

ECサイトの閲覧履歴に基づく商品購入断念要因の推定

小林 亮太[†] 豊田 哲也[‡] 大原 剛三[‡][†]青山学院大学 大学院理工学研究科[‡]青山学院大学 理工学部

1. 背景

現在, Amazon などの EC サイトにおけるネットショッピングが一般化しており, 多くの EC サイトでは商品推薦システムが導入されている[1]. これまで, それらのシステムの推薦精度を向上させるために, 商品閲覧履歴に基づいた, 商品選択条件の抽出法[2]などが提案されているが, 多くの場合, 実際に購入に至った事例における閲覧履歴を対象としている. これに対して本研究では, ユーザが商品購入を断念した事例に着目する. 一定時間検討した結果, 購入を断念した商品には, 購入を見送らせた要因があり, それはユーザにとって譲れないこだわり, すなわちユーザの嗜好に結びつくと考えられる. このような考えの下, 本研究では, 商品購入時に比較検討した複数商品の閲覧時間と閲覧頻度に基づいて購入断念要因となる商品仕様を推定する方法を提案する.

2. 閲覧履歴に基づく購入断念の要因推定

本研究では, ユーザが長時間閲覧した商品ほど興味度が高い[3]という仮定の下, 各商品属性の属性値に対するユーザの興味度を推定する. ここで商品属性とは, 価格や容量などの商品仕様を指す. 商品ごとの閲覧時間は, 商品詳細ページの閲覧開始時間と, そこからサイト内の別のページに移動した時間との差分とする. 同一の商品を複数回閲覧した場合, 閲覧時間はその合計とする. このとき, 閲覧時間に基づいて商品を 2 つのグループに分類する. 具体的には, 閲覧時間の長さ上位半分を長時間閲覧グループ, 下位半分を短時間閲覧グループとする.

直観的には, 連続した商品閲覧において, 閲

覧時間の長い商品の多くに共有される属性値ほど, ユーザの興味度は大きいと考えられる. この考えに基づき, 商品属性 A における特定の商品属性値 a に対するユーザ u の興味度 $S(a, u)$ を以下のように定式化する.

$$S(a, u) = \frac{N_{long}(a, u)}{N_{long}(u)} \quad (1)$$

ここで, $N_{long}(u)$ はユーザ u の長時間閲覧グループにおける商品数, $N_{long}(a, u)$ はその中で商品属性値 a をもつ商品の数である. 商品属性値 a の閲覧商品数が多いほど $S(a, u)$ の値が大きくなり, 興味度が大きいと推定する.

一方, 閲覧した商品間で異なる商品属性値を持つ場合を考える. 商品属性によっては, 閲覧した商品属性値が多岐にわたるものがある. そこで, 特定の商品属性値をもつ商品を高頻度で閲覧しているか否かを推定するためのスコアである変動係数 $CV(A, u)$ を式(2)によって算出する.

$$CV(A, u) = \frac{\sigma_A}{\bar{S}(A, u)} \quad (2)$$

ここで, $\bar{S}(A, u)$ は, 商品属性 A の各商品属性値に対するユーザ u の興味度の平均であり, 式(3)で求められる.

$$\bar{S}(A, u) = \frac{1}{|V_A|} \sum_{a \in V_A} S(a, u) = \frac{1}{|V_A|} \quad (3)$$

また, σ_A は商品属性 A の各属性値に対するユーザ u の興味度の標準偏差であり, 式(4)で求められる.

$$\sigma_A = \sqrt{\frac{1}{|V_A|} \sum_{a \in V_A} \{\bar{S}(A, u) - S(a, u)\}^2} \quad (4)$$

ここで, V_A は商品属性 A に対する属性値集合を示す. 式(2)で求められる値が大きいほど, 特定の商品属性値を高頻度で閲覧しており, 興味のある商品属性値がはっきりしていると判断する.

商品属性の例として, テレビのインチ数を考える. あるユーザに対する長時間閲覧グループのうち, インチ数が 24, 32, 42 である商品数がそれぞれ 8, 3, 1 であった場合, 各属性値の興味度 $S(a, u)$ と, この商品属性の変動係数 $CV(A, u)$ の算出結果は表 1 に示すような値となる. 表 1

Estimation of Online Shopping Cart**Abandonment Factor Based on Browsing History**Ryota KOBAYASHI[†], Tetsuya TOYOTA[‡] and Kouzou OHARA[‡][†]Graduate School of Science and Engineering, Aoyama Gakuin University[‡]College of Science and Engineering, Aoyama Gakuin University

表1: テレビのインチ数における興味度のスコアと変動係数の例

インチ数 (インチ)	商品数 (点)	$S(a, u)$	$CV(A, u)$
24	8	0.67	0.74
32	3	0.25	
42	1	0.08	

からわかるように、24インチの興味度が0.67という最大値をとることから、このユーザ u は特に24インチの商品に興味を示しているといえる。また、 $CV(A, u)$ が0.74と高い値を示していることから、24インチの商品に興味が集中していると仮定できる。一方、仮に、すべての商品属性値の出現回数が等しい場合、 $CV(A, u)$ は最小の0となる。変動係数の値が高くなるような、ユーザの興味は特に集中している商品属性は、商品購入などの成功事例からも推定できる要因である。このような顕在的なこだわりに対して、変動係数が低い商品属性にも、ユーザがこだわりの属性値が存在し得る。具体的には、ユーザがある商品属性の特定の値に興味をもっていたとしても、その値が他の要求仕様と同時に満たされない場合、その商品属性に対するどの属性値の閲覧頻度も、突出して高くない場合がある。ここでは、そのような潜在的なこだわりが存在すると仮定し、それを購入断念の要因として、変動係数の値を基に推定を試みる。実際の推定では、変動係数が閾値以下のときに、その商品属性を断念要因であると推定する。

3. 評価実験

3.1. 実験設定

本実験では、楽天商品検索 API¹が提供する商品データを利用して架空 EC サイトを構築し、24歳から28歳の被験者11名による閲覧履歴を収集した。使用した商品ジャンルと商品属性を表2に示す。各被験者には表2中の5つの商品ジャンルから2つを選択してもらい、商品ジャンルごとに自由に商品閲覧してもらった。なお、変動係数の閾値は、予備実験の結果に基づいて0.5に設定した。また、購入の意思がない、閲覧商品数が2以下、断念せずに購入に至ったような履歴は、推定の対象外とした。

3.2. 実験結果と考察

提案手法に基づく分析を、被験者ごとに行った結果、3名の被験者に対して合計10個の商品属性を断念要因と推定した。これらの被験者が

表2: 商品ジャンルと商品属性の一覧

商品ジャンル	商品属性
液晶テレビ	価格, 質量, チューナー数, インチ数, 解像度, HDMI端子数
デスクトップPC	価格, インチ数, 解像度, 容量, メモリ, CPU種類
ノートPC	価格, インチ数, 解像度, 容量, メモリ, CPU種類, 質量
コンパクト デジカメ	価格, 質量, F値, 画素数, 大きさ
デジタル一眼レフ	

実験後のアンケートで購入断念要因として回答した商品属性は合計4個あり、それらはすべてこの推定結果に含まれていた。

得られた結果から、提案手法で推定した断念要因には、潜在的なこだわりがある商品属性が含まれることが明らかになった。しかし、変動係数が低い商品属性には、単にこだわりのない属性と、潜在的なこだわりとなる属性の2つが混在しており、その区別ができないため、誤って推定した属性があったと考えられる。

4. 結論

本研究では、購入に踏み切れなかった要因がユーザの嗜好に結びつくという考えの下、商品購入を見送ったユーザの閲覧履歴から購入断念要因を推定した。実験結果から、被験者が回答した購入断念要因はすべて同定できているものの、その他の商品属性を誤って要因として推定する可能性が高いことがわかった。

今後、変動係数が0.5以下の商品属性の中から、他の商品属性との組み合わせなどを考慮して、真にこだわりのあるものを絞り込む方法を検討する必要がある。

謝辞 本研究では楽天株式会社の楽天商品検索APIが提供する商品データを利用した。

参考文献

- [1] 神畷敏弘: 推薦システムのアルゴリズム(1), 人工知能学会誌, Vol. 22, No. 6, pp. 826-837 (2007).
- [2] 江崎健司, 倉島健, 市川祐介, 高屋典子, 内山匡: ネットショッピングにおける商品選択条件の推定, 日本データベース学会論文誌, Vol. 11, No. 1, pp. 13-18 (2012).
- [3] Morita, M. and Shinoda, Y.: Information Filtering Based on User Behavior Analysis and Best Match Text Retrieval, Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '94), pp. 272-281 (1994).

¹ <http://webservice.rakuten.co.jp/>