

2. 2. 1. 顧客と商品のカテゴリマイニング

(1) ダイレクトメールを用いた顧客アンケート調査と POS データとの関連付けデータの作成

流通量販店における顧客のライフスタイルやパーソナリティと商品の購買傾向から生活者視点のカテゴリを調査するため、ダイレクトメールを用いた顧客アンケート調査と、顧客の ID 付き購買行動履歴である ID 付き POS データ（以下、ID-POS データ）との関連付けデータの作成を行った。これらのアンケート調査と ID-POS データの提供は兵庫県を事業エリアに 146 店舗を展開する流通量販店である生活協同組合コープこうべの協力を得て実施された。

(1-1) ID-POS データの抽出とカテゴリマイニングサーバの構築

ID-POS データの内容は、コープこうべの店舗で記録された 2008 年 10 月 1 日から 2009 年 9 月 30 日の期間における購買履歴データを利用した。会員カード番号は完全にスクランブル化され、個人が特定できない状態で分析された。コープこうべでは会員カードを利用した ID-POS システムを導入しているため、顧客 ID と商品の購買履歴をデータとして記録してある。この期間での全データのトランザクション数は約 6.7 億件（669511467 件）である。本 ID-POS データにはポイントカード ID、店舗、利用日時、各商品に対して 1 対 1 対応の商品コード、購買数、購買価格の情報が含まれている。また、商品コードと関連付けることができる流通量販店独自の商品大分類・中分類・小分類・商品名のマスタが利用可能で、それぞれの分類数は 41、154、950、364397 である。提供されたデータの構造図は図 2. 2. 1-1 の通りである。

また、この大規模なデータに対しても分析加工作業が実行できるよう、カテゴリマイニング用の高性能データサーバー（各スペックはプロセッサ 2×2.93GHz Quad-Core Intel Xeon、メモリ 32GB 1066Hz DDR3）を構築し、そこでデータ環境の検証を行った。そのデータベース内の構造を図 2. 2. 1-2 に示す。

(1-2) ダイレクトメールアンケートの実施

顧客のライフスタイルやパーソナリティを把握するため、顧客へのアンケート調査を実施した。2009 年 12 月下旬に同流通量販店の会員約 16,900 人に対し図 2. 2. 1-3 の内容のダイレクトメールを用いてアンケートを送付し、1 月上旬までにその内 3981 名から回答を得た。アンケートの項目数は全 35 問であり、デモグラフィック特性として年齢・性別・家族構成・家族人数・職業を設定した。また、来店頻度、食生活、健康と食への意識、ダイエットへの意識、消費傾向などの質問項目も設定した。さらにビッグ 5 法を用いパーソナリティを把握するための質問項目を設定した。これらは消費者のライフスタイルや価値観に着目し、様々な先行研究を元に設計したものである。回答は質問項目への当てはまりの強度順に“良く当てはまる”から“全く当てはまらない”までの 4 カテゴリの選択式である。4 つの選択肢に対し強度順に 4、3、2、1 の得点を与えている。

このダイレクトメールアンケートの顧客 ID と ID-POS データの顧客 ID を関連付けることで、3981 名分の顧客アンケートと商品購買履歴を関連付けたデータを作成した。



図 2. 2. 1-1 抽出された ID-POS のデータベース構造図

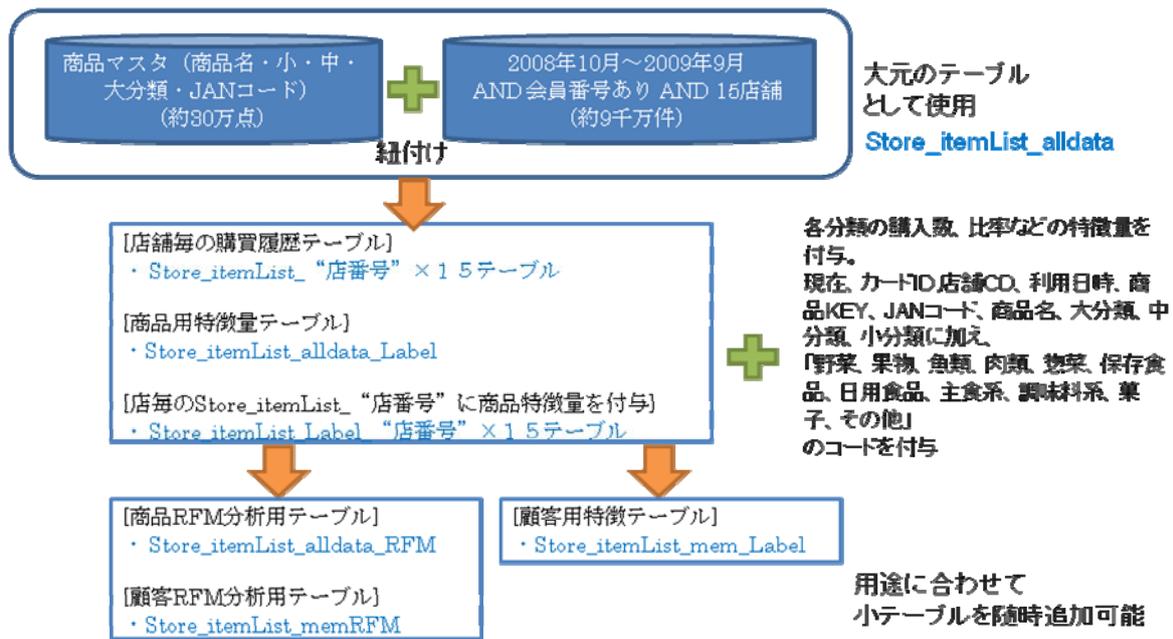


図 2.2.1-2 カテゴリマイニングサーボ内のデータベース構造

アンケートご協力のお願い

印刷用アンケートをご利用いただきありがとうございます。アンケートでは、より魅力的なお店づくりを目指し、組合員の皆様からご意見を伺う取り組みを行っています。

そこで今回、より良いお店づくりを反映させていただきたく、組合員様のご意見を伺えるアンケートを送付させていただきます。ご記入いただいた回答と組合員様の利用状況やご住所などをとらえ、毎日の生活と実際の買い物との関連性を調べ、アンケートの改善や商品づくりの参考にさせていただきます。ご協力いただけます場合は、大変お手数を重ねさせていただきます。右のアンケートについて、来店バリエーションにご記入いただき、12/25(金)までにポストへ投函してください。

※この調査の集計・統計処理は個人が特定できない方法で実施し統計情報として扱います。

※アンケートは、アンケート用紙を印刷していただいた組合員様のうち約20,000名に送付させていただきます。

生活協同組合コースこうべ

〒658-8555 兵庫県神戸市東灘区住吉本町1丁目3番19号
 無料ダイヤル 0120-602876 (10:00~18:00) 011-2511-1111(日祝祭日)

郵便はがき
 6 5 8 8 7 9 1 0
 兵庫県神戸市東灘区住吉本町1丁目3番19号
 生活協同組合コースこうべ
 MD政策推進室 行

アンケートのご回答は右の回答欄にご記入ください。

あなたのことや普段の生活について教えてください。
右の回答欄の該当する番号に○をご記入下さい。

① 性別 (1. 男性 2. 女性)
 ② 年齢 (1. 10代 2. 20代 3. 30代 4. 40代 5. 50代 6. 60代 7. 70代以上)
 ③ 家族構成 (1. 単身 2. 夫婦のみ 3. 親子 (下の子が小学生以下) 4. 親子 (下の子が中学生以上) 5. 親子 (下の子が大学生以上) 6. 二世帯世帯 7. 三世帯世帯)
 ④ 家族人数 (1. 1人 2. 2人 3. 3人 4. 4人 5. 5人 6. 6人以上)
 ⑤ ご職業 (1. 主婦 2. 会社員 3. 学生 4. 自営業 5. パートタイム/アルバイト 6. その他)
 ⑥ どのくらいの頻度で食料品を買いに行きますか?
 (1. ほぼ毎日 2. 2日に1回 3. 3,4日に1回 4. 週に1,2回程度 5. 週末のみ)
 ⑦ 食料品の購入に関して、どのくらいクーポンを利用されていますか?
 (1. ほとんど利用しない 2. 特売の日だけ 3. 月に1回程度 4. 月に2,3回 5. 週に1回 6. 週に2,3回 7. それ以上)
 ⑧ どのくらいインターネットセンターを利用されていますか?
 (1. ほとんど利用しない 2. 特売の日だけ 3. 月に1回程度 4. 月に2,3回 5. 週に1回 6. 週に2,3回 7. それ以上)

⑨ どのくらいドラッグストアを利用されていますか?
 (1. ほとんど利用しない 2. 特売の日だけ 3. 月に1回程度 4. 月に2,3回 5. 週に1回 6. 週に2,3回 7. それ以上)
 ⑩ コープをご利用いただく際の主な交通手段をお教えてください。
 (1. 徒歩 2. 自転車 3. バイク 4. 自動車 5. バス 6. 電車)
 ⑪ 平日1日のアレシと休憩時間はどのくらいですか?
 (1. ほとんどない 2. 1時間未満 3. 1,2時間 4. 3,4時間 5. 5時間以上)
 ⑫ 家事 (食事、洗濯、買い物など) に平日どのくらいの時間を費やしていますか?
 (1. ほとんどない 2. 1時間未満 3. 1,2時間 4. 3,4時間 5. 5,6時間 6. 7,8時間 7. それ以上)
 ⑬ ご自宅夕食をとられるのは、週に何日ですか?
 (1. ほぼ毎日 2. 4日~6日 3. 1日~3日 4. ほとんど外食)
 ⑭ ご自宅朝食をとられるのは、週に何日ですか?
 (1. ほぼ毎日 2. 4日~6日 3. 1日~3日 4. 食べない)
 ⑮ ご自身や家族の健康で特に心配していることはありますか?
 差し支えない範囲でお答えください。(複数回答可)
 (1. 肥満 2. 高血圧 3. 高血糖 4. 内臓脂肪 5. アルコール 6. 関節痛 7. 神経痛 8. 腰痛 9. その他 10. 特にない)

⑯ 画面回答欄
 ⑰ 画面回答欄
 ⑱ 画面回答欄
 ⑲ 画面回答欄
 ⑳ 画面回答欄
 ㉑ 画面回答欄
 ㉒ 画面回答欄
 ㉓ 画面回答欄
 ㉔ 画面回答欄
 ㉕ 画面回答欄
 ㉖ 画面回答欄
 ㉗ 画面回答欄
 ㉘ 画面回答欄
 ㉙ 画面回答欄
 ㉚ 画面回答欄
 ㉛ 画面回答欄
 ㉜ 画面回答欄
 ㉝ 画面回答欄
 ㉞ 画面回答欄
 ㉟ 画面回答欄
 ㊱ 画面回答欄
 ㊲ 画面回答欄
 ㊳ 画面回答欄
 ㊴ 画面回答欄
 ㊵ 画面回答欄
 ㊶ 画面回答欄
 ㊷ 画面回答欄
 ㊸ 画面回答欄
 ㊹ 画面回答欄
 ㊺ 画面回答欄
 ㊻ 画面回答欄
 ㊼ 画面回答欄
 ㊽ 画面回答欄
 ㊾ 画面回答欄
 ㊿ 画面回答欄

回答欄

あてはまる数字に○をつけて下さい。

| | |
|------------------------|-----------|
| ① 1・2 | ⑮ 1・2・3・4 |
| ② 1・2・3・4・5・6・7 | ⑯ 1・2・3・4 |
| ③ 1・2・3・4・5・6・7 | ⑰ 1・2・3・4 |
| ④ 1・2・3・4・5・6 | ⑱ 1・2・3・4 |
| ⑤ 1・2・3・4・5・6 | ⑲ 1・2・3・4 |
| ⑥ 1・2・3・4・5 | ⑳ 1・2・3・4 |
| ⑦ 1・2・3・4・5・6・7 | ㉑ 1・2・3・4 |
| ⑧ 1・2・3・4・5・6・7 | ㉒ 1・2・3・4 |
| ⑨ 1・2・3・4・5・6・7 | ㉓ 1・2・3・4 |
| ⑩ 1・2・3・4・5・6 | ㉔ 1・2・3・4 |
| ⑪ 1・2・3・4・5 | ㉕ 1・2・3・4 |
| ⑫ 1・2・3・4・5・6・7 | ㉖ 1・2・3・4 |
| ⑬ 1・2・3・4 | ㉗ 1・2・3・4 |
| ⑭ 1・2・3・4 | ㉘ 1・2・3・4 |
| ⑮ 1・2・3・4・5・6・7・8・9・10 | ㉙ 1・2・3・4 |
| ⑯ 1・2・3・4 | ㉚ 1・2・3・4 |
| ⑰ 1・2・3・4 | ㉛ 1・2・3・4 |
| ⑱ 1・2・3・4 | ㉜ 1・2・3・4 |
| ⑲ 1・2・3・4 | ㉝ 1・2・3・4 |
| ⑳ 1・2・3・4 | ㉞ 1・2・3・4 |
| ㉑ 1・2・3・4 | ㉟ 1・2・3・4 |
| ㉒ 1・2・3・4 | ㊱ 1・2・3・4 |
| ㉓ 1・2・3・4 | ㊲ 1・2・3・4 |
| ㉔ 1・2・3・4 | ㊳ 1・2・3・4 |
| ㉕ 1・2・3・4 | ㊴ 1・2・3・4 |
| ㉖ 1・2・3・4 | ㊵ 1・2・3・4 |
| ㉗ 1・2・3・4 | ㊶ 1・2・3・4 |
| ㉘ 1・2・3・4 | ㊷ 1・2・3・4 |
| ㉙ 1・2・3・4 | ㊸ 1・2・3・4 |
| ㉚ 1・2・3・4 | ㊹ 1・2・3・4 |
| ㉛ 1・2・3・4 | ㊺ 1・2・3・4 |
| ㉜ 1・2・3・4 | ㊻ 1・2・3・4 |
| ㉝ 1・2・3・4 | ㊼ 1・2・3・4 |
| ㉞ 1・2・3・4 | ㊽ 1・2・3・4 |
| ㉟ 1・2・3・4 | ㊾ 1・2・3・4 |
| ㊱ 1・2・3・4 | ㊿ 1・2・3・4 |

※この調査の集計・統計処理は個人が特定できない方法で実施し統計情報として取り扱います。

アンケート締め切り
12/25(金)までに投函してください。

STEP 1 記入頂きましたら、ハガキ部分を切り取って、ポストにご投函ください。

STEP 2 アンケートにご回答いただき...

STEP 3 ポストへ投函

ご協力ありがとうございました。
 COOP 生活協同組合コースこうべ

図 2.2.1-3 ダイレクトメールアンケート

(2) カテゴリマイニング技術開発
 (1) で作成したデータを用いて、顧客のライフスタイルやパーソナリティを取り込んだ顧客と

商品カテゴリマイニング分析手法の開発を行った。ここでは図 2.2.1-4 に示すような顧客のライフスタイルと商品の購買傾向の関係を示し、ベイジアンネットを用いて各変数の構造を確率モデル化した。

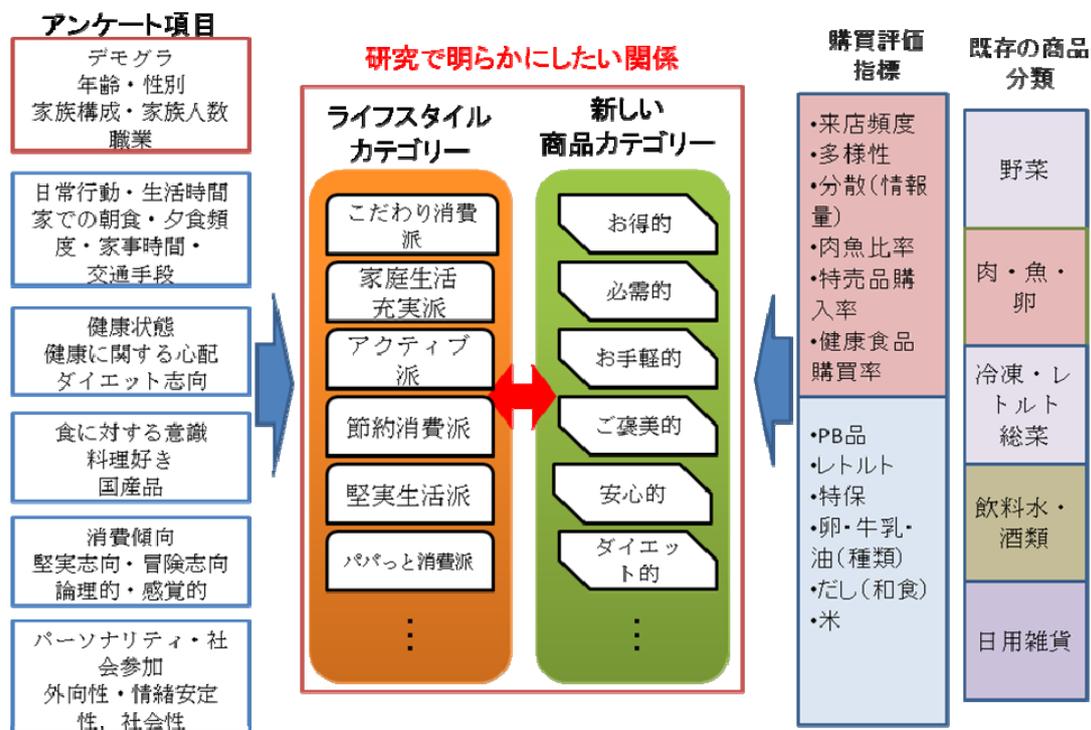


図 2.2.1-4 生活者モデルの基本構造

(2-1) 因子分析による顧客ライフスタイルカテゴリーの抽出

因子分析によりアンケートデータから顧客の消費・生活因子の抽出を行った。ここでは Kaiser の正規化を伴うバリマックス法を用いた。その結果、以下に示す特徴的な 6 つの因子を抽出した。各因子得点を表 2.2.1-1 に示す。ここでは、それぞれの因子を以下のように特徴付け、6 つのライフスタイルに関わるカテゴリーを抽出した。なお、カテゴリーの名称は、現在、暫定的に付与したものであり、今後、精査が必要である。

[第 1 因子]「こだわり消費派」：高くても健康に良いものを選び、産地への関心、こだわりのブランドがある。

[第 2 因子]「家庭生活充実派」：料理が好きで食事も生活も充実している。気分も安定している。

[第 3 因子]「アクティブ消費派」：外向的で、新商品や話題の商品は試しに買ってみる。ただ無駄遣いは多い。

[第 4 因子]「節約消費派」：チラシを見てお得な商品を買う。安ければ少々遠い店にも行く。高い商品は買わない。

[第 5 因子]「堅実生活派」：几帳面で家計簿をつけ、無駄遣いはしない。毎日の献立はスーパーに行く前に決める。

[第 6 因子]「パパッと消費派」：スーパーでの買い物はできるだけ早くすませたい。お弁当を作ることがある。

以上の 6 因子を顧客の消費・生活因子として顧客のライフスタイルカテゴリーと定義した。

表 2. 2. 1-1 各質問項目の因子得点

| | 回転後の因子行列 ^a | | | | | |
|-----------------|-----------------------|------------|------------|------------|------------|------------|
| | 因子 | | | | | |
| | 1 | 2 | 3 | 4 | 5 | 6 |
| Q16バランスの良い食事 | 0.2336805 | 0.5484452 | 0.0408709 | 0.0037327 | 0.1557524 | 0.0594485 |
| Q17毎日の生活が充実 | 0.0855746 | 0.6026855 | 0.1764881 | -0.0332687 | 0.0331666 | 0.0867818 |
| Q18料理好き | 0.1962424 | 0.5076026 | 0.2117116 | 0.0818154 | 0.0677555 | -0.0316921 |
| Q19お弁当 | -0.0331312 | 0.1239969 | 0.130132 | 0.1034101 | -0.0367461 | 0.1442449 |
| Q20低カロリー | 0.3140799 | 0.1959978 | 0.0373392 | 0.097736 | 0.0732867 | 0.0579539 |
| Q21家計簿 | 0.0979278 | -0.0170043 | 0.0381553 | 0.1094744 | 0.382077 | 0.0663099 |
| Q22寝れば遠くても | 0.0126347 | -0.0280407 | 0.1280185 | 0.5836937 | 0.0560822 | 0.0859867 |
| Q23チラシお得 | 0.1101945 | 0.0037253 | 0.0488645 | 0.7052306 | 0.0795292 | 0.0516476 |
| Q24スーパー早くすませたい | 0.0565589 | -0.0340723 | -0.0053603 | 0.0636841 | 0.0214505 | 0.4885447 |
| Q25献立はスーパーで | 0.0558405 | -0.0295026 | 0.1124582 | 0.0424565 | -0.2193077 | 0.1272927 |
| Q26無駄遣い | 0.0780097 | -0.2121282 | 0.2667936 | -0.0149716 | -0.421225 | 0.0515053 |
| Q27新商品 | 0.2121196 | -0.0596531 | 0.4309869 | 0.0830721 | -0.1797376 | 0.0462985 |
| Q28コップでしか買わない商品 | 0.3818731 | -0.0139799 | 0.0518595 | 0.0712535 | 0.0438735 | 0.0683186 |
| Q29高くても健康 | 0.7100942 | 0.0974914 | 0.1466559 | -0.1320333 | 0.0088925 | 0.0021707 |
| Q30産地レシビ開心 | 0.5397713 | 0.1537658 | 0.230683 | 0.0721101 | 0.0494689 | -0.0635689 |
| Q31にぎやかな所が | 0.0622048 | 0.0035026 | 0.5018811 | 0.0635499 | -0.0247638 | -0.0184657 |
| Q32さちょうめん | 0.2001072 | 0.0943064 | 0.103723 | 0.0435794 | 0.3645636 | 0.0496431 |
| Q33気分が変わりやすい | 0.0396625 | -0.2796193 | 0.0775764 | 0.0435094 | -0.0142338 | 0.0936903 |
| Q34新しい体験 | 0.122036 | 0.1159401 | 0.555494 | -0.0130691 | 0.0288857 | 0.0735427 |
| Q35友人と買い物 | 0.0668866 | 0.1074661 | 0.3374588 | 0.0731719 | 0.0204875 | 0.0056362 |

因子抽出法: 主因子法
回転法: Kaiserの正規化を伴うハリ
マックス法

因子分析の結果より各顧客に対して、6つの因子の得点を付与することができる。ここでは、アンケートの各質問項目に対して、因子の絶対値が最大値をもつライフスタイルカテゴリを、その質問項目が所属するライフスタイルカテゴリと設定した。そして、各ライフスタイルカテゴリに所属するアンケート回答の平均値を、各顧客の消費・生活因子得点として定義する。その一例を図 2. 2. 1-5 に示す。ただし Q25、Q26、Q33 に関しては負の因子であるため、アンケート回答の重みを逆にして結果に反映させている。顧客 A は第 1 因子のこだわり消費派の傾向が強いことが、図中の顧客 B は第 4 因子の節約消費派の傾向が強いことがわかる。また、表 2. 2. 1-2 に各ライフスタイルカテゴリに対して最大の得点をもつ顧客の人数と、そのライフスタイルカテゴリに属する質問項目の平均値を示す。最大得点が複数のカテゴリで同得点の場合は、複数のカテゴリへ人数を加算している。また、これにより、図 2. 2. 1-6 に示すような顧客プロフィールの作成が可能となった。

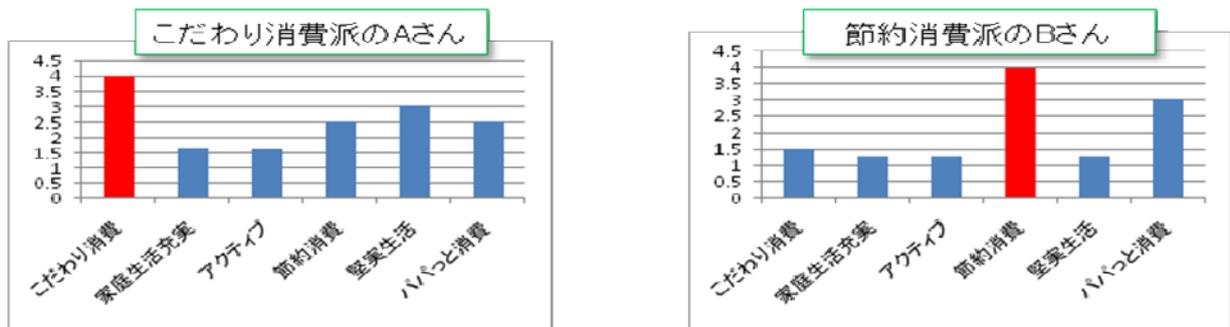


図 2. 2. 1-5 各顧客の消費・生活因子得点の例

表 2.2.1-2 各ライフスタイルカテゴリの特徴

| | 所属人数 | 平均得点 | 男女比 | 平均年齢 | 平均家族人数 |
|---------|------|------|-----|------|--------|
| こだわり消費派 | 1670 | 3.16 | 94% | 57歳 | 2.7人 |
| 家庭生活充実派 | 1385 | 3.12 | 93% | 56歳 | 2.8人 |
| アクティブ派 | 384 | 2.63 | 95% | 53歳 | 2.9人 |
| 節約消費派 | 707 | 2.50 | 95% | 54歳 | 3.0人 |
| 堅実生活派 | 364 | 2.52 | 85% | 58歳 | 2.6人 |
| パパッと消費派 | 869 | 2.58 | 96% | 49歳 | 3.3人 |

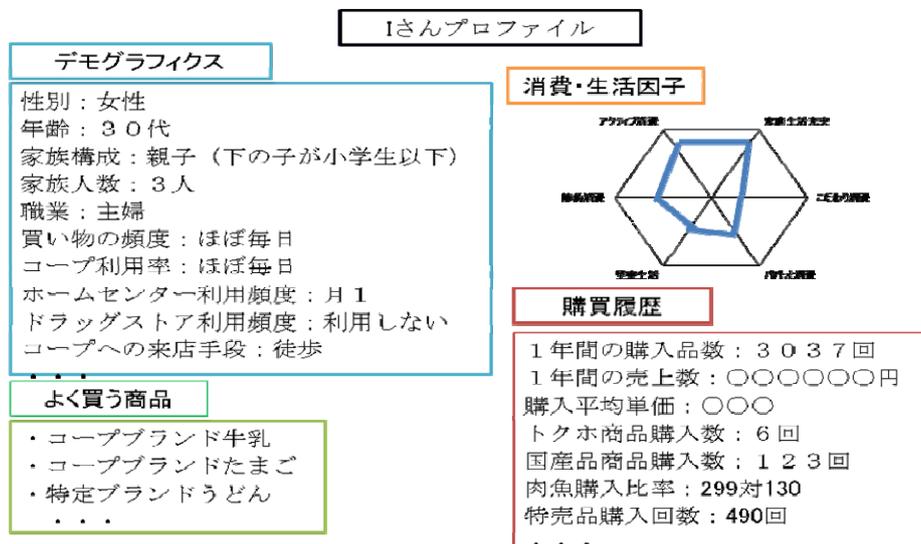


図 2.2.1-6 顧客プロフィールの一例

(2-2) 確率的潜在意味解析による顧客－商品の同時分類

ID-POS データから、どのような顧客がどのような商品を購入しているのか定量的に把握するための顧客－商品の同時分類を行った。ここでは、顧客が商品に対して求める意味カテゴリを潜在クラスとして仮定し、顧客と商品を同時クラスタリングする。この問題は潜在意味解析法 (PLSI) と同様のアルゴリズムで計算が可能となる。PLSI は自然言語処理の分野で文書と単語行列から文書の意味を推定するために提唱された潜在クラス分析の一種である。その概念図を図 2.2.1-7 に示す。この手法では、顧客と商品の購買履歴から顧客－商品の購買共起行列から、顧客がある意味カテゴリに属する確率と商品がある潜在意味カテゴリに属する確率を同時に計算し共にクラスタリングすることができる。また、潜在クラスの数は情報量規準により決定することができる。

ここでは、対象顧客をアンケート回答者 3891 人とし、対象商品を 1 年間の売上個数の上位の 1000 商品として同時分類を実行した。この 3891 人が主に購買をする 15 店舗での購買履歴をデータベースから抽出し、3981×1000 の顧客－商品購買共起行列を作成した。この該当トランザクション数は 4175441 件である。また、共起行列の充填率は 25.6%であった。PLSI の確率推定結果は初期値による影響を受けるため、[2, 4, 8, 10, 15, 20, 25, 30, 35, 40, 45, 50] の潜在カテゴリ数で情報量規準の一つである BIC の値の変化を調べ、その後 15 から 25 までの 11 潜在カテゴリ数に対し