.4%)	7.06億円 (25.5%)
離脱顧客	107,057人 (33.0%)	5.68億円 (20.5%)
継続顧客	70,398人 (21.7%)	14.96億円 (54.0%)

を占めている。一般的な小売業と同様に、当該分野においても、継続取引顧客を維持・拡大することは、長期的な取引の観点から重要であることが確認される。

基礎分析の結果から、本データについては、購買の年代と購買ジャンルに大きな偏りがあり、継続的に安定した購買を行う顧客の重要性が確認された。スーパーマーケットのような小売業の購買データであれば、購買額の変動や同時関連購買の分析などにより有益な情報をマイニングすることも可能であるが、本データはCDの購買であるため、平均購買間隔が非常に長く、また、一度に複数商品を同時購入することは必ずしも一般的な傾向とはいえない。そのため、百貨店の購買データのように²年単位の比較的長期間での併買行動を分析する必要がある。もう1つの特徴は、商品単価に多少変動はあるものの、ほぼアルバムとシングルの2種類しか存在せず、高額商品や廉価商品の区別がないということである。したがって、どのような分類、またはどのようなアーティストのCDを併買しているかということが分析の鍵となる。当初、所与のジャンルの分類やアーティストをもとに顧客の併買行動を分析したが、ジャンルで見た場合は、J-POPに購買が偏りすぎているため、またアーティストレベルでは、アーティスト数が非常に多く、購買が多様化しすぎているため、有益な知識をマイニングすることが困難であった。

その原因は、大分類J-POP内の中分類区分にあると考えた。所与の分類は、購買の分析とは無関係であり、アーティストの申告などによって決定されているため、音楽性などを考慮し明確に分けられているわけではない。また1人のアーティストが複数のジャンルに係わる場合もある。そこで有益な併買行動を分析するためには、消費者の購買状況から新たにアーティストを再分類することが有益であると考えられる。

アーティストを1つの点、点を結ぶ枝の強さを併買関係によって表現すれば、全体を1つのグラフとして

表現できる。そして、このグラフをいくつかのサブグラフに分割することで、消費者の購買状況を考慮したアーティストの分類が可能となる。ただしその際、すべての点間に枝が存在しないこと、また関係が強いものだけを集めて分類すると分類の偏りが大きくなるという懸念がある。本稿では、グラフ分割問題 (graph partitioning problem) を利用して、併買関係の強さから、偏りの少ない分類区分を発見する。そして、これを利用して顧客の併買行動を分析する。

3. グラフ分割によるアーティストの分類

グラフ分割問題は、分割後の各グラフの頂点の大きさが均衡するという制約を満たし、異なる集合を横断する枝のコスト和が最小になるように、グラフの頂点を分割する問題であり、NP-困難であることが知られている[1]。その主たる応用は、VSLIの設計や、ネットワーク設計などであり、ビジネス分野のID付きPOSデータなどへの適用事例は知る限り報告されていない。本稿では、アーティスト間の併買関係をグラフで表現して k 個の部分集合へと分割することで、併買関係を考慮した新たなアーティストの分類を提案する。

本稿で扱うグラフ分割問題は、以下のように定義される。無向グラフ $G=(V, E)$ (V は頂点集合、 E は枝集合)、各点の重みと各枝のコストをそれぞれ $w(v)$, $c(e)$ とする。正整数 k が与えられたとき、グラフの頂点集合を以下のような均衡に関する制約を満たす k 個の部分集合 $V=V_1+V_2+\dots+V_k$ (非交和) に分割する。

$$w(V_i) \leq \rho \cdot \frac{w(V)}{k} \quad (1)$$

上式の ρ は後述のグラフ分割に利用したツール METIS[4] 内で設定されているパラメータであり、 $\rho=1.03$ である。ここで、分割によって生じる異なる部分集合間を横断する枝の総コストは、以下のように設定され、これを最小化することが目的となる。

$$\sum_{(x,y) \in E, x \in V_i, y \in V_j, i \neq j} c((x,y)) \quad (2)$$

以上の問題における分析上適切な点の重みと枝のコストを提供データから検討する。まず枝のコスト $c((x,y))$ は、アーティスト同士の関連の強さを表す値が望ましい。商品間の関連の強さを表す代表的な値としては、サポート (Support)、信頼度 (Confidence) や Jaccard 係数[3] などの指標が考えられる。しかし信頼度や Jaccard 係数は、アーティ

² 百貨店の購買データを分析した論文[5]でも購買周期が長かったため、長期間での併買行動を分析している。

スト x と y の購入者が少数である場合、その値が非常に高くなる傾向がある。特にアーティスト x と y を一人だけが購買し、それが同時購買である場合には、これらの値は共に 100% になる。これらの値を用いた場合、本分析データでは、約 150 の枝が 100% の値となってしまう、グラフ分割に強すぎる影響を与えることになる。一方、サポートは分子の数にその値が依存するため、あるアーティスト間だけの関係に着目した場合に、高い確率で発生する併買関係を過小評価してしまう傾向がある。このような関係が単独で存在する場合は、グラフ分割の過程において、その枝は切断されてしまう可能性が高い。しかし、どちらかのアーティストを基点として、複数の接続関係が存在していれば、それら個々の接続関係は弱くても切断される枝の総コストが増加するため、分断される可能性は小さくなる。

以上の考察の結果、本データに対しては枝の値にサポートを用いることが適当であると判断し、以下の式で定義する。

$$c((x, y)) = \frac{\text{アーティスト } x \text{ と } y \text{ を購買した顧客数}}{\text{分析対象全顧客数}} \quad (3)$$

次に各点の重みは、分割後、そのグループ内の合計値がほぼ均等になることが期待されるため、何らかのアーティストの相対的な重要性を示す値が望ましい。そこで、点の重みは各アーティストが購買された数を用いることにする。既存の分類では大ジャンル、中ジャンル共に、購買数について J-POP に大きく偏っている。既存の分類を利用して異なる 2 つの顧客集合間での併買関係の違いを観察する観点からは、このような偏りが存在すると、密なグループにおける関係は集約されすぎ、また疎なグループでの関係は希薄になりすぎる。そのため関係は確認できるが、その違いを示すことは難しくなる傾向にある。逆にその違いを表現するには、できるだけ偏りのないグループを作ることが 1 つの方法であると考えられる。そこで本研究では、前述のグラフ分割問題を設定し、分割後の各グループの購買数を均衡させることによって、後述の分析に望ましいと考えられる新たな分類を作成する。

3.1 CD 購買データへの適用

後述の分析との整合性を保つため、2003 年 9 月～2004 年 8 月の 1 年間に、大ジャンル J-POP に属するアーティストを 2 種類以上購買したことがある顧客 54,745 人を分析対象顧客として新たな分類を計算す

る。このとき購買された大分類 J-POP 内に属するアーティスト数 (点の数) は 3,096、それらの間の併買関係を表す枝の数は 224,789 であった。したがって、多くのアーティスト間には、併買が起こっていないことがわかる。

新たな分類を作成するために、分割数を与えなければならないが、最適な分割数を決定することは、簡単ではない。この場合、分割数が多くなると、切断される枝の総コストが増加するというデメリットが存在する。また分割数が多くなると、1 つの分割されるグループの頂点の重みの総和が小さくなり、1 つのグループ内に存在できる点の制約が厳しくなる。本データでは、頂点の最大の重みは Mr. Children で J-POP 全体の購買数に占める割合の 4.4% であり、次いでオムニバスが 3.8%、宇多田ヒカルが 3.2%、浜崎あゆみが 1.8% であった。そのため分割数が多くなると、例えば分割数が 10 の場合は、1 つのグループの重みの和の上限値は、全購買数に占める割合の 10% となり、導出される分割におけるアーティストの配置が固定的になる恐れがある。一方、2 つの顧客集合の違いを明確にする観点からは、分割数がある程度多い方が望ましいといえるが、そこにはトレードオフが存在する。そこで、分割数を 3 から 8 まで計算したところ、6 以上では分割が固定的になる傾向が確認された。残り 3 つの分割数の中では、分類されるアーティストを確認したところ分割数 4 のケースが最も音楽的に近いグループに分けられていると思われたため、ここでは 4 つに分割することにした。したがって、分割後の各グループに属する頂点の重みの総和は、全体の購買数の約 25% ずつになることが期待される。またこのことから、上位 4 アーティストの割合を見ても、購買枚数の多いアーティストの移動が制約されることはなく、上位のアーティストが同一グループに属して分割される可能性もあると考えられる。したがって、アーティストの移動が、枚数によって固定的になりすぎるといった状況にはないといえる。

分割の計算にはフリーで公開されているグラフ分割ツールである METIS³ を用いた。METIS は近似アルゴリズムであるが、計算の複雑性は $O(|E|)$ であり、計算実験からは比較的高速で、良好な結果が示されている [4]。

³ <http://glaros.dtc.umn.edu/gkhome/metis/metis/overview>

表2はMETISを用いて4つに分割した際の頂点数と重みの和を示している。各部分グラフ(名前は後述)の頂点数はかなり違いが見られるが、頂点の重みの和は、ほぼ均等になっていることが確認できる。

図1~4は、分割された各部分グラフの特徴を明らかにするために、各部分グラフ内で接続関係の強い上位20の枝とそれに接続する点を図示している。図中の各枝の値は(3)式により求めた枝のコストである。全体的に、各グループで著名なアーティストが出現し、それらの併買関係がよく表現されている。

図1は、浜崎あゆみやB'zなどいくつかのアーティストを中心に、それに関係の深いアーティストからなる複数のサブグループが形成されていることが確認できる。またこれらのアーティストは、音楽的な能力だけでなく、ダンスやルックスなど、その視覚的な表現力が強いことから、この部分グラフをビジュアルメインと呼ぶことにする。サブグループについては、モ

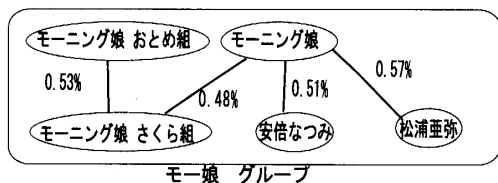
ーニング娘に関するアーティストの集合をモー娘グループ、浜崎あゆみを中心とした集合は、avexに所属しているアーティストが多いため、女性avexグループ、B'zに関するアーティストの集合はビーズグループ、そしてそれ以外の3つの集合は、ペアグループと呼ぶことにする。

図2は宇多田ヒカルとMr.Childrenを中心として、平井堅、平原綾香など歌唱力の高いアーティストが多数出現している。そこでこの部分グラフをボーカルメイングループと呼ぶことにする。

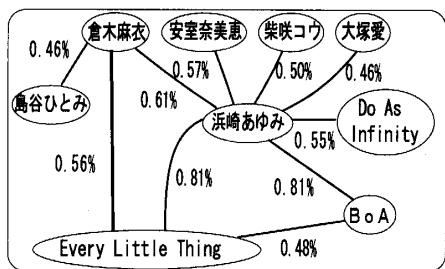
図3は、オムニバスを中心に複数のアーティストが放射状に接続している。このオムニバスはアーティスト名として与えられているものであるが、様々な音楽をまとめた作品集の総称であると考えられる⁴。したがって、オムニバスと何らかのアーティストを顧客が併買している関係が強く表現されている部分グラフであると考えられる。これをオムニバスメイングループと呼ぶことにする。

表2 4分割の値

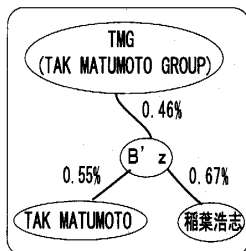
部分グラフ	点数	点の重みの和
ビジュアルメイン	235	62,406
ボーカルメイン	538	62,427
オムニバスメイン	1,176	62,525
インディーズメイン	1,147	62,263



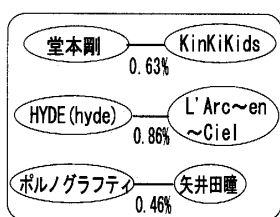
モー娘グループ



女性avexグループ



ビーズグループ



ペアグループ

図1 ビジュアルメイングループ

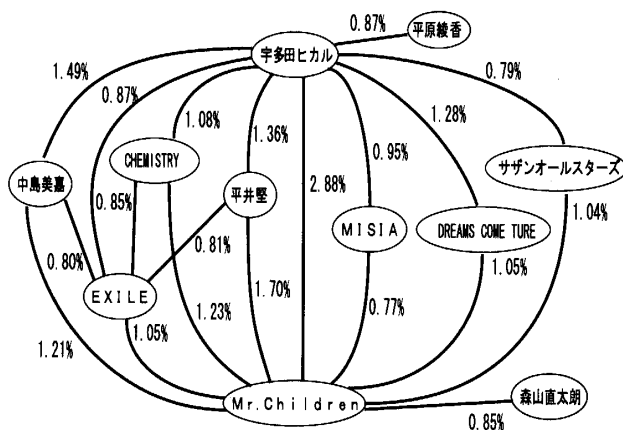


図2 ボーカルメイングループ

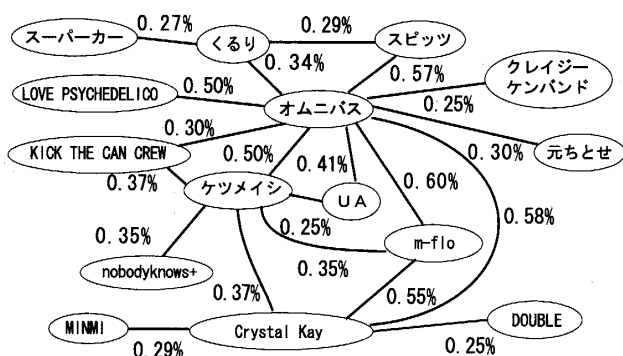


図3 オムニバスメイングループ

⁴ 中ジャンルにも同名のオムニバスという分類名が存在するが、ここではアーティスト名として与えられているオムニバスだけを利用している。

最後に図4は、MONGOL 800, ロードオブメジャー, HY, RIP SLYME など、2003年インディーズ売上ベスト20⁵にランクインするアーティストが多数出現していることから、この部分グラフをインディーズメイングループと呼ぶことにする。

以上のような新分類からいくつかの考察が得られた。1つは、著名なアーティスト間で必ずしも強い併買関係が存在していないということである。例えばもし、上位の売上を占めているビジュアルメインの浜崎あゆみとボーカルメインのMr. Childrenの接続関係が他の接続関係を切断するよりも強力であるならば、1つの分類になったはずである。この現象は他の分類に属している著名なアーティスト間の接続関係についても同様である。分割する前は、もっと有名な少数のアーティストが偏って分割されることを危惧していたが、結果として、購買者は単に有名なアーティストだけを併買していないことを意味している。逆に言えば、顧客は好みにしたがった購買を行うという結果を反映しているものと考えられる。

もう1つは、特定のファン層の行動も十分に反映されているということである。ビジュアルメインを見ると、例えばビーズグループのように、明らかに関係の深いと思われるアーティストのみを接続しているサブグループが出現している。また女性avexが、他の分類の女性アーティストよりも強く接続しているということは、顧客の何らかの好みの共通性を意味していると考えられる。今回検証することはできないが、店舗における商品配置などにも、これらの接続関係を利用した方法が十分応用できると考えられる。商品配置に

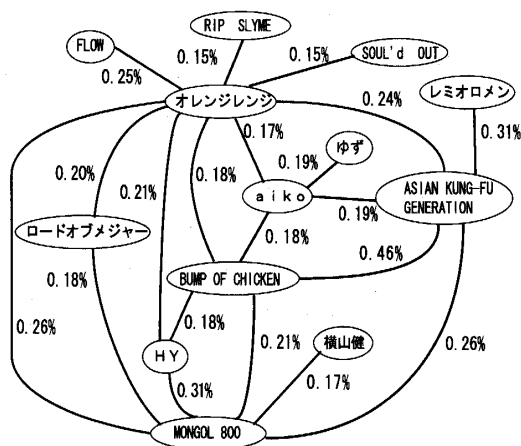


図4 インディーズメイングループ

については、ビジュアルメインだけでなく他の分類においても次数の高いアーティストに着目することが重要である。例えば、ボーカルメインの宇多田ヒカルやMr. Children, そしてインディーズメインのオレンジレンジやMONGOL 800は、それぞれ接続しているアーティストと隣接した商品陳列を検討すれば、効果的な結果を得ることが予想される。

以下では、優良顧客と考えるコレクター顧客をデータから定義した後、これらの新分類を利用しながら購買の特徴を明らかにする。

4. 分析対象顧客の特定と購買特徴の抽出

優良顧客と考えるコレクター的な性質を有する顧客を特定するため、2年間継続して購買のある顧客70,398人を対象に、各顧客の2年間のデータを1年目(2003年9月~2004年8月)と2年目(2004年9月~2005年8月)に分け、これら2年分の全顧客の標準化したRFM値を用いて、K-means法によって、全顧客を4つのセグメントに分類した。ここでRecencyの基準日は、各月の来店回数に極端な差はなかったので、2年間のデータを1年毎に分ける観点から、1年目、2年目共に8月末日を最終日として用いている。各グループに属する顧客のRFM値の分布を図5に示す⁶。図の各軸は標準化したRFMの値を表している。図よりRecencyが小さく、FrequencyとMonetaryの値が最も大きいセグメントは優良顧客集合と位置づけることができ、このセグメントをコレクター顧客と呼ぶことにする。Monetaryは悪化するが他の値が比較的良好なセグメントを潜在顧客、そ

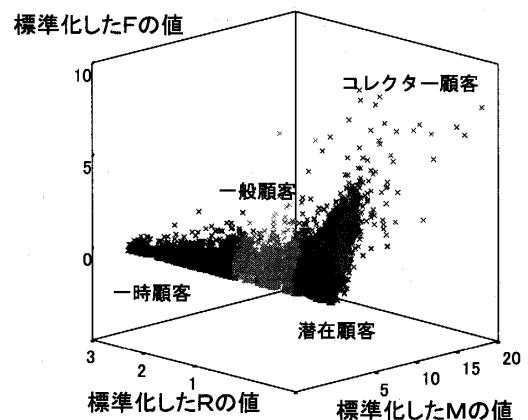


図5 RFMを基準にした顧客のセグメント

⁵ オリコン年間2003年版を参照。

⁶ この図には、各顧客についてそれぞれ2点がプロットされていることに注意されたい。

れに隣接するセグメントを一般顧客、そしてRFM値が最も悪いセグメントを一時顧客と呼ぶことにする。

表3は、1年目のデータを用いて、各セグメントに属する顧客の基礎集計を行った結果である。コレクター顧客は、月に約1回の来店があり、年間約4万円の購買を行う魅力的な顧客集合であることが確認できる。潜在顧客は、コレクター顧客に次いで良い顧客集合であり、最新来店を見ると、31.9と直近で約1ヶ月以内の来店が確認できることから、潜在顧客をいかにしてコレクター顧客へと導けるかが、当該CD販売店にとって1つの重要な課題であると考えられる。

図6は、各セグメントの1年目から2年目の移動を図示している。楕円の大きさは各セグメントの人数を相対的に反映させている。コレクター顧客と一時顧客に属する人数は、1年目に比べ2年目には若干増加しているが、潜在顧客と一般顧客は減少している。2年間継続してコレクターである顧客も存在するが、数が少ないため5%以上の移動としては確認できない。これはコレクター顧客から他のセグメントへ移動した場合も、また潜在顧客以外からコレクター顧客になる場合についても同様である。したがって、コレクター顧客になる可能性を最も秘めているのは、潜在顧客セグメントであり、約3,500人の移動が確認できる。その

一方で潜在顧客から一時顧客へと利用が減少する移動も7,200人程度確認され、このセグメントは移動が激しいことがわかる。また、セグメント間の推移確率⁷の観点から移動を分析すると、1年目コレクター顧客から2年目コレクター顧客への移動で45.9%が最も高く、次いで一般顧客から潜在顧客の移動が38.0%であった。逆に推移確率が低い移動は、1年目一時顧客から2年目コレクター顧客への移動で2.8%、そして1年目一般顧客から2年目コレクター顧客への移動が5.7%であった。これらの結果から、いったんコレクター顧客になれば2年目も継続の可能性が大きく、また一時顧客や一般顧客がコレクター顧客になることは稀であることが確認できる。以上の考察から、潜在顧客からコレクター顧客に効率よくプロモーションを行うことがCD販売店に望まれる1つのシナリオであると考えられ、以後の分析では潜在顧客に焦点を当てて分析を行うことにする。

以下では、1年目に潜在顧客である顧客に着目し、共通する1年目の購買データだけを利用して、将来コレクター顧客になる顧客と一時顧客になる顧客の特徴を分析する。1年目だけのデータを用いるため、購買している量や来店頻度などRFMの観点からは、これらの2つのグループの顧客はほぼ等しい。そこで以下では、前節で計算した新たなJ-POPの分類と、それ以外の既存の大ジャンルを全て分析に用いて、この両者を分ける特徴を発見する。

4.1 コレクター顧客と一時顧客の比較

両者を分けるような特定の分類（またはアーティスト）の購買は、基礎分析から確認できなかった。そこで併買分析のため、分析対象顧客を2種類以上のアーティストの購買経験があり、1年目は潜在顧客に属し、2年目にコレクターとなる顧客（3,044人）と2年目に一時顧客となる顧客（4,862人）に限定して以下の分析を行う。上述のように、分析は両方の顧客集合が同じ潜在顧客に属している前半1年間の購買データのみを用いて行う点に注意されたい。

J-POP内の4つの新分類における両者の購買者比率を計算したのが表4である。表より、ビジュアルメインは購買者比率に14%程度の差があることが分かる。また、インディーズメインも、ビジュアルメインほどではないが、10%弱の違いが確認できる。一方、

表3 セグメントの基礎集計

セグメント	購買合計 (百万円)	人数	平均購買額	平均来店回数	最新来店 (日数)
コレクター	253	5,757	43,927.6	12.6	17.9
潜在顧客	287	31,331	9,149.5	3.0	31.9
一般顧客	170	21,250	8,041.4	2.6	118.5
一時顧客	58	12,057	4,888.7	1.6	248.9

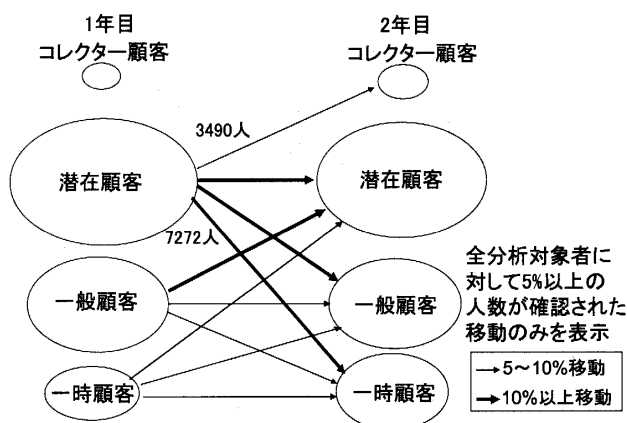


図6 1年目から2年目へのグループ間の移動

⁷ あるセグメントからあるセグメントへの移動人数/移動元のセグメントの顧客数

表4 新たなJ-POPの分類における購買者比率の違い

部分グラフ	コレクター	一時顧客	差
ビジュアルメイン	41.4%	27.2%	14.2%
インディーズメイン	47.4%	37.6%	9.7%
オムニバスメイン	57.1%	51.3%	5.8%
ボーカルメイン	43.1%	37.5%	5.6%

*上記の値は有意水準5%で有意な結果のみを表示

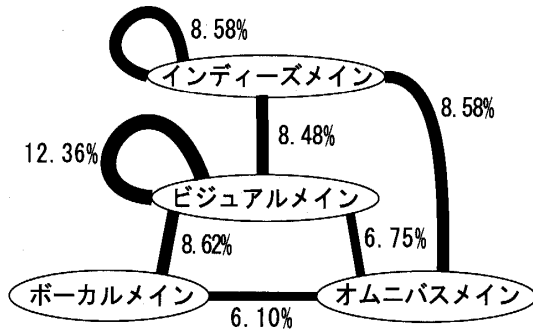


図7 コレクター顧客と一時顧客の購買者比率の差

オムニバスメインやボーカルメインは、あまり顕著な違いは見られず、また各顧客集合の購買者比率も大きい。したがって、これらは多くの人に一般的に好かれるアーティストが集まっていると考えられる。

次に、分類間の併買関係について分析を行う。図7は、コレクター顧客と一時顧客の併買における購買者比率の差（コレクター顧客から一時顧客の購買者比率を引いた値、以下同様に算出）を枝の値としてグラフで表現している。ここでは差が顕著であると思われる絶対値6%以上の接続関係のみを図示している。また、図中のループは、同じ分類を複数回購買した購買者比率を表す自己ループを意味している。ここでは、J-POP以外の大ジャンルとの関係も分析しているが、強い結果としては現れていない。図より、すべての差が正の値であるため、コレクター顧客の購買者比率のほうが大きいことが分かる。最も強力な接続はビジュアルメインの自己ループであり、またビジュアルメインは、他のすべての分類にも比較的強い接続関係を有していることが確認できる。インディーズメインは、ボーカルメイン以外の分類と強力な関係を持っている。これらは、上記の新分類における購買者比率とも整合している結果であり、最も特徴を有しているのは、ビジュアルメインの購買履歴がある顧客であると考えられる。

以上の結果を踏まえ、コレクター顧客の識別を目的とした決定木分析を行う。説明変数として、各顧客の

表5 男性顧客の特徴的な併買行動

併買関係	コレクター	一時顧客	差
女性 avex とその他	46.6%	30.2%	16.4%
女性 avex と宇多田ヒカル	17.6%	5.7%	11.9%
その他と宇多田ヒカル	15.5%	3.8%	11.8%

*上記の値は有意水準5%で有意な結果のみを表示

新しい分類間の併買関係の有無、および、顧客属性やアーティスト購買種類数などRFM値とは関連の少ない説明変数を14個用意した。

決定木モデルを生成するために利用したツールは、Weka⁸のJ48アルゴリズム[7]であり、信頼係数CFの値はデフォルト値の0.25を用いた。ランダムサンプリングによって同数のサンプル数からの識別を出発とし、10回の交差検証法によるモデルの精度は63.8%であった。これは同条件で、新しい分類の購買に関する説明変数を利用しない場合のモデルに比べて、約2%モデルの精度を改善していることが確認された。モデル内で出現したルールを見ると、新しい分類を説明変数に利用しない場合は、年間購買数量が6枚以上というルールが66%（判別対象数2,028）で最も判別精度の高いルールであったが、新たな分類を説明変数として使用することで、以下のような74%（判別対象数436）の判別精度をもつ、より強力なルールを発見することができた。

ルール 年間購買数量が5枚以上で、ビジュアルメインを4枚以上購入する顧客

したがって、購買数量の多い顧客は、将来コレクター顧客へとなることが期待されるが、購買に占めるビジュアルメインの購入の多さが、2つのルールに関する精度の差を生じさせているといえる。以下では、より実用的な分析結果を示す1つの例として、決定木の中で一番強い判別結果が確認された上記ルールに属する顧客436人（コレクター顧客323人、一時顧客113人）に着目して、より詳細な分析結果を示す。

表5は、ビジュアルメイン、ボーカルメイン、インディーズメインの間関係について、コレクター顧客と一時顧客の購買者比率で差の確認できた併買関係を表している。ビジュアルメインについては、前述の4つのサブグループとそれ以外を“その他”という1つのサブグループで表現し、ボーカルメインとインディ

⁸ <http://www.cs.waikato.ac.nz/ml/weka/>

ーズメインは、アーティストを用いてその関係を分析した。その際、男性顧客の結果が女性顧客の結果に比べて顕著であったため、ここでは最も顕著であった男性顧客の結果を示している。

最も差のある併買関係は、女性 avex とその他の併買であり、約16%の差が確認できる。また、女性 avex と宇多田ヒカル、その他と宇多田ヒカルの接続関係も12%弱の差が確認できており、これらの併買関係は、男性のコレクター顧客に顕著な特徴である。また、統計的に有意とまではいえないが、一時顧客に見られる相対的に強い特徴としては、モー娘とモー娘や女性 avex と女性 avex などクローズなアーティストに対する併買を行う傾向が確認できた。これらのことを併せて考えると、コレクター顧客は音楽的な関心の範囲が広く、逆に一時顧客は、音楽的な関心の範囲が狭い傾向があるのではないかと考えられる。すなわち音楽に対する嗜好や興味の多様性が、将来、購買意欲の大きな優良顧客として成長するのではないかと推測される。

単にアーティスト間ではなく、また所与の分類でもなく、購買者が類似していると考えているグループを横断する多様な購買は、小売店が顧客に対するプロモーションを行う上でのキーポイントになることが考えられる。具体的には、顧客全体に何かをプロモーションするのではなく、モー娘グループのような特定のサブグループ内のみを購買する顧客に対して、他の顧客のデータを参考にし、併買する可能性のあるアーティスト（またはサブグループ）をピンポイントでプロモーションしていく方法が有効であろうと考えられる。

このような結果は、クローズなアーティストグループの大きさが様々であるため、単にアーティストの購買種類といった変数で説明することは困難である。また所与の分類は、顧客の購買に関する類似性を反映させた分類ではないために、その違いを明確にすることは難しい。これは今回、グラフ分割手法を利用した分類を利用することによって得られた成果であると考えられる。

5. おわりに

本稿は、グラフ分割問題をCDの購買履歴データに適用することで、既存のアーティストの分類とは異なる、顧客の併買行動を考慮したアーティストの分類を示し、これを利用した顧客の併買を分析した。分割された部分グラフは、顧客の併買行動をよく表現したグラフであり、併買に関するいくつかの特徴を抽出できることを示した。また、顧客に関する分析では、新たな分類を利用することで、コレクター顧客と一時顧客の購買行動の違いを発見し、グラフ分割を適用することの有効性を示した。

本稿で提案したグラフ分割による分類方法は、他のID付きPOSデータにも十分に適用可能であると考えられる。スーパーや百貨店でも商品は実に多様であり、既存の分類は、必ずしも顧客の嗜好によって分類されているものではなく、その多くは業務を効率良く遂行するために作成された分類である場合が多い。ロイヤル顧客など特定の顧客層の購買が、他の顧客層と何か違いがあるならば、商品レベルや所与の分類レベルで相違点を発見することが困難であったとしても、今回示したような併買関係を考慮した新たな分類は、何らかの示唆を与えてくれるのではないかと考えられる。今後、様々なデータへ適用し、実用的な結果を導きたいと考えている。

参考文献

参考文献

- [1] M. R. Garey, D. S. Johnson and L. Stockmeyer: "Some simplified NP-complete problems," *In Proceedings of the sixth annual ACM symposium on Theory of computing*, pp. 47-63 (1974).
- [2] 広島ガス: <http://www.hiroshimagas.co.jp/gas/lip/pre/anq0204.htm>
- [3] A. K. Jain and R. C. Dubes: *Algorithms for clustering data*, Prentice Hall (1988).
- [4] G. Karypis and V. Kumar: "Multilevel k-way Partitioning Scheme for Irregular Graphs," *Journal of Parallel and Distributed Computing*, Vol. 48, No. 1, pp. 96-129 (1998).
- [5] 中原孝信, 森田裕之: "百貨店のクレジット購買データを用いた関連購買による顧客特徴分析," *オペレーションズ・リサーチ*, Vol. 50, No. 7, pp. 488-494 (2005).
- [6] 社団法人日本レコード協会: "2005年度音楽メディアユーザー実態調査報告書," (2006).
- [7] I. H. Witten and E. Frank: *Data Mining*, Morgan Kaufmann (2000).