

ネットワーク流量推定手法および文字列解析手法を応用したブランド購買パターンからの知識発見

Knowledge Discovery in Brand Purchasing Pattern using Network Flow Estimation and String Pattern Analysis

大阪産業大学 *羽室 行信 HAMURO Yukinobu
01104684 京都大学 加藤 直樹 KATO Naoki

1. はじめに

企業のマーケティング活動において、自社ブランドに対する顧客のロイヤルティを高めることは、安定した売上げと利益をもたらす、さらにはマーケティングコストの引き下げに貢献すると言われており[1]、非常に重要なテーマと認識されている。一方で近年、小売店においては顧客の詳細な購買履歴の蓄積が可能となっており、企業はこれらのデータを活用し、1)全体としてのブランドスイッチの傾向を的確に把握し、さらには、2)ロイヤルティを示す顧客の購買特徴を把握することによって、顧客のブランドロイヤルティを高めるためのマーケティングアクションの有効なサポートを可能にすることが期待できる。そこで本稿はこれら2点について、ネットワーク流量推定手法および文字列解析手法をブランド購買パターン分析に応用する方法について提案し、その有効性を検証することを目的としている。

まず全体としてのブランドスイッチの傾向を把握するための手法として、連続する2つの期における顧客のブランド購買数量をノードとし、またブランドスイッチ確率をエッジとするネットワーク流を考え、平均的な顧客のブランドスイッチ確率を求めることによりブランドスイッチの傾向を把握する手法を提案する。次に、ブランドの購買パターンを文字列として捉え、分子生物学や遺伝子解析の分野で研究が盛んな文字列解析手法を利用し、あるブランドにロイヤルになる顧客のブランド購買ルールの発見手法を提案する。

なお本稿では、あるスーパーマーケットから提供されたID付きPOSデータ¹を用い、粉末洗濯洗剤に焦点をあてて分析を進めていく。

2. ネットワーク流量の推定手法²の応用

ある商品カテゴリにおけるブランド集合を $B = \{1, 2, \dots, m\}$ とする。ある期 t において顧客 i が購買したブランド $j \in B$ の購買数量を $a_i^t(j)$ とする。ここで連続する二つの期 t_1, t_2 における顧客 i の全てのブランドの購買数量が与えられたとき、顧客 i の期 t_1 から t_2 におけるブランド j から k へのスイッチ確率 $p_i(j, k)$ を推定する問題を考える。

説明の簡略のために顧客 i が期 t_1, t_2 に購買した総個数 $a_i^{t_1}$ 、

$a_i^{t_2}$ は等しいと考え³、図1(a)に例示されるように、期とブランドのペアで表されるノード(ノード内の数字は購買数量を表す)と、合い続く期のノード間に枝集合を持つ有向グラフを考えると、各ノードの流量が既知のときの各枝を流れる流量 $f_i(j, k)$ を推定し、その流量の割合をブランドスイッチ確率として考える。

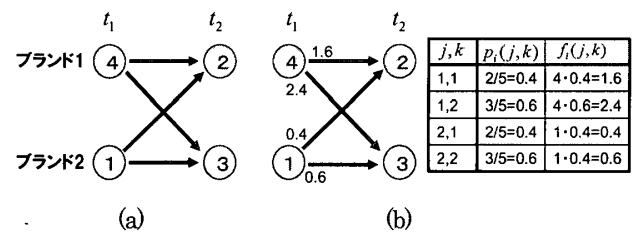


図1 ネットワーク流量の推定問題

この時ブランドスイッチ確率は t_1 における購買個数に左右されず $p_i(j, k) = a_i^{t_2}(k) / a_i^{t_2}$ と推定するのが妥当で、また流量 $f_i(j, k)$ は $a_i^{t_1}(j) \cdot p_i(j, k)$ で求められる(図1(b)に例示)。

そして、このようにして求めた全顧客のブランドスイッチ確率から、平均的な顧客の j から k へのブランドスイッチ確率 $p(j, k)$ を次の問題の解として定義する。

$$\text{minimize } \sum_i a_i^{t_1}(j) (p_i(j, k) - p(j, k))^2 \quad (1)$$

式(1)より $p(j, k)$ は式(2)で求められる。

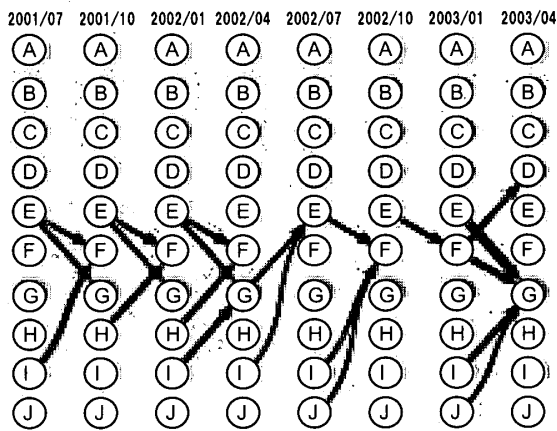
$$p(j, k) = \frac{\sum_i a_i^{t_1}(j) p_i(j, k)}{\sum_i a_i^{t_1}(j)} \quad (2)$$

以上の考え方を粉末洗濯洗剤の10個のブランド(A~J)に対して適用した結果を図2に示す。ここでは四半期を1期とし、2001年7月期から2003年4月期までの8期におけるブランドスイッチ確率を矢印の太さとして表している。結果を考察すると、多くの期においてブランドEからFもしくはGへのブランドスイッチ確率が高いことがわかり、F、GはEの代替可能な特性を有していると推測され、かつEブランドに何らかの問題があると考えられる。逆にA,B,C,Dからのスイッチ確率の高いブランドはなく、これらのブランドロイヤルティが比較的高いことがわかる。

¹日本マーケティングサイエンス学会ID付きPOSデータ活用研究部会平成15年度データ解析コンペティションより提供されたデータを用いた。

² [3]の研究に基づき発展させたものである。

³ $a_i^{t_1} \neq a_i^{t_2}$ の場合は、他店での購買を仮定することによって解決している。



同一ブランド間のスイッチ確率は省略している。
スイッチ確率が0.18以上のエッジのみ表示している。

図2 洗濯洗剤10ブランドについてのブランドスイッチ

3. 文字列解析手法の応用

ブランドをアルファベット1文字で表すと、ある顧客のブランド購買パターンは文字列で表現できる。例えば、文字列「AAACBBBB」は、Aブランドを3回連続購買した後にCを購入し、その後Bを4回連続購買したことを示す。

このような文字列で表現されるデータからの知識発見については、分子生物学や遺伝子解析の分野において様々な手法が開発されており、例えば BONSAI は文字列で表現された複数のケースを含む正事例と負事例について、文字列の正規パターンをノードの条件にもつ決定木による分類モデルを構築する[4]。

我々はBONSAIをビジネスデータに応用するために機能を拡張したE-BONSAIを開発してきた[2]。オリジナルのBONSAIからの主な改良点は次の通りである。1) 文字列属性だけでなく、数値属性やカテゴリ属性も説明属性として同時に利用可能とした。2) 正規パターンに文字列の出現位置情報を加えた。3) 複数の文字列属性を扱い可能とした。次に、このE-BONSAIをブランドスイッチ分析にどのように応用したかについて説明する。

顧客のブランド購買パターンにおいて、ある同一ブランド $g \in B$ を n 回(実験においては $n=3$)以上連続して購買した期間をブランドロイヤル期間 t 、そのブランドをロイヤルブランドとする(ただし期間 t の一回前の購買ブランドは g とは異なる)。また期間 t の前 m 回(実験では $m=5$)の購買期間 s におけるブランド購買パターン集合(文字列集合)を $BPP(g)$ とする。このとき、異なる二つのロイヤルブランド $j, k \in B$ に関する $BPP(j)$ 、 $BPP(k)$ をそれぞれ正事例、負事例としてE-BONSAIを適用し分類モデルを構築する。このことにより異なる二つのブランド j, k について、ブランドロイヤルティ形成期間の前にどのような購買行動の違いがあるかについてのルールを発見することが可能となる。

さらに、ブランド購買パターン(BPP)以外にも、期間 s における購買価格帯パターン(PP)、期間 t における購買ブランド j, k の購買価格帯平均(SP)、そして顧客属性として年齢(AGE)、

性別(GEN)を説明属性として利用した。なお価格帯とはブランドごとに、販売価格をできるかぎりトランザクション数が均等になるように0,1,2の3段階に分割したものである。以上、5つの説明属性を含むデータセットのサンプルを表1に示す。顧客IDが1についてみると、期間 t の連続購買ブランドはFで、その時の購買価格帯平均は0.76と比較的安い金額で購入している。また期間 s においてAを3回連続購買した後、Fを1回購買し、そして再びAを購入しており(BPP)、いずれの購買においても最も安い価格帯で購入している(PP)。

表1 データセットサンプル

顧客	ロイヤルブランド	BPP	PP	SP	GEN	AGE
1	F	AAAF	00000	0.76	女	45
2	F	GFDFD	21021	1.5	男	63
3	D	EDDEA	12121	1.3	女	28
4	D	CDDEC	21001	0.56	女	34

以上に示した手法をブランドDとFについて適用した結果が図3に示されている。この結果より、ブランドDにロイヤルになる多くの顧客は、それ以前の期間 s においてDを2回以上連続購買していることがわかる(リーフ①)。またブランドFにロイヤルになる多くの顧客は、Fを比較的高い価格帯で購入し、かつ期間 s においてDを2回連続購買したことがないことがわかる(リーフ②)。より詳細な解釈については当日の報告にて行う。

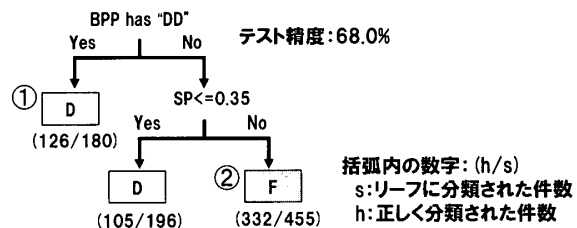


図3 ブランドDとFに関するルール

4. むすび

本稿では、ブランド購買パターンの分析について、ネットワーク流量推定手法および文字列解析手法を応用することによって、実用的なルールの発見が可能であることを示した。

参考文献

- [1] Aaker, D. A. *Building Strong Brand*, The Free Press, 1996.
- [2] Hamuro, Y, Katoh, N, Yada, K, "A Machine Learning Algorithm for Analyzing String Patterns Helps to Discover Simple and Interpretable Business Rules from Purchase History", *Lecture Notes in Artificial Intelligence*, Vol.2281, Springer, 2001, pp. 565-575.
- [3] Katoh, N., Hamuro, Y., Yada, K., "Discovering Purchase Association among Brands from Purchase History", *Proceedings of SSGRR 2002w*, 2002, in CD-ROM.
- [4] Shimozono, S., Shinohara, A., Shinohara, T., Miyano, S., Kuhara S. and Arikawa, S., 1994, Knowledge Acquisition from Amino Acid Sequences by Machine Learning System BONSAI, *Trans. Information Processing Society of Japan*, Vol. 35, pp.2009-2018.