

# アクセスログデータ可視化の試み

正木 俊行, 伊豆永 洋一, 佐藤 俊樹, 鮎川 矩義, 石濱 友裕,  
田中 彰浩, 中島 雄基, 舟橋 史明

EC サイトの運営者にとって最も重要な課題の一つは、日々蓄積されていくアクセスログデータから、当該サイトの改善につながる有用な知見を発掘することである。本稿では、このような課題の解決を支援する道具の一つとして、サイトを訪問する個々のユーザの閲覧行動を可視化する方法を提案する。

キーワード：アクセスログ解析、可視化、消費者行動、One to One マーケティング、CRM

## 1. はじめに

EC (E-Commerce) サイトの管理者にとって最も重要な課題の一つは、日々蓄積されていくアクセスログデータから、当該サイトの改善につながる有用な知見を発掘することである。

アクセスログデータとは、サイトを訪問したユーザのアクセス情報を集めたものであり、いつ、誰が、どこから来て、どのようなページを、どれだけの時間閲覧したか、などの情報を含んでいる。非常に膨大かつ複雑なデータであるため、ひとまず、専用のアクセスログ解析ツールを通して俯瞰されるのが一般的である。そこで得られた解析結果は、図やグラフの形式で計算機画面上に出力され、管理者がその後の分析方針を決定するための重要な判断材料となっている。そのため、どのような解析ツールを用いるか、さらには、得られた結果をどのように表示するか、といった問題は応用上重要である。

従来のアクセスログ解析ツールは、基本的な集計を行うものであり、具体的な指標としては、ページビュー数、訪問者数、訪問回数、滞在時間などがある。これらの解析ツールは、当該サイトの「要約された」アクセス傾向を把握する際には有用であるが、「個々のユーザの動向に着目した」知見の発掘や分析方針の立案に役立てるのは困難である。

われわれは、このような問題点の解決を目的とし、個々のユーザの閲覧行動を可視化する方法を提案する。1人のユーザが指定されたとき、(当該サイトへの)一つのアクセスを計算機画面上の一つの点で表現し、一連のアクセス、すなわちセッションを2次元平面上の「折れ線グラフ」のような形状でとらえることを目指す。従来のアクセスログ解析ツールが、「広く・浅く」データを俯瞰するのに対し、われわれが提案する可視化手法は、データを「狭く・深く」とらえることを目指している。

本研究で提案する手法のように、基礎分析の段階で人間の知恵を組み込むための「情報可視化」は、探索的データ解析における基本的な手法として近年注目を集めている [1]。また、われわれが提案する手法のように、個々のユーザに着目した分析方針は、One to One マーケティングや CRM の実現のためには必要不可欠である [2]。

本稿では、上述の方針に基づいてわれわれが開発した可視化ツールを紹介する。さらに、このツールを実社会のアクセスログデータに適用し、そこで得られた知見について議論する。

## 2. 使用データ概要

本研究では、経営科学系研究部会連合協議会主催、平成 23 年度データ解析コンペティションで提供されたゴルフ用品を販売する EC サイトのアクセスログデータを用いる。概要は以下のとおりである。

期間：2010 年 7 月 1 日～2011 年 6 月 28 日

アクセスログの項目：イベント発生日時、リファラーページ、アクセスページ、受注 ID、セッション滞在時間、商品カテゴリ（中分類・小分類）など。

会員情報：会員の性別、生年、ハンディキャップ、メー

まさき としゆき, いずなが よういち, さとう としき,  
いしはま ともひろ, たなか あきひろ, なかしま ゆう  
き, ふなはし ふみあき  
筑波大学 システム情報工学研究科  
〒 305-8577 茨城県つくば市天王台 1-1-1  
すけがわ のりよし  
東京工業大学大学院社会理工学研究科  
〒 152-8550 東京都目黒区大岡山 2-12-1

表 1 商品カテゴリ（中分類・中小分類・小分類）

中分類	中小分類	小分類
クラブ	クラブ	アイアン、ウェッジ、クラブセット、シャフト、ドライバー、など。
用品・小物	練習・リペア	練習器具、リペアグッズ、距離計測器、練習器具、など。
	バッグ・カバー	クラブケース、ヘッドカバー、キャディバッグ、など。
	ラウンド用品	ラウンド小物、ボール、シューズ、ティー、傘、グローブ、など。
ウェア	トップス	アウター、ベスト、レインウェア、ワンピース、セーター、など。
	ボトムス	パンツ、スカート、ハーフパンツ、ロングパンツ、など。
	ファッション小物	ウェア小物、くつ下、帽子、サングラス、ベルト、など。
その他	その他	カレンダー、DVD・チケット、家電、健康グッズ、書籍、など。

表 2 アクセスログデータの例

イベント発生日時	商品カテゴリ	セッション ID
11/25 22:54	ボール	1121
11/26 19:12	クラブ	1314
11/26 19:13	ボール	1314
11/26 19:13	ボール	1314
11/26 19:18	ウェア	1314
11/26 19:22	クラブ	1314
11/29 20:03	ボール	1563

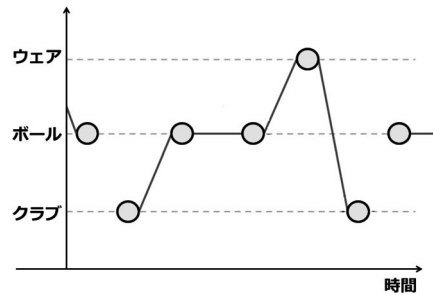


図 1 表 2 におけるセッション 1314 の可視化イメージ

ルマガジン登録、会員登録日など。

受注情報：受注 ID、メーカー名、受注数量、受注金額、ポイント・クーポン情報など。

### 3. 提案する可視化手法

われわれは、ユーザの閲覧行動のうち、特に、「どのような商品に興味を持っているか」に着目して可視化を行う。具体的には、ユーザが閲覧した商品ページに付随する「商品カテゴリ」に着目する。

縦軸に「商品カテゴリ」、横軸に「時間」をとる 2 次元平面を考え、各アクセスをこの平面上の点として表現する。平面上の 2 点は、対応するアクセスが、与えられたアクセスログデータにおいて連続し、さらに、同一セッション中に生じたものであるとき、直線で結ばれる。これにより、一連のアクセス、すなわち一つのセッションが、2 次元平面上の「折れ線グラフ」として表現される。

われわれが提案する可視化手法の基本方針は以上である。詳細な要望に対しては、開発する可視化ツールのオプションによって対処する方針を取る。

### 4. 商品カテゴリ

元のデータには、「中分類」と「小分類」という二つの商品カテゴリが与えられている。中分類は 4 項目、小分類は 48 項目で構成されている。可視化を行う際、中分類では非常に大まかな傾向しかつかみず、逆に、小分類では、分類が比較的細かいため、可視化された結果がわかりづらくなる可能性がある。そこでわれわれは、これらの折衷案として、項目数が 8 つの「中小分類」を用意する。表 1 に、これらの分類の関係を示す。

### 5. 具体例

ある 1 人のユーザに着目し、表 2 に示すようなアクセスログデータが得られたとする。このとき、われわれの提案する可視化のイメージは図 1 のようになる。縦軸における商品カテゴリの順序や、横軸における単位は、分析者の目的に合わせて変更するものとする。

縦軸としては、たとえば、傾向の似たカテゴリを近くに配置することが考えられる。この場合、可視化した際に現れた折れ線グラフが「ギザギザ」していれば、着目しているユーザがさまざまな商品に興味を持って

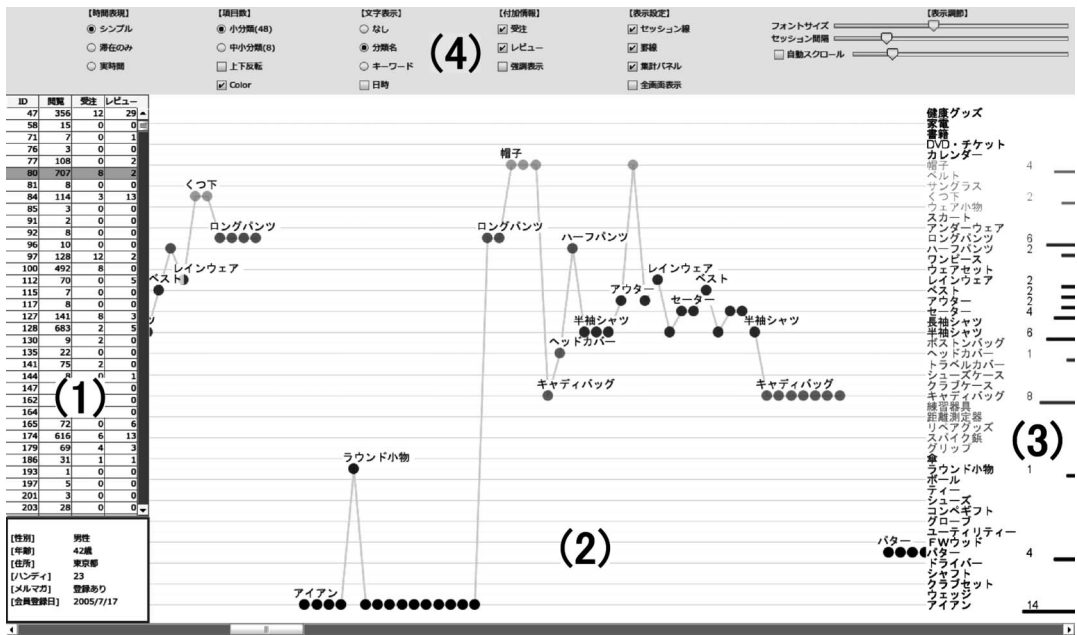


図2 提案する可視化ツールの動作画面（縦軸を48項目の小分類にした場合）

いる、もしくは、ただ単に回遊している、ということが予想できる。また、ほぼ一直線上に乗っていれば、選択肢が限定されていることが予想できる。

図2は、上述の方針に基づいてわれわれが開発した、アクセスログデータ可視化ツールの動作画面の一例である。ここでは、縦軸として小分類をとり、中小分類によって点の色を変えている。開発言語にはJavaを使用し、データを扱いやすい形式にするため、Microsoft Access 2010による前処理を施している。

## 6. 機能の詳細

提案する可視化ツールの画面は4種類のパネルから成り、(1)「ユーザー一覧パネル」、(2)「表示パネル」、(3)「集計パネル」、(4)「設定パネル」がある。図2中にある括弧つきの数字は、これらに対応する箇所を示している。各パネルの詳細は以下のとおりである。

### 6.1 ユーザー一覧パネル

各ユーザーのユーザーID、閲覧回数、受注回数、レビューページの閲覧回数が表示される。ユーザーを選択すると、(次に説明する)表示パネルに、そのユーザーの閲覧行動が表示される。このパネルの下部には、そのユーザーの性別、年齢、住所、ハンディキャップ、メルマガ登録の有無、会員登録日が表示される。

### 6.2 表示パネル

選択されたユーザーの閲覧行動を可視化した結果が表示される。各点は、対応する商品が属する中小分類ご

とに色分けされている。マウスホイールによる時間軸方向の伸縮、マウスドラッグによる左右のスクロールが行える。

### 6.3 集計パネル

表示パネルに登場する点に対応するアクセスを、商品カテゴリごとに集計した棒グラフが表示される。これにより、選択されたユーザーが（その時点で）興味を持っている商品カテゴリを把握できる。表示パネルに登場しているアクセスに関する集計であるため、表示パネルの拡大縮小・スクロールに連動して動的に変化する。

### 6.4 設定パネル

表示パネルに出力される点や縦軸を変更するためのパネルであり、6つの項目がある。詳細は以下のとおりである。

**閲覧時間評価:** 表示パネルに登場する点の大きさを変更する。点の大きさを一定にする場合(OFF)と、対応するアクセスの滞在時間の対数に比例させる場合(ON)の2種類がある。図4は閲覧時間評価をONにした場合の一例である。

閲覧時間だけではなく、連続する2つのアクセスが、実時間で、どの程度離れているかを表示パネルに反映させることも考えられる。しかし、今回われわれが開発した可視化ツールではこれらの情報をまだ反映できておらず、今後の課題の1つとなっている。

**項目数:** 縦軸の商品カテゴリとして、48項目の小分類

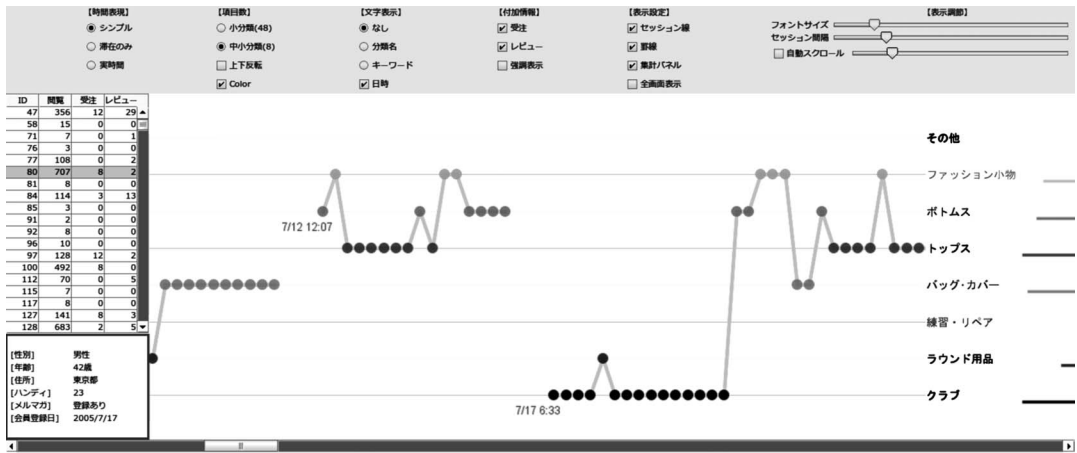


図 3 縦軸を 8 項目の中小分類にした場合

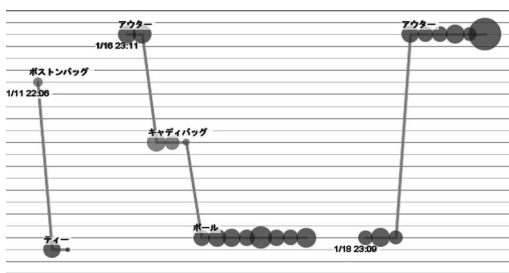


図 4 閲覧時間評価を ON にした場合

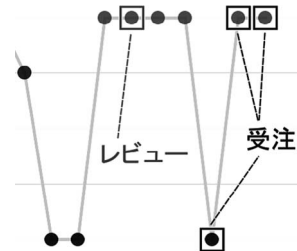


図 5 レビューページの閲覧や受注の情報を表示した場合

(図 2) と 8 項目の中小分類 (図 3) のどちらかが選択できる。

**文字表示:** 表示パネルに付加的な文字情報を載せるかどうかを選択できる。文字情報としては、対象とするアクセスに対応するカテゴリ名、そのセッションへの流入キーワード、アクセス日時が表示できる。

**付加情報:** レビューページの閲覧に対応するアクセスであること、または、受注 (ユーザ目線では、発注) に対応するアクセスであることを明示するかどうかを選択できる。レビューページの閲覧に対応する点は青い四角、受注に対応するそれは赤い四角で囲われる。図 5 はこれらを表示した場合の一例である。

**表示設定:** 罫線の有無や集計パネルの表示・非表示、全画面表示への切り替えなどが設定できる。また、フォントサイズや、セッション間のスペース、自動スクロールの速度を調節できる。

## 7. 可視化によって得られた知見

本節では、第 2 節で説明した EC サイトのユーザの閲覧行動を可視化した際に得られた知見について議論

する。

**知見 1:** 「高価格な商品ほど、発注に至るまでに長い時間が費やされる。」

低価格の商品から高価格の商品まで、幅広い価格帯の商品を発注した経験のあるユーザを選択し、その閲覧行動を可視化した際に観測された結果である。その際、アクセス回数と閲覧時間の両方を加味している。すなわち、(図 4 に示した) 閲覧時間評価を ON にして可視化した際、高価格帯の商品 (たとえば、ドライバークラフやキャディバッグ) に対しては、たくさんの、なおかつ比較的大きな、点が続く、その後で発注に至る、という傾向が確認された。

**知見 2:** 「レビューページの利用頻度はユーザによって大きく異なる。」

図 6 と図 7 は、同程度の閲覧回数と発注回数を持つ、ある 2 人のユーザの閲覧行動を可視化したものである。どちらのユーザも、ほとんどすべての (中小分類の意味での) 商品カテゴリの商品を発注した経験がある。ここでは、「受注」に対応する点を囲うことはしていない。すなわち、これらの図の中で囲われた点は、どれもレビューページの閲覧に対応している。

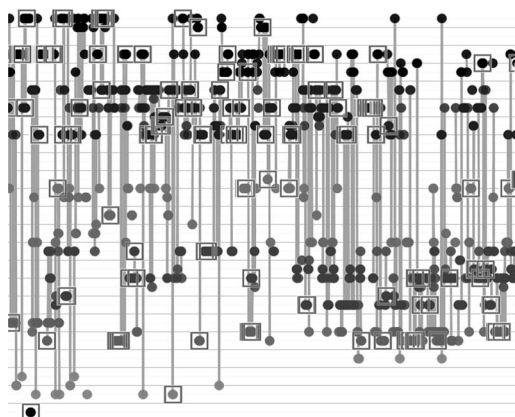


図6 レビューページを頻繁に閲覧するユーザーの閲覧行動

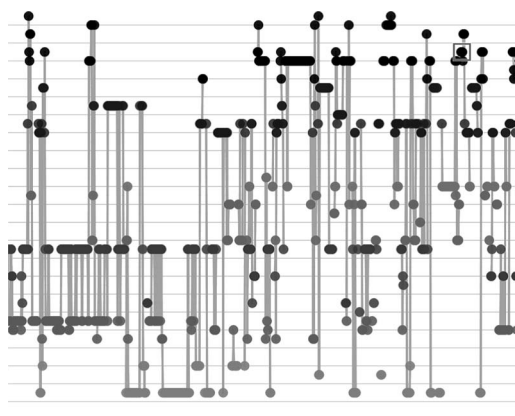


図7 レビューページをあまり閲覧しないユーザーの閲覧行動

図6に対応するユーザーは、(商品を発注する際に)レビューページを参考にする傾向が強く、反対に、図7に対応するユーザーは、ほとんどレビューページを見ないことがわかる。

また、商品カテゴリによってもレビューページにアクセスされる割合は大きく異なる。中小分類でいえば、「練習・リペア」や「クラブ」、「ラウンド用品」といった使用感が重要視される商品カテゴリにおいて、レビューページへのアクセスの割合が高い。

**知見3:**「発注の直前での閲覧行動は、ユーザーによって大きく異なる。」

図8と図9は、発注の直前での閲覧行動が特徴的、なおかつ対照的な2人のユーザーの閲覧行動を可視化したものである。

図8に対応するユーザーは、発注の直前に、その商品のページを継続的に閲覧している様子が確認できる。3つ目のセッションは、購入を視野に入れている3種類

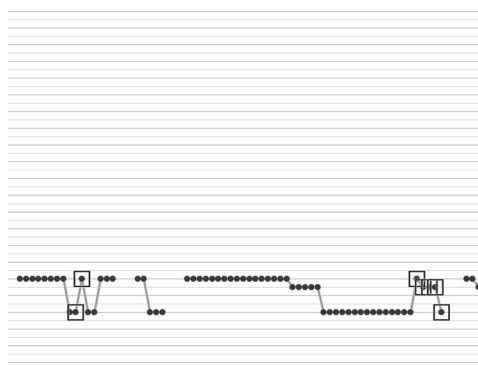


図8 商品ページを継続的に閲覧した後に発注するユーザーの閲覧行動

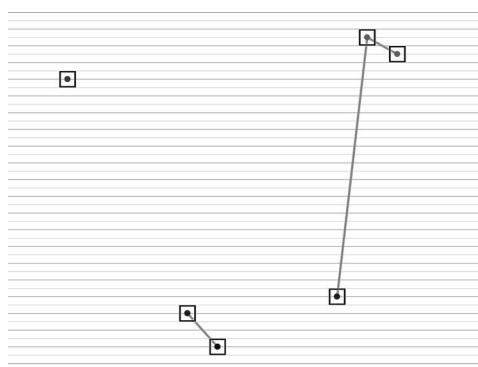


図9 商品ページをあまり閲覧することなく発注するユーザーの閲覧行動

の商品があり、どの商品を購入するかを決定(おそらく「買い物カゴ」や「ブックマーク」などを利用)し、最後に、それらをまとめて発注した様子を表しているものとみなせる。

一方で、図9に対応するユーザーは、発注する商品のページをほとんど閲覧することなく受注に至っている。このようなユーザーは、発注する商品が決まった後に当該サイトを訪問している可能性が高い。たとえば、実店舗で使用感を確かめ、購入を決めた商品名を記憶し、その後で当該サイトを訪れるユーザーはこのような閲覧行動をとるであろう。

また、ある商品を継続的に、かつ頻繁に閲覧しているながら、ある時期を境にアクセスが途絶えるユーザーの存在も確認された。これらのユーザーは、当該サイトで商品を比較し、他のECサイトや実店舗で対象商品を購入している可能性がある。

## 8. おわりに

本稿では、One to OneマーケティングやCRMの



コンセプトに基づき、アクセスログデータを可視化する手法を提案し、サイトの管理者が調整できるいくつかのパラメータを備えた可視化ツールを実際に開発した。さらに、実社会から得られた生のアクセスログデータに適用し、当該サイトの改善につながる知見の発掘を試みた。

われわれが提案した可視化ツールは、個々のユーザーが興味を持っている商品カテゴリに着目しており、その遷移を視覚的に把握するのに役立つものである。基本的な集計を目的とした従来の解析ツールとは異なり、個々のユーザーの閲覧行動に着目して得られるような知見の発掘を目指している。受注の情報とともに確認することで、ユーザー特有の受注パターンを把握する手がかりとなり得る。

しかしながら、われわれが提案した可視化ツールを実際の分析に用いる際、どのユーザーの閲覧行動を可視化すればよいのか、という問題は、今後解決すべき問

題の1つである。たいていの場合、当該サイトを訪問するユーザーの人数は膨大であるため、注目すべきユーザーを効率良く選ぶ必要がある。

また、ある商品を買ったこと、ブックマークに保存したことなど、商品ページ以外の情報を可視化ツールに反映させることができれば、ユーザーのより詳細な行動パターンが把握できると考える。

**謝辞** データをご提供いただき、研究発表の際には、多くの貴重なご意見をくださったデータ解析コンペティション関係者の皆様に心からの感謝の意を表します。

#### 参考文献

- [1] D. A. Keim: "Information Visualization and Visual Data Mining," *IEEE Transactions on Visualization and Computer Graphics*, **8**, 1-8, 2002.
- [2] 古川一郎, 守口剛, 阿部誠, 『マーケティング・サイエンス入門』, 有斐閣, 2003.