

Visual Mining Studio を利用したレコメンドロジック

～ パーソナライズにより大幅なCVRアップを実現 ～

山川義介 菅由紀子(株式会社 ALBERT)

1. はじめに

BtoC 向けの EC(電子商取引)の市場規模は 1998 年に 650 億円程度でしたが、2002 年は 1 兆円を超え、2010 年には 7.7 兆円になる見込みです。EC 市場はまだまだ伸び続け、2014 年には 12 兆円に到達すると予測されています[1]。この EC を支えるテクノロジーとしてなくてはならないのが、アマゾンで有名になった、「この本を買った人はこの本も買っています」というレコメンド機能です。ここで用いられるレコメンドロジックは、アソシエーション分析が基本であり、消費者1人ひとりに最適な商品をレコメンドするパーソナライゼーションは真の One to one マーケティングを成功に導く鍵となっています。

レコメンドーションとは、対象者にとって価値があると思われるコンテンツ(商品や情報)をより個別的に提示することです。必ずしもインターネットの世界だけではなく、初めて店頭を訪れた顧客に「今日はサンマが旨いよ!」という行為も、マクドナルドで「一緒にポテトもいかがですか?」という店頭でのお薦めもレコメンドーションのひとつと考えられます。

一方、レコメンドエンジンというと、主にインターネット上で用いられるレコメンドーションを指す場合が多く、一般的には対象者の何らかのアクションに対して、判断し予測するステップを経てコンテンツをリアルタイムに提示するものが主流です。新商品の案内や単純な売れ筋ランキングの提示もレコメンドーションの一種ですが、これだけでは「レコメンドエンジン」とは言えません。

レコメンドーションは、能動的にユーザーにアクションを求める場合もありますが、完全な受動でも完全な能動でもなく、「ユーザー(またはそれに紐づく属性情報)とのインタラクション」によって推薦されるものもあります。また、店頭の売れ筋情報など、ユーザーに何ら能動的なアクションを求めない場合もレコメンドーションのひとつと考えられます。

この様に考えると、Google やヤフーなどの検索エンジンも、ユーザーが能動的に知りたいワードを入力すると、そのアクションに対してユーザーが知りたい情報である価値があるコンテンツ(情報)を提示するという一種のレコメンドエンジンといえます。

また、検索ワードに関連した広告が出るリスティング広告や、行動ターゲティング広告なども、対象者にとって価値があると思われるコンテンツをより個別的に提示する広告であることから、レコメンドーションのひとつと考えられます。

レコメンドーションは、パーソナライズ(個別化)されている場合もありますが、されていない場合もあります。パーソナライズされているレコメンドーションのほうが高機能・高精度であり、よりマッチングがよいため、高いコンバージョンが期待できます。従って、レコメンドロジックの開発においては、単純なアソシエーション分析だけではなく、いかにパーソナライズをするかを十分に考慮する必要があるわけです。

2. アソシエーション分析

アソシエーション分析とは、データマイニングの手法のひとつで、ある対象者が商品閲覧または購入したデータと、対象者以外がチェックまたは購入したデータを用い、その購入パターンから商品間の同時閲覧ま

たは同時購入などの相関分析をする手法です。このアソシエーション分析を基本としたレコメンデーションをALBERTの開発したACKマトリクス[2]では「アクションアソシエーション型」と呼び、どんな対象者であっても同じ商品をチェックまたは購入すれば同じ商品が提示されるという点で、個人の過去の購入履歴を全て学習しパーソナライズされている『協調フィルタリング』等の「履歴アソシエーション型」とは一線を画しています。クロスセルを促す場合、「アクションアソシエーション型」は同じ商品カテゴリ内で複数購入される商品には適していますが、「この洗濯機を買った人はこの洗濯機も買っています」に意味がないように、耐久消費財には向いていません。耐久消費財の場合は、購入検討時に比較した同時検討データを用い、この洗濯機を検討している人は、この洗濯機にも興味があります」という使い方をします。

一方、代表的なレコメンドロジックとして有名な『協調フィルタリング』は「履歴アソシエーション型」のひとつであり、ある対象者が商品をチェックまたは購入したデータと、対象者以外がチェックまたは購入したデータを用い、その購入パターンから人同士の類似性、または商品間の共起性をアソシエーション分析で導出し、対象者個人の行動履歴を関連づけることでパーソナライズされた商品を提示する手法です。

「アクションアソシエーション型」と似ていますが、対象者個人の履歴データを用いてパーソナライズしているところに違いがあります。『協調フィルタリング』は、同じような好みや購入履歴を持っている人は同じようなものが好きであるという仮説に基づいており、対象者の行動履歴と似たような行動履歴を持つ人々を抽出し、その人々が興味を持っている商品や買っている商品で対象者がまだ出会っていないと考えられる商品を提示する方法です。似た人々を定義し、かつ彼らが購入しているが対象者が購入していないものを提示するところに特徴があります。

しかし、商品数が非常に多い場合では対象者の購入履歴と似た購入履歴を持つ顧客を探すことが困難であるため、大規模な購入履歴データがあることが前提となること、また、まだ他の誰も使っていない新商品を推奨することができないこと、同じカテゴリの中で繰り返し購入されない商品では適用しづらいこと、他人のために何かを購入した場合であってもそれが対象者の嗜好を反映したものとして扱われてしまうことなどが短所として知られています。

		レコメンドに用いる対象者データ				
		対象者を特定しないデータ	対象者を特定するデータ			
			アクションデータ	履歴データ	申告データ	
		対象者全てに同様なレコメンデーション	パーソナライズされたレコメンデーション			
レコメンドに用いる対象者以外の情報	モノ属性ベース (Attribute)	商品属性、情報データベース	A-1 スペックランキング型	A-2 商品関連性評価型	A-3 コンテンツベースフィルタリング型	A-4 ニーズインプット型
	人ベース (Consumer)	対象者以外の履歴データ	C-1 人気ランキング型	C-2 アクションアソシエーション型	C-3 履歴アソシエーション型 (協調フィルタリング)	C-4 アンケートベース型
	ナレッジベース (Knowledge)	ナレッジの (信頼性・妥当性の担保が必要)	K-1 ナレッジベース 独断型	K-2 ナレッジベース 反応型	K-3 ナレッジベース 観察型	K-4 ナレッジベース 診断型

図 1. ACK マトリクス[2]

3. 共起性とレコメンドロジック

共起性は2つの集合の類似度として考えることができます。集合の類似度は要素の一致度で表すことができますが、一口に類似度と言っても様々な尺度があります。最も基本となる尺度は一致している要素の数、すなわち共起数ですが、そのほかにも以下のような係数が提案されています。

$$\begin{aligned}\text{共起頻度} &= |X \cap Y| \\ \text{Jaccard 係数} &= |X \cap Y| / |X \cup Y| \\ \text{Simpson 係数} &= |X \cap Y| / \min(|X|, |Y|) \\ \text{Cosine 係数} &= |X \cap Y| / \sqrt{|X| |Y|}\end{aligned}$$

これらの共起性には方向性がなく、レコメンドロジックはパンを買った人がその後バターを買う確率というように、方向性がある場合が多いので、レコメンドロジックとしては以下の様な指標が用いられます。

$$\begin{aligned}\text{Confidence} &= |X \cap Y| / |X| \\ \text{Support} &= |X| / |A| \quad (\text{Support} = |X \cap Y| / |A| \text{ とすることもある}) \\ \text{Lift} &= (|X \cap Y| / |X|) / (|Y| / |A|)\end{aligned}$$

Confidence は信頼度、Support は支持度、Lift はルールの改善度などと訳されます。Confidence とは、ルールの条件 X が発生したときに、結論 Y が起こる割合を示します。Confidence が高いほど、ルールの条件と結論の結び付きが強いことを意味しています。

Support とは、条件と結論を同時に満たすトランザクション $|X \cap Y|$ が全トランザクション $|A|$ に占める割合をいいます。つまり、ルールそのものの出現率です。好ましいレコメンドルールは、Confidence が高いことが重要なのは当然ですが、Support も一定の大きさが必要です。Support が極端に小さいということは、そのルールがめったに起こらない商品の組み合わせということを意味するので、あまりよいルールとはいえません。

一方、Lift はルールの改善度とかルールの意味の大きさといわれますが、そもそもよく売れている商品は、どんな商品とも共起してしまいます。有名で最もよく売れるパンがあったとします。このパンはお店を訪れた人の 70% の人が購入すると仮定しましょう。このお店の北海道バターを買った人の 60% がこのパンを買ったとすると、Confidence は 60% ということになり、なかなかよいルールだなど思ってしまうかもしれません。しかし、そもそも何の条件もない場合は 70% の人がこのパンを買っているのに、北海道バターを買ったという条件がつくと 60% の人しかパンを買わないとすれば、北海道バターを買った人にパンをお薦めするのはあまりよいルールではなく、ルールが改善されたとはいえないわけです。

このように、ある条件下でどの程度ルールが改善されるかを表す指標が Lift であり、そもそも何の条件もない時に買われる確率 $(|Y| / |A|)$ に対して、X を購入したという条件下で買われる確率 $(|X \cap Y| / |X|)$ の比を評価したものです。

レコメンドエンジンに用いられるレコメンドロジックは、これらの指標を組みあせて設計されており、ALBERT ではさらにランキングデータなどを加味した独自の総合指標を用いています。



図 2. ALBERT のレコメンドロジック

4. パーソナライズの必要性

分析で得られたレコメンドルールを用い、その商品を購入しようとしている顧客に対し「こちらはいかがですか？」と商品を推薦することは、もちろん有効です。しかし、推薦する商品がその顧客がこれまでに全く購入したことのないカテゴリの商品であった場合はどうでしょうか？

たとえば、生成されたアソシエーションルールでは、「白い靴下」という商品に対して「ストッキング 3 足組」が共起しているけれど、その顧客はその店舗で「ストッキング 3 足組」を一回も購入しておらず、T シャツばかりを買っているという場合です。この顧客には「ストッキング 3 足組」を推薦するよりも、生成されたルールの中からこれまで購入実績のあるカテゴリや、買ったことはないが買う可能性の高いカテゴリの商品ルールを優先して推薦したほうが効果的です。

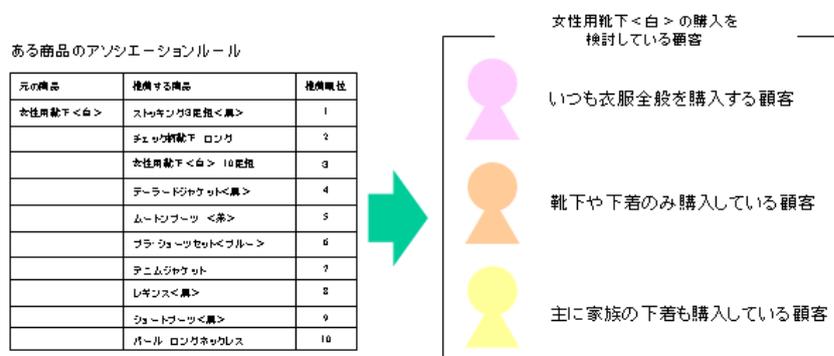


図 3. 生成されたアソシエーションルールと推薦する対象者

アソシエーション分析でよく例に挙げられるのが、「おむつとビール」の関係です。アソシエーションルールの分析では、商品 ID (SKU) 単位で共起関係を測りますが、すべての商品の相関関係を明らかにするには大量のデータが必要となり、困難です。多くの商品を扱う場合は特に商品の中には購入されたことのない商品も存在します。そこで用いるのがカテゴリの情報です。カテゴリは普遍的であるため予測が可能です。商品 ID 単位の相関よりカテゴリ単位の相関のほうがはるかにパワフルで精緻な購買予測が可能となります。カテゴリ間同士にどのような相関があるか、という情報は商品間のルール以上に重要なルールとなります。

5. 顧客セグメントの手法

顧客をセグメントするためには多くの手法がありますが、用いるデータや業種・業界などによってその手法を使い分けることが重要です。今回のように、顧客の購買履歴データを用いる場合には、デシル分析、RFM 分析、クラスター分析などの手法が代表的です。購入金額や購入商品数、購入日や購入頻度などを数値化し、顧客をセグメントしていきます。

デシル分析は、顧客を 10 のグループに等分に分ける分析手法です。購入金額や購入商品数などの合計値が高い順に 10 等分し、グループごとにどのような顧客コミュニケーションを行なうかを検討します。非常に手軽に出来る分析手法ですが、あくまで現状把握と簡単な施策が出来るにすぎません。

RFM 分析は、特に小売業などでよく活用される分析手法で、Recency (最終購入日) Frequency (頻度) Monetary (購入金額) の 3 つの指標で顧客を分類します。それぞれの指標を何段階かに分け、 $R \times F$ 、 $R \times M$ といったようにクロスさせて顧客を分析します。

たとえば、 $R \times F$ のクロス結果で「購入頻度は多いが直近の来店がずいぶん前である」という顧客には来店を促すメール配信を行なう、といったようにセグメント別に顧客コミュニケーションの課題と施策を明確化することができます。

ただし RFM 分析は「ある瞬間の顧客分析」であり、時期が変われば顧客も入れ替わり継続性がありません。また、金額のみで優良顧客かどうかを分類しているので、「何を買ったか」という点について顧客の特性を把握しているとはいえません。

クラスター分析は、個体をいくつかのグループに分類する手法で、大別すると「階層型」と「非階層型」の 2 つがあります。階層型は分類する対象が少なく、結合過程にも意味がある場合に用いられます。非階層型は分類する対象数が多く、結合過程があまり重要でない場合に用いられます。

分析を行なうためには、統計解析ソフトや類似度計算を行なうプログラム等が必要となりますが、似た特徴を持つ顧客を的確に分けることが可能です。ただし、分類した顧客グループがどのような特徴を持っているか、といったことや、それぞれのグループに対してどのようなコミュニケーション戦略をとるべきか、ということは独自に考案する必要があります。

6. クラスタ分析を活用した顧客ごとのパーソナライズルール

顧客がどのカテゴリの商品を購入したかというデータを用いたクラスタ分析の実例をご紹介します。

用いたデータ

- ・某大手インターネット通販サイトの購入履歴データ
- ・約 400 万行、9 万商品 131 カテゴリ
- ・顧客数:約 6 万

顧客ID	カテゴリ	カテゴリ	カテゴリ	カテゴリ
1	0	0	1	2
2	2	4		1
3	4	2	4	2
4	2	4	1	2
5	2	2	0	2
6	2	2	0	0
7	3	2	0	0
8	2	2	0	0
9	2	2	4	0
10	2	2	0	0
11	2	2	0	0

計算方法

- ・クラスタ分析 k-means 法
- ・距離計算法: cosine
- ・繰り返し最大数: 100

顧客×購入したカテゴリの商品数のデータから、2種類以上のカテゴリの商品を購入している顧客をk-means法でクラスタ分析

計算の結果得られたクラスタごとに、購入しているカテゴリとの関連を見ると、下図のような特徴が現れました。

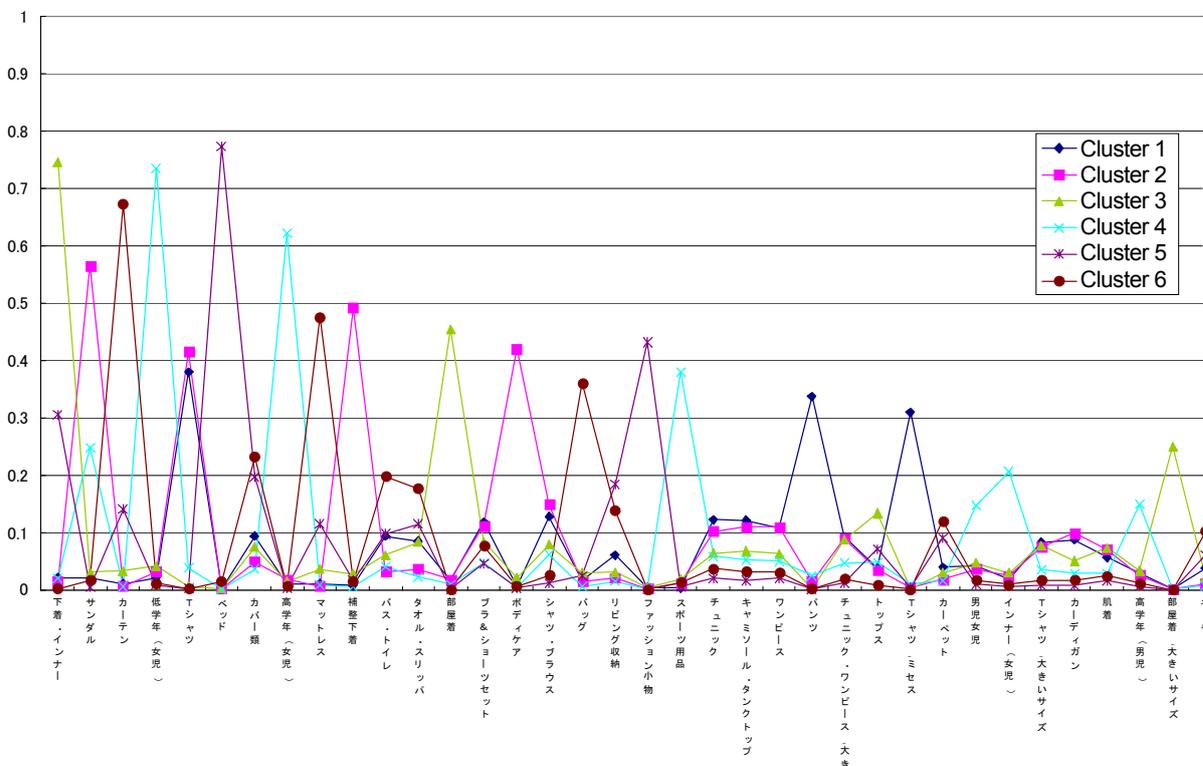


図 4. クラスタ別購入カテゴリ

クラスター1 はTシャツやパンツを多く買っています。男性ものや子供用品などをあまり購入していないことから、比較的若い女性が多いクラスターだと思われます。クラスター2 は補正下着やボディケア商品などを購入しており、靴やサンダル、服のほか子供服も時々購入しているので、子供がいるけれど働いていて外見に気を使っている女性のクラスターではないかと思われます。

クラスター3 はちょっと太めの奥様で、大きめサイズの部屋着、インナーなどのほか、最も購入数が多いのがメンズの下着、インナーとなっており、ご主人用に多く購入している層だと思われます。

クラスター4 は、男児女児用の商品全般を多く購入しており、ご自身のサンダルや服なども時々購入している、ヤングミセス層です。クラスター5 は、ベッドやメンズ小物やデスク、家具などを多く購入しており、比較的若い夫婦や男性層がここに属すると思われます。

最後のクラスター6 はカーテン、マットレス、カーペット、カバー類、バッグが多いことが特徴で、あまりアパレル関係を購入しない層です。服はあまり通販で購入せず、実用品は通販でという比較的若い女性を中心にしているクラスターと言えるでしょう。このクラスターに属する顧客には、カーテン、バッグ、ミセスのTシャツを優先して推薦し、それ以外のカテゴリについては優先度を下げることで、顧客ごとに効果的なレコメンドを行なうことが可能です。

1 種類のカテゴリしか購入していない顧客については、購入したカテゴリと共起性の高いカテゴリの人気商品を推薦します。顧客の購入履歴をカテゴリデータ単位でアソシエーション分析し、カテゴリごとに共起性の高いカテゴリを抽出しておくことで、顧客の好みに合う商品が推薦されやすくなります。



図 5. クラスタープロフィールイメージ

参考文献

- [1] 株式会社野村総合研究所:NEWS RELEASE (2009.12.21)
- [2] ALBERT ホームページ:<http://www.albert2005.co.jp/technology/ACK.html>