

二重螺旋モデルを用いたスーパーの顧客行動変化の予兆発見

Double Helical Chance Discovery from Customers-Behavior in Supermarket

大澤幸生 (筑波大学・科学技術振興事業団) 臼井優樹 (筑波大学・日東紡)

福田寿 (CRCソリューションズ) 松尾豊 (産業技術総合研究所) 松村真宏 (科学技術振興事業団)

高山美和 (松下電器産業 (株) システム創造研究所) 相馬浩隆 (筑波大学) 佐橋官 (筑波大学)

連絡先: 〒112-0012 東京都文京区大塚 3-29-1 筑波大学ビジネス科学研究科

osawa@gssm.otsuka.tsukuba.ac.jp

あらまし

本論文ではスーパーマーケットにおける消費者行動を例にとり、まだ求める需要者の少ないニーズも含めて多様なニーズのマーケットにおける意味を理解するような知識発見プロセスを定式化する。チャンス発見の基本的なプロセスモデルである二重螺旋について述べ、データマイニングコンペにおいて提供されたスーパーマーケットデータを基にこの二重螺旋プロセスを展開した事例を示す。得られた店舗経営にとって重要な知識をまとめ、検証を行った。

Yukio Ohsawa (University of Tsukuba, and Japan Science and Technology Cooperation (JST))

Masaki Usui (Nottobo, Inc. and University of Tsukuba) Hisashi Fukuda (CRC Solution Inc.)

Yutaka Matsuo (AIST, Japan) Naohiro Matsumura (JST), Miwa Takayama (Matsushita electronics

Hiroataka Soma (University of Tsukuba) and Tsukusa Sahashi (University of Tsukuba)

Contact Address: 3-29-1 Otsuka, Bunkyo-ku, Tokyo Japan 112-0012, Tel: +81-3-3942-7141

Abstract

The behaviors of consumers buying in a supermarket are dealt, for understanding rare but significant needs among various data from the market. The meaning of consumer-behaviors are underlying a set of products to be bought, and we make a system for visualizing such sets and the relations between sets by KeyGraph. By the Double-Helix Model for chance discovery, we show how the researchers of the market achieved the discovery of unfamiliar, i.e., hidden demands of potential customers and of potential increase in the money of user, in the market. The success of these discoveries are validated statistically.

1. はじめに：決定論的システムの同定とチャンス発見情報を取り入れて行動するまでのヒトは、決定論的システムであるが、このシステムは複雑で未知の要因が非常に多く入力として影響する。それ故、ヒトというシステムは同定も制御も困難を極める。従来、マーケティングにおいて顧客をセグメンテーションしたり商品を関連付けたりする場合は、たとえば次のような作業をその場の必要に応じて組み合わせて実施されてきた。

- 選択式あるいは記述式の質問を一定個数与え、これら質問への回答間の相関を計算する意識調査。
- 顧客属性情報を有限項目について登録してもらい、属性と消費行動の相関を見るもの。
- さまざまな条件（気温や価格など、通常は所与の変数）において同時に購入する商品のセットから消費者の関心をとらえるもの。

これらは、ヒトの行動要因は有限数の項目を知ればほぼ理解できるという近似をヒトの集団について導入し、ヒトなるシステムを同定する方法といえる。この近似は、大規模な集団では多様な個性が相殺しあうか、または統計的にはノイズとして扱えるという仮定のもとでは正当性を確保できる。

ところが近年、インターネットを通して自己主張が容易に公開できるなどの情報インフラの変化に伴い、これまでは明示的に理解されなかったような個性を持つ人々が社会に影響する度合いが強まっている。このような背景から、仮に少数であっても社会に影響する個性的な、しかも言葉になって現れにくい嗜好（マーケティングにおけるチャンス）を発見し、その意義を理解できるマーケティング調査手法が求められるようになった。

本論文ではスーパーマーケットにおける消費者行動を例にとり、まだ求める需要者の少ないニーズも含めて多様なニーズのマーケットにおける意味を理解するような知識発見プロセスを定式化する。2節でチャンス発見の基本的なプロセスモデルである二重螺旋について述べ、3節以降でOR学会「データマイニングコンペ」において提供されたスーパーマーケットデータを基にこの二重螺旋プロセスを展開した事例を示す。4節に、この結果得られた店舗経営にとって重要な知識をまとめ、検証を行う。

2. チャンス発見プロセスの二重螺旋モデル

チャンス発見の実行において最も重要となるのは、発見プロセスを設計することである。現在のところ、計算機が単独で人にとって重要なチャンスを見出すようなアルゴリズムは存在しない。それは、チャンス発見を行う人がある事象をチャンスとして受け入れる心理状態にいるかどうかという計算機と独立の条件が、チャンスかも知れない事象の遭遇に気づくことかどうかに影響を及ぼすからである。それ故、チャンス発見プロセスにおける、人の心理状態の遷移モデル構築が課題となる。

チャンスに基づく行動の効用は、人と環境の相互作用の中で獲得されるものであり、初期から一意に定められるものではない。しかし、全くチャンスの良さを定めないことには、チャンス発見のプロセスを設計する指針を立てることができない。そこで[2]では、チャンス进行评估するメタな指標として(1)行動の提案可能性、(2)気付かれにくさ、そして(3)提案が採択され実行されていくという成長性である。[3]ではこれらを向上させるために一人ではなく数人のグループを構成し、グループディスカッション(GI)における人と人との間でのチャンスに基づく提案や採択を通してチャンスを選択してゆくプロセスを実現した。ここでは、意思決定環境についての環境データ(販売者における購買データなど)に対する視覚的データマイニングの出力図を見て、少数参加者から主観的に評価されたチャンスへの評価が、他の人にも共有されてゆくグループダイナミクスを利用したわけである。

このような方法は、データマイニングを用いるプロセスを研究対象としている点ではFayyadの知識発見プロセス[4]の考えと共通しているが、未知要因を直接観察したり、人の関心を新たなチャンスに向ける心的コンテキストの遷移には研究の余地があった。一方、チャンス発見プロセスは次の各状態からなるものとする。

- ・ **関心**：遭遇したチャンスについて、自分の意思決定にとっての意義を新たに理解したいと感じる状態。
- ・ **理解**：ある特定のチャンスが自分の意思決定に対して持つ意義を理解した状態。
- ・ **発案**：チャンスを利用した行動を発案し、場合によりグループの他者に提案する。
- ・ **実行か模擬**：実際または仮想的に行動を行う。
- ・ **評価と選択**：実行または模擬の結果に基づき、行動を評価し選択する。この評価において、新たな関心が生まれる場合は次のサイクルの「関心」へ遷移。

さらに、主体の思考過程を捕らえた**主体データ**を併せ用いることにより、人が自分自身の考えを表出化することを支援する過程が図1の二重螺旋モデルである。

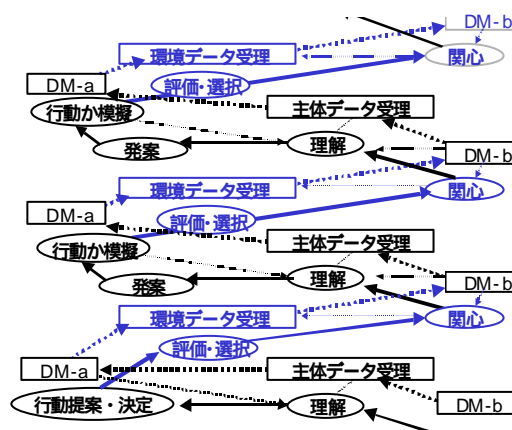


図1 チャンス発見過程の二重螺旋モデル（図中1-hなどはこの後の事例の各ステップに相当）

この二重螺旋モデル（[5,6]に準拠）では、人のチャンス発見の螺旋過程（関心 理解 発案 行動または模擬評価と選択 関心という繰り返し）と平行して、その人が関心を寄せる環境データとその人の思考内容についてのテキストデータのマイニングを計算機が繰り返す（人が環境データのマイニング結果を理解する最中も、その理解過程までデータとして取り込む点で「二重」と呼ぶ）。

実際、環境データのマイニング結果（DM-a）についてのグループインタビュー（GI）における議論の各参加者の考えを、カードに思い当たるだけ書きこんでもらい、そのテキスト内の単語の相関関係を図示する視覚的なテキストマイニング（DM-b）の出力図を見ながらの議論を通してチャンス表出化してゆくことができた。このシステムで顧客像を把握した結果、顧客の購買（POS）データからスーパーマーケットの購入金額増加の鍵となる商品や、その店の経営状態のおおまかな変化を示す予兆を発見することができた。その内容を以下に述べてゆく。

3. スーパーマーケット顧客の潜在ニーズの理解

この問題の要点は、消費者の購入が金額として増加したり減少したりする原因を突き止め、これを助長または抑制する策を求めることである。その原因や策は、顧客行動の変化を実現させるものであることが望ましい。われわれの採った方法は、以下の二重螺旋プロセスである。

Step-1 DM-b：視覚的データマイニング

（KeyGraph[7]：後述）を、商品を大雑把に分類した販売履歴データに適用。

Step-2 **理解・発案**：KeyGraphの出力図を見て、消費者のおかれていた情景を想定するなどして、消費者行動の動機に関する仮説、それに基づく同スーパーへの経営戦略をグループワークで乱立させる。これらの仮説を接合した文章を生成。

Step-3 DM-aによる理解・発案の洗練 評価：

Step-2の文章へのKeyGraph適用により、互いに似た消費者情景を想定した人達と同グループとなり分散的に議論して情景についての仮説を深化させ、支持に値する仮説を残してゆく。それら仮説に基づいて実際に行動をとった場合のリスクなどを評価し、新たに考慮すべき問題点を抽出。

Step-4 **新たなデータの獲得から螺旋上の進歩**：新たに考慮すべき点が提案されればStep-1に戻る。

Step-5 **仮説の検証**：Step-3で得られた仮説のうち、入手できるデータから立証できるものは仮説実証型のデータ解析で立証。矛盾が発生した場合は未知仮説の存在する可能性があるため、必要に応じてStep 1に戻る。

この手順に従い、われわれ著者7名がグループワークとなって進んでいった。以下、各ステップの詳細を述べる。

Step 1~2: 初期状態は購入金額の変化への関心から開始し、環境データとしては、購入金額が急激に増加している16人および減少している12人の顧客について購買履歴（POS）データを用いた。データは、一行ごとに購入した一商品が並んでおり、「大分類、中分類、小分類」という順にカンマで区切られて与えられたテキストファイルである。このように蓄積されたデータのうち関心の対象となっている一部を、次例（実際のデータからの抜粋ではない）のように適切な個所で区切り、区切り個所にピリオド（.）を挿入した文章のような形式にして、これをKeyGraphによって解析した。区切りの間は文と呼ばれ、同じ文に出現する商品は同時購買されたとみなす。

シコウヒン, ペットインリヨウ, ペットリヨクチャ 500ML
シコウヒン, ピンカンインリヨウ, リポビタンD 100ML
サニタリ-, セイリヨウヒン, ソフィボデイフィットスリム
サニタリ-, エイセイヨウヒン, コバヤシメガネクリ-ナ
ブタニク, ブタスライス, SPFブタコマキレ 250G, 250, 1
KeyGraphに期待したのは、同時購買されることの多い商品集合、そして集合と集合の間に潜む、潜在的ニーズをわれわれが想像できるようにすることである。すなわち、同時に買われることの多い商品の集合を「購買の土台となる、顧客にとって当たり前の習慣化された購入パターン」とみなす。これらは、黒いノードで図示される商品たちと、それらの間を結ぶ黒いリンクから構成される。

そして、土台の商品セットと同時に購買される商品（屋根）は、頻度が少なくても顧客の習慣的な生活と広範に絡み合った潜在的に重要なニーズであると解釈できる。屋根は、赤いノードの商品として図示され、黒ノードと黒リンクのクラスとは、赤いリンクで連結される。そして、赤いリンクのうち特に重要なものが集まるノードは、土台中あっても屋根であってキーノードと見られ、緑色の円で囲まれて図示される。この視覚化を、図1のDM-bから「理解」に至るステップに適用した。

図2は、最初のKeyGraph適用結果である。購入金額の増加者と減少者では、購入パターンが異なることがわかる。これを見て、本研究グループのメンバー7人で3つずつの仮説をメールで集めた。例えば金額増加者につき「オカズ」「フクロソクセキメン」「ベントウ」などものすごく忙しさが現れているように思えます。子供がいる若い世帯も、会社勤めの兼業主婦もどちらも忙しく、食事を作る時間を短縮したい人々です。やはり「ソウザイ」など出ていますねとの仮説が出たが、別のメンバーは逆に「野菜をメインとして、出汁をきちんと作るということは、ちゃんと料理している家庭らしい」と仮説を提出した。問題は、これらのうち、どの仮説が支持に値するかである。

異なる食習慣の人たちが図2の各図に含まれるとすれば、複数の矛盾する仮説があってもよいことになるので、いずれか一仮説を選択するのではなく、それぞれの仮説

の支持可能性を、実際に想定できる家庭をあらわしたものとになっているかという視点で評価することになる。

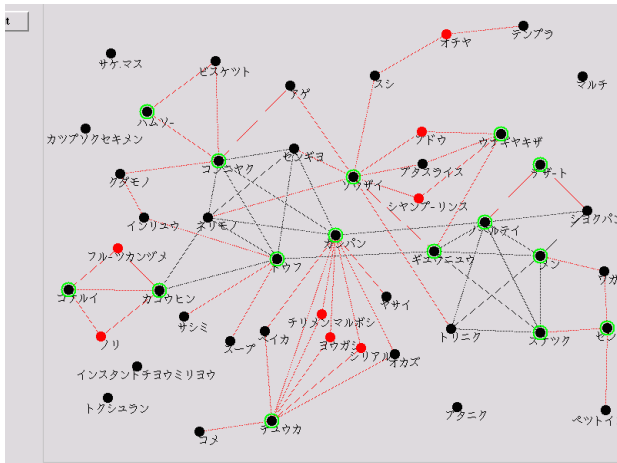
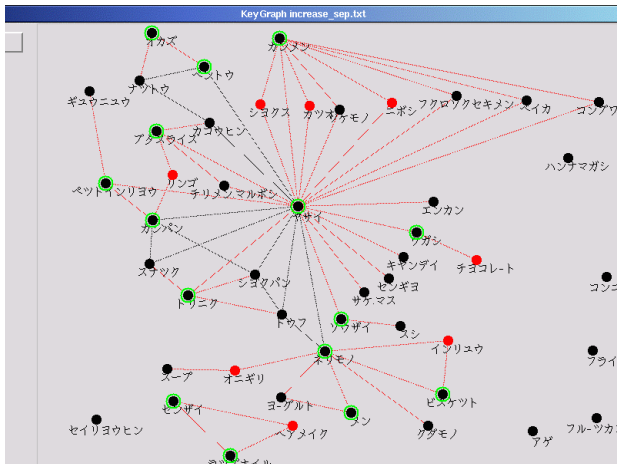


図2 購入金額の単調増加者（上）と単調減少者（下）について、購入金額変化期間の最初の3日の購買を一文として合成したデータへのKeyGraphの結果。概要を把握するため、各商品は大分類名で表したデータを用いた。

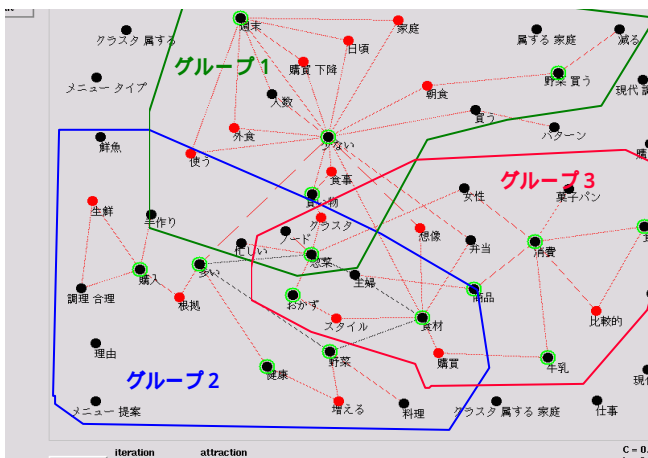


図3 持ち寄った仮説に対するキーグラフの結果

Step 3 7名それぞれの考えた顧客像を仮説として電子メールで送ってもらい、これらをテキストとして接合したものを主体データとしてキーグラフで処理した。この場合は各単語がアイテムで、各仮説が共起の単位となる。この結果、図3を得た。

各人はこの図で自分の仮説の該当する位置を図で示し、その位置に応じて図3の枠で囲んだ3つのサブグループに分かれ、それぞれの想定する顧客像を言葉で表現できるようになってから再度で集合して議論した。これは、7人のうち単に声の大きな人の意見が通るようなことを避け、様々な顧客像をできるだけ考慮して接点を探るために、自分の考えた顧客像を議論者が一旦整理しておく必要があるからである。再会しての議論の結果

- ・購入金額が減る場合
 - ?近い店がメインディッシュをとってしまった
 - ?ついで買い
 - ?減りつつも、惣菜に固執
- ・購入金額が増える場合
 - ?野菜を中心に家庭で食事を作る
 - ?大家族で消費する家庭
 - ?子供がいる家族

というように、両方のタイプの顧客について顧客の特徴が想定された。また、惣菜が図2で重要な位置につけている点を全員が認め、

- 惣菜のプロモーションに注力
 - 作る人の支援として献立推奨
- などの戦略が実行可能案として指示された。

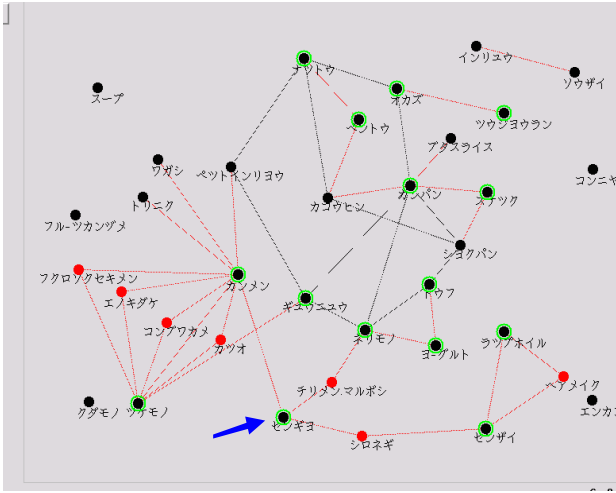
Step 4 Step 3の理解と戦略を具体的に評価する上で、

- 野菜という大分類が具体的に何をさしているのか（大根なのか、にんじんなのか等）
- 時間軸方向の変化を考えると、各時期にそれぞれの人の購買内容はどう変化したのかを知る必要が生じた。

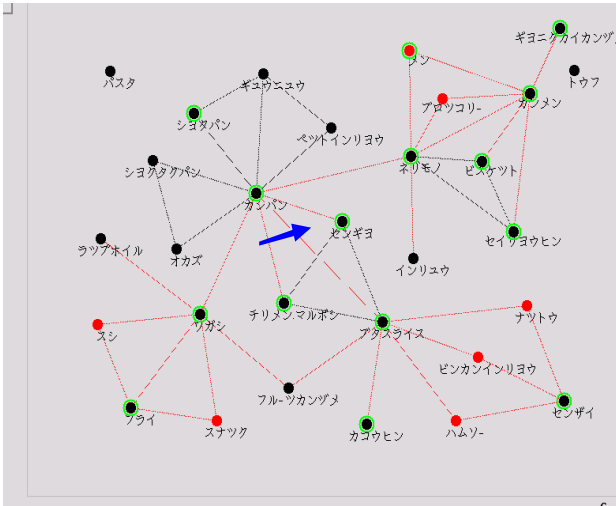
Step 1 ~ 3（二週目） 野菜を中分類まで詳細化したのに加え、購入金額の単調増加または減少した期間のうち、さまざまな個所の期間（3日間）について購買履歴データを解析した。例えば、購入金額がある時期単調増加した客それぞれの、増加期間の最初の3日の購買データを接合し、各顧客の当該3日の購買を一文とした文章データをKeyGraphに処理させた結果、図4最上図を得た。さらに図4の他の図と、同様のことを購入金額が減るケースについての図5を得た。

図4からは、太矢印を付けたセンギョ（鮮魚）が全体の中で占める位置づけが中心クラスタの一部に以降し、最終的に食卓のデザインに欠かせない食材になっていることがわかる。逆に図5では、最初は二重丸（元図では、緑の丸で囲まれていた）であった鮮魚が、そのうちローカルな位置づけでキーノードではなくなっている。

a. 増加期間の最初の1~2日



b. 増加期間の最初の2~4日



c. 増加期間の最初の4~6日

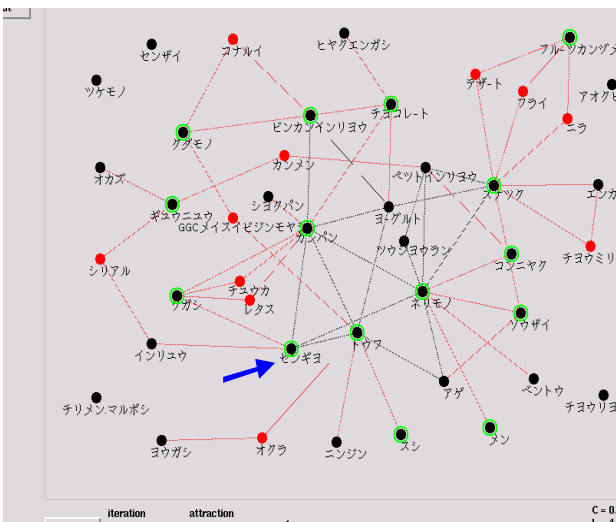
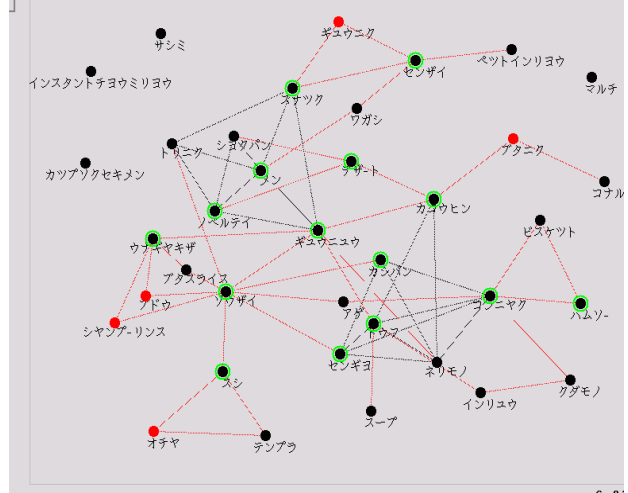
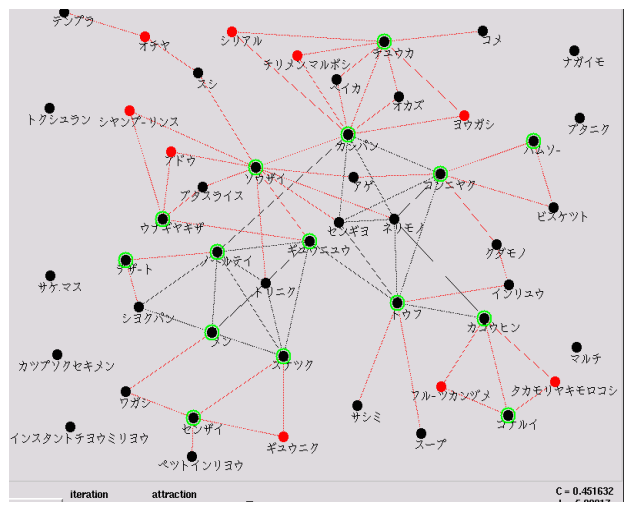


図4 購買金額が単調増加した時期の顧客たちの、購買データをKeyGraphで処理した結果。

a. 減ってきた人のラスト4,5,6回目の購買



c. 減ってきた人のラスト3,4,5回目の購買



d. 減ってきた人の最後の3購買を別々に一文に .

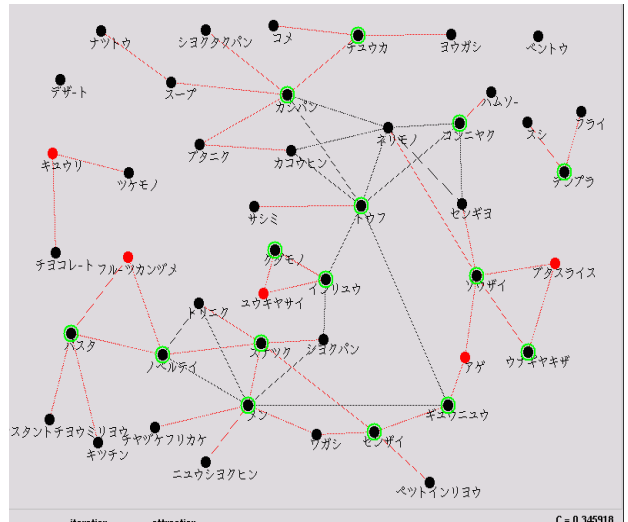


図5 購買金額が単調減少した時期の顧客たちの、購買データをKeyGraphで処理した結果。

そこで、7人の仮説はひとつには「このスーパーで定着して行く人は、鮮魚に魅力を感じているらしい」に落ち着いた。同様のことを、購入金額が増える場合にも行った結果、次のような仮説の提案または受理に行き着いた。

・購入金額が減る場合

- 近い店が若者のメインディッシュをとってしまった
- 若い人のいる家庭が減っていく
- 鮮魚への関心を失っていく人が離れていく
- ついで買い・忙しい

・購入金額が増える場合

- ?食卓のメインは野菜で、菓子は減少している
- ?忙しくても料理は家で、ダシまで作る
- ?子供のいる家庭やシニアのいる家庭など多様な家庭
- ?鮮魚を好む客である

というように、両方のタイプの顧客について顧客の特徴がさらに具体的に想定された。そして実際、このスーパーのある地域は魚料理が非常にメジャーであること、若者がいきがちなスーパーがこのデータ期間中にこのスーパー付近に出来ていたという外部情報があとで入り、われわれの仮説はこれらの事実と一致していることがわかった。このような結果から、魚料理の献立推奨や、シニア向けのマーケティング戦略をとるなどの提案を得た。

ここまでで述べたのがチャンス発見の本質を見せる部分プロセスである。以上のように、仮説検証の前に行うべき仮説創出の支援が可能となった。

4. 仮説の検証と今後の課題

以上において生まれた顧客の潜在的な需要の発見は、次の5つの内容であり、まだ仮説の段階である。このまま提案として受け入れて意思決定を行うことも可能であるが、可能であれば仮説を検証することも、意思決定の信頼性を確保する意義はある。

仮説1：鮮魚は購入金額の高い家庭の典型消費財である。

仮説2：購入金額が上昇する家庭は、比較的高年齢である。

仮説3：購入金額が減少する家庭は、比較的若年である。

仮説4：購入金額が上昇する人は、だし原料であるカツオ、昆布、若芽の購入比率が高い。

仮説5：ネリモノは、購入金額が上昇する家庭で料理に用いられる。

以下、各仮説について統計的に検証を試みた結果を示す。

仮説1：「店に定着して購入金額の高い客は、使用金額に対して鮮魚の購入比率が高い」という仮説が支持された。

仮説2,3：t-検定によると有意水準5%で支持されたが、これが実際にスーパーマーケットにおいて持つ意味は明らかではない。というのは、家庭にシニアの構成員がいても若年層がいても、専業主婦であれば構成員と無関係に主婦が買い物に来るからである。むしろ、「購入金額が下降する人は、使用金額のうちスナックの購入割合が高い」「購入金額が下降する人は、使用金額のうち菓子パ

ンの購入割合が高い」がともに有意水準5%のt検定で支持されたことの方が、商品を家庭に持ち帰ったあとの消費を捉えている点で仮説2, 3の検証としては根拠になる。

仮説4, 5：詳細は省略するが有意水準5%で支持された。以上のとおり、3節で発見された顧客の潜在ニーズの傾向は、データによって改めて別の角度から検証された。しかしチャンス発見は、データ中の有意な傾向から外れたところに意思決定のきっかけを見出す狙いである。その意味からすると、仮説が統計的な検定で支持され、新知識としての信頼性が裏付けられたことよりも、提案された未経験の戦略「魚料理を含む献立推奨サービスをする、シニア市場を増やせる」などを検証する方法の開拓がポイントになる。現在では市場でテストする方法がマーケティング分野で盛んに開発され始めているが、そのような動向と連携して進めることが今後の課題である。

5. 結論

二重らせんモデルというのは、チャンス発見を全体を動かすための、モデルの一例に過ぎない。チャンス発見システムを時計にたとえると、計算機がつかさどるデータマイニングのアルゴリズムは秒針を動かす一つの歯車に過ぎない。この時計のぜんまいは人である。

謝辞：本研究は、オペレーションズリサーチ学会2001年度データマイニングコンペにおいて提供された市場データを基に行いました。データ提供者に記して感謝します。

参考文献

- [1] Simon, H.A., *Administrative Behavior*, (1945) 松田武彦ほか訳『経営行動』ダイヤモンド社 (1965)
- [2] Ohsawa, Y.: Chance Discoveries for Making Decisions in Complex Real World, *New Generation Computing*, Vol.20 No.2 (2002)
- [3] Fukuda, H. and Ohsawa, Y., Discovery of Rare Essential Foods by Community Navigation with KeyGraph - An Introduction to Data-based Community Marketing, in *Proc. KES2001* (2001)
- [4] Fayyad, U., Shapiro, G.P. and Smyth, P., From Data Mining to Knowledge Discovery in Databases, *AI magazine*, Vol.17, No.3, 37--54 (1996)
- [5] 大澤・臼井ほか：OR学会・マーケティングエンジニアリング部会研究会データ解析コンペティション第一回報告回資料（入手困難な場合は本論文著者に連絡下さい）(2001)
- [6] Yukio Ohsawa and Yumiko Nara, Understanding Internet Users on Double Helical Model of Chance Discovery Process, Invited to talk in *the 17th IEEE International Symposium on Intelligence Control (ISIC'02)*, Vancouver
- [7] Yukio Ohsawa, Nels E. Benson and Masahiko Yachida, *KeyGraph: Automatic Indexing by Co-occurrence Graph based on Building Construction Metaphor*, *Proc. Advanced Digital Library Conference (IEEE ADL'98)*, pp.12-18 (1998)