

顧客個人の購買行動データに基づく

バラエティ・シーキング行動予測手法の提案

Prediction of Variety-Seeking Behavior Based on Individual Customer

Purchase Data

浦 杏弥¹ 後藤 裕介¹

Amane Ura¹ and Yusuke Goto¹

¹岩手県立大学

¹Iwate Prefectural University

Abstract: In this study, we tried to predict the brand switching behavior (variety-seeking behavior) for stimulating customer's purchases at the store in real-time. We used customer purchase data from Japanese retail stores for evaluating our proposed method. Our proposed method includes some additional explanatory variables from existing studies, which considers purchasing pattern in a time series. The results showed that our proposed method increased the prediction accuracy of brand switching behavior. Our proposed method includes some additional explanatory variables from existing studies, which considers purchasing pattern in a time series. The results showed that our proposed method increased the prediction accuracy of brand switching behavior. Furthermore, our proposed method can be used for various stores and products because experimental results showed that our proposed method's effectiveness is independent of stores and products.

1 はじめに

近年、小売店ではオンラインショップが普及され、実店舗では競争が激しく、新規顧客の獲得や、既存顧客の売り上げ向上が難しい。現在行われている、顧客に購買を促す取り組みとして、店舗内において商品プロモーションやクーポンの配信等の働きかけをしている[1][2]。これらの取り組みに関する問題点として必ずしも顧客の関心のあるタイミングや好みで広告を行っておらず、顧客の買い物時に購買意欲を高められていないことが挙げられる。

現状では会計後にクーポンを発行しているが、これでは当日の購買を促すことができないことや、使用期限などと言った理由から使用できない場合もあると考えられる。そこで来店時にクーポンを発行するよう体制を変えることで、当日の購買を促すことができることや、期限を気にせず、今顧客が欲しいと思う新しい商品の購買を促すことができると考察した。

ここでより多くの人にクーポンを反応してもらうために、クーポンを配布することで購買意欲が向上する可能性が高い顧客について検討を行った。その

結果バラエティ・シーキング型の顧客である可能性が高いことが考えられた。バラエティ・シーキングとは何かものを選ぶ際に、特定のブランドだけではなく、様々なブランドを購入しようとする消費者の行動特性のことである。バラエティ・シーキング型の顧客はブランドスイッチを行う。ブランドスイッチとは顧客が今まである会社の製品を使っていたものの、あるきっかけを元に競合他社の同じ製品に乗り換えてしまうことを指す。

以上より顧客の過去の購買データからブランドスイッチを行うタイミングを予測することで来店前に各顧客にあったクーポンの配布を適切に行うことができ、顧客の購買意欲の向上につながる可能性を見出した。

先行研究[3]では、直前の購買行動は考慮されているが、各顧客の全体の購買行動、商品の違い、店舗の違いは考慮されていない。また外国での店舗を対象としており、日本の小売店舗でも同様の結論となるかどうかはわからない。

本研究では、より正確性の高いリアルタイムプロモーションを提案するために、顧客の購買行動の全体と店舗や商品の違いを考慮したバラエティ・シー

キング行動予測手法の提案を行い、その有効性を日本の小売店舗の購買データから評価する。第2章では先行研究について概説する。第3章では先行研究の比較や、商品ごとの購買の仕方についての基礎分析の結果を示す。第4章ではバラエティ・シーキング行動予測手法を示し、第5章では分析結果を示している。第6章は考察を示している。第7章はまとめである。

2 先行研究

先行研究[3]では顧客の多様性と、製品の特性がバラエティ・シーキング行動にどれだけ影響があるかが評価されている。これにより、購入頻度が高い顧客ほど、バラエティ・シーキングを行い、ブランドスイッチの頻度が増えることがわかっている。しかし結果として関連研究の予測精度はリピート購入が81%、ブランドスイッチが53%、全体として70%という結果になった。

先行研究[4]では店舗の購入データに基づいて、バラエティ・シーキング行動が起きやすい商品の分析や、実際にその商品に対する顧客のバラエティ・シーキング行動の違いを分析している。結果として、バラエティ・シーキング行動をする人はブランドの種類が多い商品で多く、商品の仕方に差が生まれることがわかった。

先行研究[5]では過去の購買データだけでなく、当日の製品との直接的な顧客のデータを考慮して購買行動の多様性を分析している。結果は女性客がリピーターになる可能性や、女性かつ若い顧客は、購買行動を変更して購買を増加させる可能性が高いことがわかった。

先行研究[6]では顧客行動の閲覧/購買した商品ジャンルの幅広さや特定のジャンルへの依存度を見ることによる、ある行動の「多様さ」を定量化した変数を用いて将来的な行動を予測分析している。結果は「多様性」を用いて分析することで精度は向上した。

しかしこれらは、日本の実店舗では考慮されておらず、日本の小売店舗でも対応するかについては記されていない。また、現在の予測方法では、ブランドスイッチの予測精度が低く、改善が必要と考えられる。精度向上のために、先行研究で考慮されていない顧客の購買行動を細かく見ることや、汎用性のために商品の違いや複数店舗からバラエティ・シーキングとの関連性を分析していく。

3 基礎分析

3.1 データの概要

本研究で使用するデータは2016年10月1日から2017年10月31日までの期間において、総合スーパー3店舗から取得された購買履歴である。購買履歴は、日、時間帯、JAN、会員ID、売単価、売上数量、商品名等の項目が記録された47,454,867行のデータである。

3.2 先行研究の比較

まず、本研究が先行研究[4]との傾向の違いについてみていく。今回、個々の顧客の購買行動について以下の累積分布関数の式で求めた。

$$Pr(K|\lambda) = e^{-\lambda} \sum_{i=0}^{\text{floor } k} \frac{\lambda^i}{i!} \cdots (1)$$

上の式は顧客が*k*個以下の製品を購入した確率を示す。商品はそれぞれ「牛乳」「トイレットペーパー」「食パン」の3点から分析を行っている。また顧客は商品を3点購買した人を対象としている。*K*は商品別顧客ごとのブランド数を示す。 λ は顧客別に購買しているブランドの種類の平均を示す。顧客別かつ商品別に、累積分布関数をし、平均をとっている。これは1.0に近いほどブランドを変更して購買している。結果は図1のようになった。

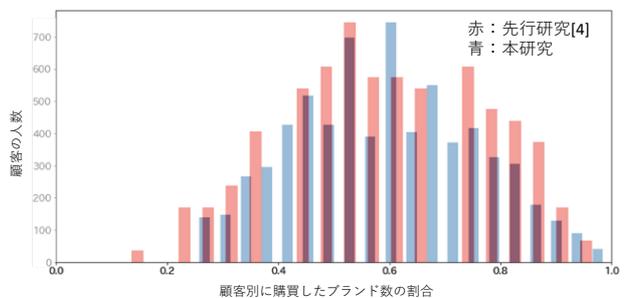


図1:顧客商品別の累積分布関数

先行研究のものは資料をもとに作成したものが、比較的購買した購買数の割合に対する顧客の人数の傾向は、本研究と似ている傾向であることがわかった。よって本研究でも先行研究のように顧客の購買行動に差がある可能性が見出された。

次はクラスタリングを行い、顧客の分類を行った。今回は3つの商品を購入した顧客のそれぞれブランドスイッチに関する平均と標準偏差(累積分布関数)の違いからクラスタリングを行った。クラスタリングはk-means法というクラスターの平均(means)を

用い、あらかじめ決められたクラスター数「k」個に分類する方法で行った。使用データは来店日数が11回以上ある顧客のブランドスイッチに関する平均と標準偏差を用いている。クラスター数はエルボー法を用いて4に設定した。以下の図2はクラスタリングを行った結果になっている。濃い青は、3商品それぞれの購買ブランド数が比較的少なく、ばらつきがあまりない顧客を示している。青は、3商品それぞれの購買ブランド数が比較的多く、ばらつきがあまりない顧客を示している。緑は、3商品それぞれの購買ブランド数が適度にあるが、3商品でばらつきがある顧客を示している。黄緑色は3商品それぞれの購買ブランド数が適度にあるが、ばらつきがあまりない顧客を示している。それぞれグラフ中のA~Dはクラスターごとに1人ずつ抜き出し、表1からA~Dのそれぞれの顧客の購買の仕方についてみていく。

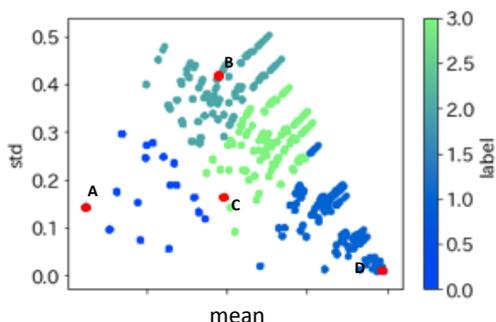


図2:累積分布関数のクラスタリング結果

まず図2よりクラスターごとのそれぞれの顧客について見ていくことで商品ごとの購買行動の違いを見ることが出来る。

表1:顧客ごとの累積分布関数の比較

	全体的な累積分布関数	牛乳	トイレットペーパー	食パン
顧客A	0.24(低)	0.21(低)	0.40(中)	0.12(低)
顧客B	0.57(中)	0.92(高)	0.67(中)	0.12(低)
顧客C	0.59(中)	0.66(中)	0.40(中)	0.70(中)
顧客D	0.99(高)	0.99(高)	0.98(高)	0.99(高)

表1より顧客C, Dの行動は3商品それぞれ累積分布関数の値がかなり一貫しており、3つのカテゴリ全てで同じレベルの多様性を求めている。そして、顧客Bは牛乳が0.92、トイレットペーパーが0.67、食パンが0.12と完全に累積分布関数の値が異なり、同じ振る舞いを示す単一のカテゴリはないことがわかった。以上より先行研究と類似した結果になり、本研究においても顧客ごとや商品ごとに購買の仕方に差が生まれることが分かった。

3.3 顧客別ブランド購買数

図3は商品ごとの購買回数が5回以上の顧客における購入ブランド数を示している。

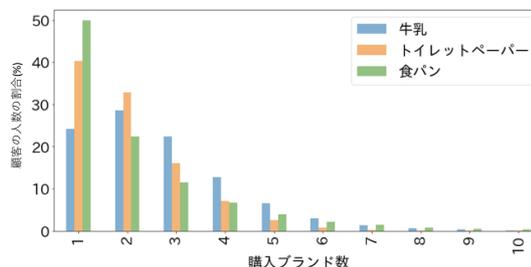


図3:顧客別各商品の購入ブランド数

図3より牛乳に関しては1つのブランドを購入し続けている顧客は25%ほどで、約75%の顧客が購入するブランドを変更している。それに対してトイレットペーパーは顧客の60%、食パンは顧客の50%と、商品によってその購買行動には差がある可能性が見出された。

3.4 商品とブランド継続の関係

継続率は顧客別にある商品の属するブランドごとの継続購買回数を表している。図4は顧客別にブランドごとの継続購買回数の最大値を求め、関係を見ていく。今回は商品によって大きな差が見られなかったので例としての牛乳を示している。

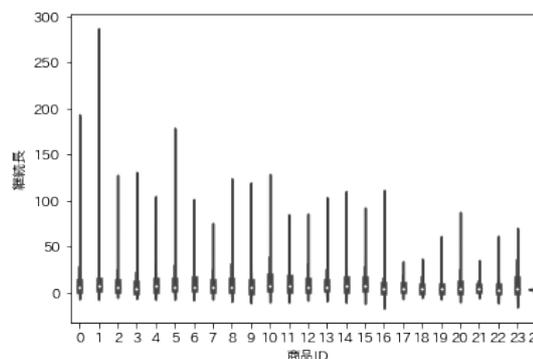


図4:商品とブランド継続の関係

図4より、ばらつきが多い商品は人気が高い商品であることがわかった。また様々な味の違うものは気まぐれに変更して購買している顧客がいることがわかった。

3.5 ブランド変更率の商品別比較

顧客別商品ごとにブランドを変更している人を以下の式を用いてみていく。

$$s = \frac{\text{Max}(len)}{a} \cdot \dots (2)$$

上の式 s は顧客別にブランドごとの継続購買回数 len の最大値を商品ごとの累計購買回数 a で除算し、ブランド変更率を求めた。1.0に近いほど同じブランドを購入し続けている。以下のグラフはこの式を基に作成している。

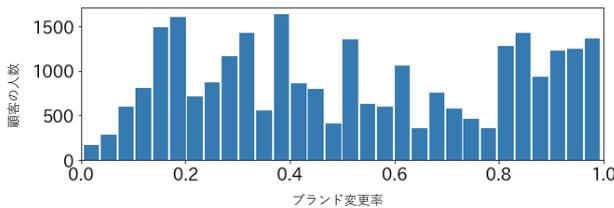


図 5:牛乳のブランド変更率の分布

図 5 より牛乳はばらつきがあるが、0.8~1.0 または約 0.2 に人数が集中していることから、お気に入りのブランドがある顧客とブランドに執着していない顧客で分かれている可能性がある。

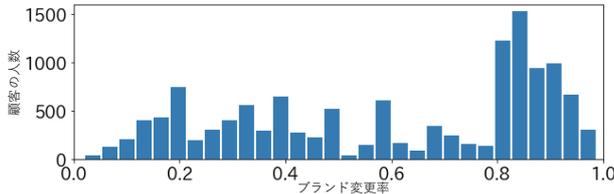


図 6:トイレットペーパーのブランド変更率の分布

図 6 よりトイレットペーパーはばらつきがあるが、約 0.8 に人数が集中していることから、お気に入りのブランドがある中でブランドを変更している顧客が多い可能性がある。

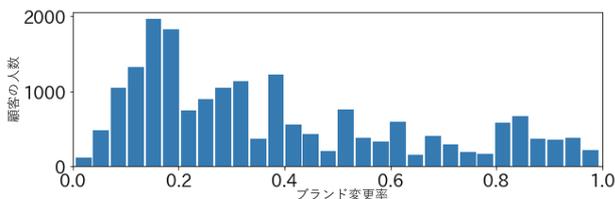


図 7:食パンのブランド変更率の分布

図 7 より食パンはばらつきがあるが、約 0.2 に顧客が集中していることやサイズの種類が豊富なことから、同じブランドでも気まぐれに味を変更するなどといったブランドスイッチが行われている可能

性がある。

以上より商品によって購買の仕方に差がある可能性が考察される。

3.6 商品別購買回数とブランド継続の関係

図 8 は購買回数とブランドを継続して購買した継続長の関係についてみていく。今回は商品によって大きな差が見られなかったので例としての牛乳を示している。

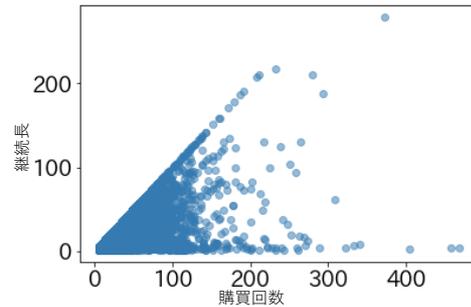


図 8:購買回数とブランド継続長の散布図

図 8 より購買回数が増加していくにつれ比較的ブランドを変更する顧客が少ない傾向にあるので、購買回数が多い人ほど自分の好みのブランドを見つけ、それを中心にブランドスイッチを行なっている可能性があると考えられる。

基礎分析の結果から、商品の違いがブランドスイッチ予測の精度に影響を与える可能性を考えた。そこで本研究では説明変数に商品を取り入れ、最終的にブランドスイッチにおける重要度を確認することとする。

4 バラエティ・シーキング行動予測 分析手法

4.1 予測分析手法の概要

4.3 の表 2 のとおり、顧客の過去の購買データを用いて、先行研究の「最終購買パターン」「購入ブランド数」「性別」に加えて、本研究の「 k 回目の購買パターン」「商品」「店舗番号」を説明変数として作成した。そしてこれらを用いて、顧客がブランドスイッチを起こす予測分析を行う。分析手法はランダムフォレスト分析を用いた。リピート購入（ある商品に対して同じブランドを購入し続けていること）を「0」、ブランドスイッチ（ある商品に対して違うブランドに変更すること）を行った場合を「1」として分類を行い、目的変数を求めた。

4.2 予測分析用のデータ加工

先行研究[3]を模倣したもので、1～(n-1)回を説明変数として利用し、n回目の購買パターン(リピート購入かブランドスイッチ)を目的変数として使用している。本研究では、商品別に11回以上購買を行った顧客を対象とし、顧客ごとに1~10、2~11...n~(n+9)といった10回組のものを作成し、説明変数とした。目的変数は(n+10)回目の購買パターンを使用している。

4.3 予測分析の説明変数

今回説明変数として使用した変数は表2の、(1)先行研究を模倣したもの、(2)本研究で追加した購買履歴の2点から分析を行った。

表2:バラエティ・シーキング説明変数

説明変数		内容
(1)	最終購買パターン	説明変数(1~n回)のうちn回目の購買行動パターン (リピート購入[0], ブランドスイッチ[1])
	購入ブランド数	顧客が商品ごとに購入したブランドの種類の数 [値域: 1~10]
	ダミー変数 性別	男, 女, 不明
(2)	k回目の購買パターン	説明変数(1~n回)のうちk回目の購買行動パターン (リピート購入[0], ブランドスイッチ[1]) k = 2~9
	ダミー変数 商品	牛乳, トイレットペーパー, 食パン
	ダミー変数 店舗番号	24, 49, 66

先行研究では1回分の購買パターンのみで分析しているが、本研究ではそれに加えて新しく、「k回目の購買パターン」で、説明変数全体の購買パターンをより細かく検討する。そうすることでパターンがより細かく分析することが可能となると考える。また、性別は総合スーパーにおいても関係があるのかを検討する。そして基礎分析より、商品によって購買行動に差がある可能性が高いことがわかった。以上より、商品や店舗間で差異があるか確かめ、本研究の予測分析が様々な場面で利用できるかを検討する。目的変数は「リピート購入」を[0], 「ブランドスイッチ」を[1]としている。

5 バラエティ・シーキング行動予測分析

5.1 予測分析の評価方法

今回は、(1)先行研究の模倣、(2)(1)と本研究で作成したものを合わせたものの2点から分析を行う。また(1), (2)は比較するため、同じ顧客を使用している。加えて、本研究では、より多くの場合に利用できる

ようにテストデータと訓練用データは顧客が重複しないように分けてある。

手法の評価に際しては、顧客のブランドスイッチするタイミングが正確にわかることでリアルタイムプロモーションが充実させることができると考えたので「recall」で各手法を比較している。また、ランダムフォレスト分析のハイパーパラメーターはグリッドサーチにより最適化を行った。

5.2 分析結果

ランダムフォレスト分析を行った結果、表3のような結果になった。

表3:それぞれの精度比較

説明変数	クラス	予測0	予測1	再現率
(1)	0	1673	605	0.73
	1	421	1100	0.72
(2)	0	1687	591	0.74
	1	301	1220	0.80

結果は(2)の本研究では(1)の先行研究を模倣したものに比べ、ブランドスイッチの再現率が高く、予測精度が向上したことを示唆した。全体の購買パターンを見ることでより細かくパターンがわかり、再現率が向上する結果になったと考察する。

ここまでで、本研究で行ったブランドスイッチの予測分析の精度が向上したことを示したが、ブランドスイッチが起きると予測される上で影響が大きかった項目を重要度として解析し、示したものが図9である。

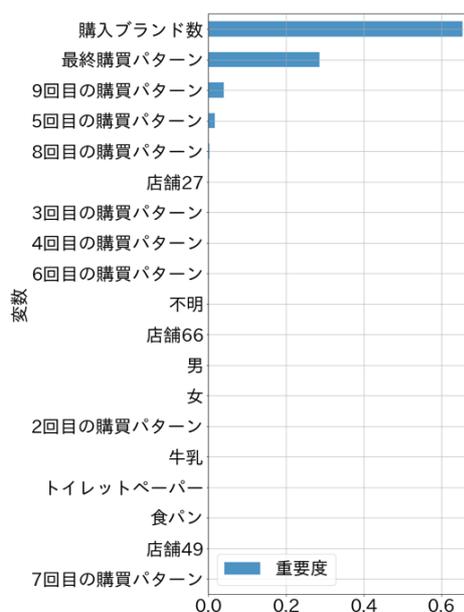


図9:重要度比較

図9より、ブランドスイッチが起きる上で重要な項目は「購入ブランド数」と「最終購買パターン」だけでなく、各回の購買パターンもブランドスイッチに影響することがわかった。つまり最終購買パターンだけでなく、全体における購買パターンを見ることで再現率が上がるということがわかった。また、商品、店舗の項目はブランドスイッチにおいて重要度が低く、差異がないことがわかった。

また本研究の予測分析により、精度は上がったことはわかったが、どのような顧客でもブランドスイッチするかどうかを予測することができるかを検討するために、今回は各顧客の過去の履歴におけるブランドスイッチの回数と予測精度の関係について分析した。以下のグラフは顧客のそれぞれのブランドスイッチ回数に対し、その中でブランドスイッチの予測が成功している顧客の割合を示したものとなっている。

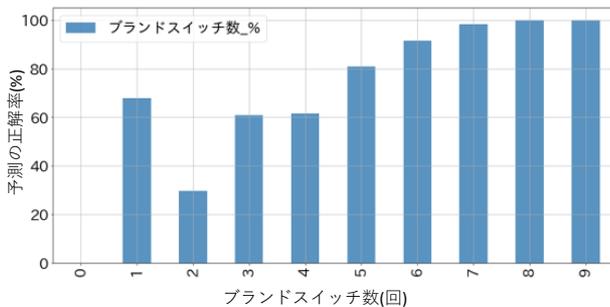


図10:ブランドスイッチ回数と予測精度の関係

図10よりどのブランドスイッチ回数の顧客でも予測はできているが、ブランドスイッチ回数が多い顧客で予測の正解率が高い傾向があった。またブランドスイッチ回数が少ない顧客では比較的ブランドスイッチの予測は難しい傾向にある。

6 考察

ここで、本研究で行った予測手法について考察する。まず精度について、以前のように最終購買パターンだけでなく、今回追加した全購買行動パターンを見ることで再現率が上がった。これはブランドスイッチを行った全パターンを網羅することにより精度が上がったと考えられる。次に汎用性について、ブランドスイッチにおいて店舗間、商品による影響がないことが重要度の検討によりわかった。この結果から、同じ系列店の多くの店舗や商品でこの予測分析が利用できる可能性が示された。

本研究で提案した予測手法によって過去の各顧客の購買データからブランドスイッチするタイミングをより正確に予測ができた。実際に店舗でこの予測

分析を用いて入店時にクーポンを発行することで顧客に対して当日の購買を促すことができると考えている。しかし、ブランドスイッチが多い顧客は全体の購買パターンで予測することはできるが、課題点としてブランドスイッチ回数が少ない顧客は予測が困難であることがわかった。また、今回は総合スーパー3店舗の購買データを使用して予測分析を作成したが、顧客の回遊データを使用した予測分析を作成したいと考える。また、予測分析に、商品の種類を増やして、汎用性の限度を知りたいと考えている。

7 まとめ

本研究では、総合スーパーの課題として、オンラインショップが普及され、既存顧客の売り上げ向上が難しいことが挙げられている。そこで顧客の当日の購買意欲を湖上させるため、ブランドスイッチを行うタイミングを予測しようと考え、顧客個人の購買行動データに基づくバラエティ・シーキング行動予測手法の提案を行った。

研究アプローチとして、データ基礎集計では顧客の購買行動の確認を行った。そして関連研究と本研究それぞれの予測分析のデータ加工と説明変数を作成し、それを用いて、バラエティ・シーキング行動予測分析の有効性の確認実験や重要度を比較した。最後にブランドスイッチ回数と予測精度の関係を見ていった。

結果は本研究で説明変数に各購買パターンを追加したことによってブランドスイッチの予測精度が上がった。また、店舗間、商品による際が生まれにくいことにより、多くの店舗や商品でこの予測分析が利用できる可能性が推察された。今後ブランドスイッチを行う顧客の特性をより詳しく分析することや、商品の種類を増やして分析を行う必要があると考えている。それによって更なるブランドスイッチの予測向上が期待でき、このような体制が構築されれば、実際に実店舗の売り上げ向上に貢献できるのではないかと考えている。

謝辞

分析データは株式会社トライアルカンパニー様よりいただいた。厚く御礼を申し上げ、感謝の意を示す。

参考文献

- [1] kinko 's お役立ちコラム
<https://www.kinkos.co.jp/wp/column/coupon-utilization/>
- [2] 【業界研究】スーパー業界の現状・課題・今後の動向・将来性 <https://mayonez.jp/topic/1106>

- [3] Esther P. Y. , Tang and Iris O. K. Chin: Analyzing Variety Seeking Behavior Using Panel Data, *Journal of International Consumer Marketing*, Vol.19, No.4, pp.7-31 (2006)
- [4] Edward Ho, Alexander Ilic : TOWARDS HIGH RESOLUTION IDENTIFICATION OF VARIETY-SEEKING BEHAVIOR, Proceedings of the European Conference on Information Systems (ECIS), pp.1-17 (2014)
- [5] Stephane Cheung, Yasuyuki Shirai, Hiroyuki Morita , Hiroyuki Takashima , Masakazu Nakamoto , Edward Hak-Sing Ip : Application of Hidden Markov Model to Analyze Enthusiasts' Dynamics of A Lifestyle Brand , 49th Hawaii International Conference on System Sciences, Vol. 1 , pp.1557-1566 (2016)
- [6] 新美潤一郎, 星野崇宏: 顧客行動の多様性変数を利用した購買行動の予測, 人工知能学会論文誌, Vol.32, No.2, pp.1-9 (2016)