

Soot を用いた解析コードの挿入による 電子商取引システムのユーザーログ生成手法

野口 建仁[†] 難波 和明[‡]

[†] 東京理科大学大学院理工学研究科経営工学専攻

大和田 勇人[‡]

[‡] 東京理科大学理工学部経営工学科

1 序論

インターネットの爆発的な成長により、インターネット上には数多くのショッピングサイトが公開され、サイトごとに様々な狙いをもった商品を売り出している。これらショッピングサイトの管理者は自分が売りたいと考えている商品についてどのように売り出していけば顧客が購入してくれるのかを考える必要があり、おのずと顧客のサイト内での行動を知ることが必要不可欠となり、サイトを利用するユーザーについて知るために、Web サーバーのログファイルを解析することに关心が高まっている。これにより、このログファイルを解析し、ユーザーの行動パターン等の情報を分析、要約、視覚化する数々の Web ログ分析ツールが開発されてきた。

LEE[1] は、オンラインショッピングサイトでどのユーザーがどのように行動したのかの足跡であるクリックストリームについて分析し、平行座標によるセッションの視覚化とスターフィールドによるパフォーマンス効果の視覚化という 2 つの異なる観点からのアプローチを提案した。大河ら [2] はこれを基に Web ログデータを動的にデータベースに格納する Web ログ・データベース変換システムの導入を行い、これにより Web ログデータをサイト管理者に手間とせることなくデータベースに格納させた。

1.1 本論文の目的

しかし、これらデータベース変換システム及び視覚化システムについて改善すべき点がいくつかある。

1. クリックストリーム取得方法.
2. “Referer”ごとの表示しかない.
3. 実装したサーバー内でしか視覚化結果を確認できない.

これらの問題に対して本研究の目的として、既存の視覚化システムの改善を行うための以下のようないくつかの改善案を提案する。

The user log generation technique of the electronic commerce system by the insertion of the analysis code which I used Soot for.

[†]Takehito Noguchi [‡]Kazuaki Namba [‡]Hayato Ohwada

[†]Department of Industrial Administration, Graduate School of Science and Technology, Tokyo University of Science

[‡]Department of Industrial Administration, Faculty of Science and Technology, Tokyo University of Science

- ユーザーごとのクリックストリーム取得、その利用 解析コード挿入によるログデータ取得から、各クリックストリームごとに正確なデータをデータベースに格納。新たにユーザーごとにテーブルを作成することにより、ユーザーごとの正確なクリックストリームを取得する。
- 視覚化システムのオンライン上での実行、表示 PHP, JavaScript を使うことでオンラインでのシステム実装を行い、サイト管理者の解析場所を制限させない。

2 提案手法

本研究の提案手法は大きく分けて二つある。一つはユーザーごとのアクセス情報を取得するための Soot を用いたユーザーログ生成手法、もう一つはこの生成したユーザーログを用いてユーザー情報を用いた平行座標によるアクセス情報の視覚化である。

2.1 Soot を用いたユーザーログ生成

Soot[3] とは Java におけるソースファイルとクラスファイルの間に位置する中間言語 Jimple 形式のファイルを運用するツールである。これによりプログラムファイルを 1 行に 1 命令の記述として明確に理解できる状態で扱うことができるため、適宜解析コードを挿入することが可能となる利点がある。

ユーザーログ生成では、Soot[3] を用いて解析コードをプログラム中に適宜挿入することでユーザーがそのページを訪れるごとに解析に必要なデータを含んだログデータを出力する。出力されたログデータとデータベース内のユーザー情報が登録されているテーブルを用いて煩雑なログデータを、ユーザー単位のユーザーログとしてデータベース内のユーザーごとのクリックストリームデータを記録するテーブルに格納する。

2.2 ユーザー情報を用いたアクセス情報の視覚化

ユーザー情報を用いたアクセス情報の視覚化では、ユーザーごとにまとめたクリックストリームデータとユーザーの情報を記録しているテーブルとをリンクさせることによりユーザーごとのクリックストリームの平行座標

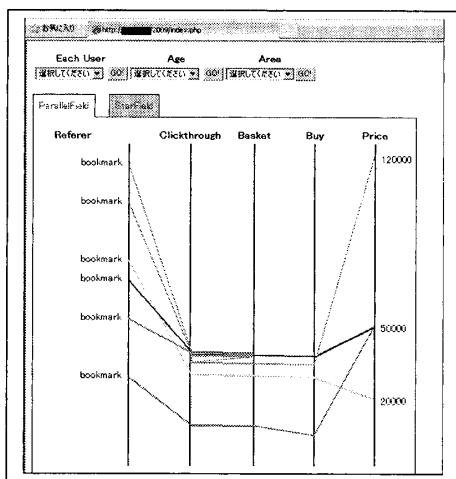


図 1: 平行座標視覚化の例

による視覚化、年代別や地域別のクリックストリームの視覚化を行う。

3 実装

本研究では平行座標を作成するための基本設計を大河ら [2] の研究から用いてこれを PHP, JavaScript で再構築する。これは、既存研究ではログ分析結果をローカルでしか確認することができなかった問題に対し、PHP, JavaScript で構築することによりオンライン上での使用を可能にすることで、解析、分析場所を制限しないことを目的としたためである。

図 1 は本研究でのクリックストリームデータの平行座標による視覚化例である。上のオプションメニューによりユーザーの絞り込みや年代、地域といった条件を設定することができる。またメインの画面では実際のユーザーの動きを大きく “Clickthrough”, “Basket”, “Buy” の三つに分け経路を分析し視覚化している。画面の左から “Referer” はこのユーザーがどこからやってきたのか (図ではブックマークからやっている)。“Clickthrough” は製品の詳細画面ページに訪れた時間、“Basket” はショッピングカートページに訪れた時間、“Buy” は商品を購入完了したページに訪れた時間、“Price” はそのクリックストリーム中で購入した商品の合計である。時系列は 24 時間単位となっていて上が 0 時、下が 23 時 59 分となっている。

サイトの管理者、解析者はこの図を見ることで、ユーザーがどこから訪れ、どのような経路を辿って購入まで行き着いたかを見ることができ、さらには購入金額からそのユーザーが自分のサイトに対してどれほど貢献してくれているのかを知ることができるようになる。

4 考察

これらの実装結果から、既存研究に対して以下の改善点が挙げられる。

- ユーザーごとのクリックストリーム視覚化

従来までは “Referer” ごとの表示しかなかったのに対し、本研究においてはユーザーごとのクリックストリームを視覚化した。これにより、“Referer” ごとにしか分析できなかったクリックストリームをユーザー単位で行うことができ、サイトを利用するユーザーごとの利用傾向、経路分析を知ることはもちろん、年代別、地域別に特定のターゲットを考えているサイト管理者は、ターゲットには商品のページを見てもらっているのか、ショッピングカートに入れてもらったのに商品購入まで進んでもらえなかった、などの結果からマーケティングの分析を行うことができる。

- オンライン上の視覚化結果の表示

従来までは Java による実装だったため、ログの分析結果、視覚化結果をローカルな場所でしか確認できず分析場所を制限されていた。これに対し、本研究では PHP, JavaScript で実装することにより、インターネット上の実行が可能となり、サイト管理者や解析者に対しログ分析を行うための場所を制限しなくなった。

5 結論

本研究では、既存研究におけるシステムの問題点の改善とその有効性を提案した。今後は、既存研究の購入商品について分析を行える視覚化技術に対しての実装も検討している。

参考文献

- [1] JUHNYOUNG LEE, Visualization and Analysis of Clickstream Data of Online Stores for Understanding Web Merchandising, Data Mining and Knowledge Discovery, 5, 59-84, 2001
- [2] 大河 隼. 平行座標とスターフィールドを用いたデータマイニング視覚化ツールの開発. 東京理科大学学士論文, 2003.
- [3] 石垣一, 松井藤五郎, 大和田勇人. JavaServlet におけるクラス間の依存情報抽出, 日本ソフトウェア科学会第 22 回大会, (2005).