

携帯電話のセッション推定に基づくアクセスログの補完について

中村正人[†] 近藤圭佑[†] 大園忠親[†] 新谷虎松[†]

[†]名古屋工業大学大学院工学研究科情報工学専攻

1 はじめに

携帯電話は、移動端末という特性上アクセス毎に IP アドレスが異なる可能性がある。また、端末によっては Cookie や JavaScript, リファラが利用できない。そのような制約の中で、携帯電話のアクセス解析はフルブラウザ用の解析と同様にユーザの閲覧行動を詳細に解析することが求められる。

携帯電話のアクセス解析は、上記の制約のため、フルブラウザ用のアクセス解析と分離して行われる傾向にある。アクセス解析の目的はユーザの動向を把握することであり、これは携帯電話用の解析とフルブラウザ用の解析において共通である。従って、両者の解析において、共通のアクセス解析機構を用いることが望ましい。本研究では、携帯電話のセッションを推定した上で、リファラやセッション ID など携帯電話からのアクセスで欠如した情報を Web サーバ上で補完し、アクセスログの整形を行う。これにより、携帯電話用とフルブラウザ用のアクセス解析を統合することが可能となる。

2 端末の識別

携帯電話のアクセス解析において、ユーザ及びセッションの識別が重要である。携帯電話のキャリアによっては、URL に特殊なパラメータ (以下 UID パラメータ) を付加することで端末識別番号を取得できる。この時、外部のサイトからリンクを経由して流入するアクセスに対し、端末識別番号を取得困難な点が課題となる。端末識別番号を利用可能な携帯電話であっても、URL に UID パラメータを付加しなければ、端末識別番号を取得できない。サーチエンジン等外部のサイトから自分のサイトに対して向けられたリンクに対して、UID パラメータを付加することはできない。そのため、サイト外部からのリンクを経由したアクセスでは端末

An Access Log Supplement Method for Mobile Website using Session Estimation

Masato NAKAMURA[†], Keisuke KONDO[†], Tadachika OZONO[†] and Toramatsu SHINTANI[†]

Dept. of Computer Science and Engineering, Graduate School of Engineering, Nagoya Institute of Technology
466-8555, Nagoya, Japan

```

SID ← セッション ID;
UID ← 端末識別番号;
UA ← ユーザエージェント;
FLAG ← TRUE;
NOW ← 現在時刻;
if SID = 有効な SID then
  OLD_UID ← 過去の SID 使用時の端末識別番号;
  OLD_UA ← 過去の SID 使用時のユーザエージェント;
  t ← NOW - SID を使用した最終アクセス日時;
  if UID = OLD_UID AND UA = OLD_UA AND t >
一定時間 T then
    FLAG ← FALSE;
    SID による最終アクセス日時 ← NOW;
  endif
endif
if FLAG = TRUE then
  SID ← 新規 SID;
  UID, UA, NOW, IP アドレス, リファラ, アクセスされた URL を SID に紐づけてデータベースに格納;
endif

```

図 1: セッション ID の生成手順

識別番号を取得することは困難である。対策として、UID パラメータのついていない URL を用いたアクセスの場合、UID パラメータを付加した URL へ転送する方法がある。しかしこの方法ではリクエストが 2 回発生する上、本来 1 回のアクセスではあるはずがページビューを 2 とカウントする。また、端末識別番号の送信はユーザの設定により拒否できる。そのため、ユーザの識別には端末識別番号以外の枠組みが必要となる。

Cookie や端末識別番号を利用できない携帯電話に対するセッションの判別には、サーバ側で発行するセッション ID を利用する他に手段がない。そのため、本研究では Web ページ出力時に動的にテキストの改変を行う。つまり、サイト内のページへ遷移する可能性のあるタグである A タグや FORM タグに対し、セッション ID を付加する。

セッション ID は、図 1 の手順に従い生成される。URL に対し端末識別のためのセッション ID を付加する手段は、他人になりすますことが可能である。それを防ぐため、過去に同一のセッション ID でアクセスがある場合に、端末識別番号を利用可能な場合は端末識

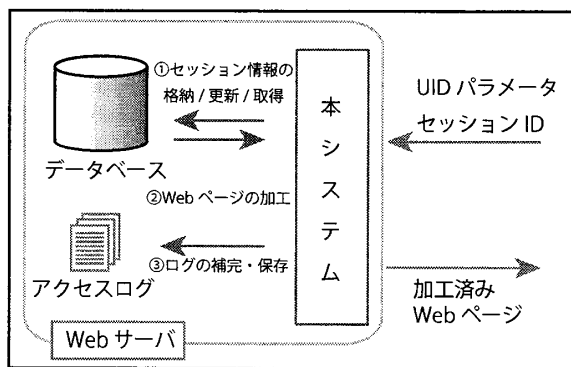


図 2: システム構成図

別番号を、それ以外の場合はユーザエージェントを使用して同一セッションの推定を行う。但し、端末識別番号が利用できず、尚かつ同一機種の端末でアクセスがある場合は、同一セッションとみなされるという欠点がある。

3 ログの補完

携帯電話からのアクセスにおいて、フルブラウザから送信される情報と比較して欠如する可能性のある情報は、Cookieとリファラである。本研究では、端末の識別にセッションIDを利用するため、Cookieは不要である。アクセス解析において必要となるのは、ユーザのページ遷移情報であり、サーバ上のログをベースに解析する場合、リファラの情報が必要となる。そこで、セッションIDを元にリファラの補完を行う。過去の最終ページアクセスを参照し、同一のセッションIDを持つURLをリファラとして追加する。また、携帯電話はアクセス毎にIPアドレスが異なる可能性があるため、同一セッションの場合は一貫したIPアドレスからアクセスがあったようにログを変更する。

既存の携帯電話向けのアクセス解析ツールでは、リバースプロキシとして動作するタイプ、アプリケーション側でセッション管理を行うタイプ、パケットキャプチャ型のタイプ、Webビーコン型などがある。本研究ではサーバ上のログを補完することを目的とするため、提案するモデルをWebサーバのプラグインとして試作した。サーバのプラグインとして試作することで、既存のWebページを変更することなく利用できる。

システム構成図を図2に示す。セッションに関する情報であるセッションID、アクセスページ、最終アクセス日時、最終アクセスページ、ユーザエージェント、IPアドレスをデータベースに保持する。Webサーバへリクエストがあると、図1の手順に従いセッションIDの発行などを行う。また、Webページに記述され

たページ遷移に関わるタグに対し、ページ遷移先が同一サーバ内である場合にセッションIDを付加する。ログ生成時には、セッションIDを元にIPアドレス及びリファラの補完を行う。

4 Webビーコン型アクセス解析への応用

フルブラウザに対するアクセス解析では、JavaScriptを用いたWebビーコン型が主流である。そのため、JavaScriptが動作しない携帯電話用ブラウザでは、フルブラウザ用のアクセス解析ツールをそのまま使用することはできない。一方提案手法では、本来フルブラウザがアクセス解析用サーバへ通知する情報をWebサーバが保持しているため、フルブラウザ用の解析ツールへ応用できる。つまり、本来フルブラウザのCookieへ保存される端末識別IDやリファラ、画面サイズなどの情報を、セッション情報を格納したデータベースを元に作成し、アクセス解析用サーバへ通知を行うことで、Webビーコン型アクセス解析を利用可能である。

5 おわりに

本稿では、フルブラウザ用アクセス解析ツールを用いて、携帯電話のアクセス解析を行う手法について述べた。現在普及している携帯電話用ブラウザは、その制約の多さからサーバへ送信する情報が少ない。携帯電話用アクセス解析としてWebビーコン型を用いる方法があるが、端末識別番号を取得できない場合ユニークユーザを判別できず、またリンクの加工を行えないため、セッションの特定は困難である。一方本モデルでは、端末識別番号が取得できない場合に対しても、ユーザが携帯電話のキャッシュからWebページを閲覧した場合であっても、ユニークユーザの判別が可能である。

携帯電話からのアクセスに対し、本モデルはフルブラウザからのアクセスのログと同様の形式でWebサーバのログを生成する。その為、フルブラウザ向けに利用されてきたWebページ推薦手法を、携帯電話用サイトに適用できる。Webサーバのログを用いる推薦手法として、山元らの研究[1]が挙げられる。この手法は、頻出アクセスパターンを検出し、ユーザに最短パスの推薦を行う。情報の転送量が少ない方が好まれる携帯電話では、このようなページ推薦が有効であると考えられる。

参考文献

- [1] 山元理絵, 小林大, 小林隆志, 横田治夫: “WebアクセスログのLCSを用いたWebページの推薦手法”, 信学技報, vol. 106, no. 148, DE2006-40, pp. 109-114, 2006.