

閲覧の起点と関連性に注目したユーザ意図の推移追跡

吉田 拓磨^{†,a} 中村 明順^{†,b} 安積 卓也^{†,c} 西尾 信彦^{†,d}

[†]立命館大学情報理工学部 [‡]立命館大学大学院情報理工学研究科

a) *yossy@ubi.cs.ritsumei.ac.jp* b) *nakamu@ubi.cs.ritsumei.ac.jp* c) *takuya@cs.ritsumei.ac.jp*

d) *nishio@cs.ritsumei.ac.jp*

概要 ユーザの Web 閲覧は集中するという仮説に基づいて閲覧履歴（ログ群）をまとめ、ユーザの意図を抽出し、活動を支援する研究が盛んに行われている。しかし関連のない複数ページの同時閲覧が考えられるため、時間的な近さからログをまとめる方法はユーザの意図の抽出に適していない。そこでリンク先への遷移やタブ切り替えといったログ間の関連性と、閲覧を開始した点に注目してユーザの意図を抽出し、この推移を追跡する。限定的な分析では、短時間内に複数意図で閲覧していた場合でも、意図ごとにログを連結できていた。今後はログ間の関連性があると判断した中でも、主題が大きく変化する場合に対処するため、ページの内容を考慮していく。

キーワード 閲覧行動履歴, ユーザ行動分析, パーソナライゼーション, 情報抽出

1 はじめに

近年, Google のパーソナライズ検索など, ユーザの Web 上での活動支援を目的として, 個々のユーザの Web 閲覧履歴 (ログ群) を用いてプロフィールを構築する技術 [1][2][3] に注目が集まっている。プロフィールを構築するために, ユーザの興味あるいは意図を抽出するには, ユーザの閲覧意図 (ユーザ意図) の推移を知ることが重要である。ユーザ意図の推移を知るために, 時間や閲覧順序の間隔などの指標によってログをまとめる手法 [4][5] が提案されている。しかし, 前述した手法ではユーザ意図の推移を捉えきれず, 複数の興味を単一の興味として扱われてしまう問題や, 単一の興味によるログ群が複数の興味として扱われてしまう問題がある。

本研究では, ユーザ意図の推移を知るために, ユーザアクションに注目したユーザ意図の推移候補を抽出する手法を提案する。ユーザアクションとは, 検索やブックマークからの閲覧, リンクをクリックして遷移することなどを指す。今回はユーザ意図が推移していない可能性が高いリンクのクリック, タブを切り替えるという行動に焦点を当てる。そして, ログを追跡し, そのログの始点 (閲覧の起点) をユーザ意図の推移候補とする。

本論文の構成は以下の通りである。2 節では関連研究と問題について述べる。3 節では提案手法と明らかになったことについて述べる。4 節では今後の課題について述べる。

2 関連研究

杉山ら [4] は, ユーザの興味には長期的な興味と 1 日限りの興味があると仮定している。1 日限りの興味を抽出するために, 固定的な時間間隔でログ群を区切る手法

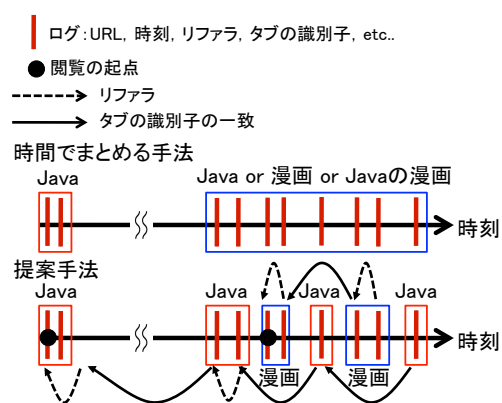


図1 時間ですまとめる手法と提案手法

を導入している。しかし, 区切られたログ群に複数の興味が含まれる場合, 複数の興味を単一の興味として扱われてしまう問題がある。

長野ら [5] は短期的な興味プロフィールの構築を目的とし, 各興味を反映したクラスタ群を形成している。平均 10 ページの閲覧ごとに興味変化が発生するという実験の結果 [6] に基づき, ページの内容が類似していたとしても両ページを閲覧した順序間隔が離れていれば, 同じ興味となる可能性が低くなると仮定している。しかし, 閲覧の順序間隔が離れたとしても, ある単一の興味が続いていることが考えられるため, 一様に同じ興味ではないと断定できない。

3 提案手法

3.1 提案手法の概要

我々はユーザがページに訪問する, もしくはタブを切り替えるという行動のログをブラウザ拡張によって取得している。ログには URL, 時刻, リンク元ページの情報であるリファラ, タブの識別子などが含まれている。

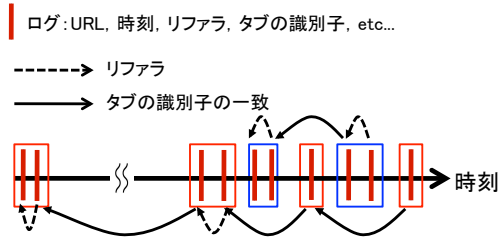


図2 閲覧の関連性によるログの連結

図1は、ユーザが「Java」と「漫画」という別々の興味に関して閲覧を行った場合の例である。2節で述べた手法では、図1の上部のように、複数の興味を区別できず、「Java」に関するログが繋がらなくなる可能性がある。本手法では、図1下部に示すように「Java」と「漫画」のそれぞれのログについて、ユーザ意図が同一である可能性が高いログを連結する。次に、連結されたログ群を辿りきることで「Java」「漫画」それぞれの閲覧の起点を探し出す。そして、「漫画」に関する閲覧の起点で、ユーザ意図が「Java」から「漫画」に推移したということを示す。

3.2 ユーザ意図の推移候補判定

ユーザ意図の推移候補を抽出するために、閲覧の起点と関連性に注目する。まず閲覧の関連性について述べる。あるログのリファラと、別のログのURLが一致する場合、もしくはログ間においてタブの識別子が一致する場合を閲覧の関連性があると判断した。前者の場合はログ間の内容が類似している、後者の場合は新しく開かれたタブを除き、すでに開いていたページであるため、これらはユーザ意図が推移していない可能性が高い。閲覧の関連性によるログの連結の概要を図2に示す。リファラの場合は点線矢印で、タブの識別子が一致している場合は実線矢印としてログを連結した。

次に閲覧の起点は、閲覧の関連性によって連結されたログ群において、閲覧の関連性で辿ることができなくなるログである。閲覧の起点が現れるポイントでユーザ意図の推移が起きている可能性が高い。したがって本研究では閲覧の起点をユーザ意図の推移候補とした。

3.3 プロトタイプ実装

図3は自身のログに対して提案手法を適用し、閲覧の関連性とページ遷移を可視化したものである。閲覧の関連性でログを連結して辿ることで、ある情報に関する閲覧がどれだけ続いていたかが明らかになった。さらに、ユーザの意図ごとにログを連結することができていたため、本手法は2節で取り上げた問題に対して効果的であるといえる。そして、閲覧の起点に注目するとユーザ意図の推移が起きていることが多く、図3上部においても閲覧の起点においてユーザ意図の推移が起きていた。こ

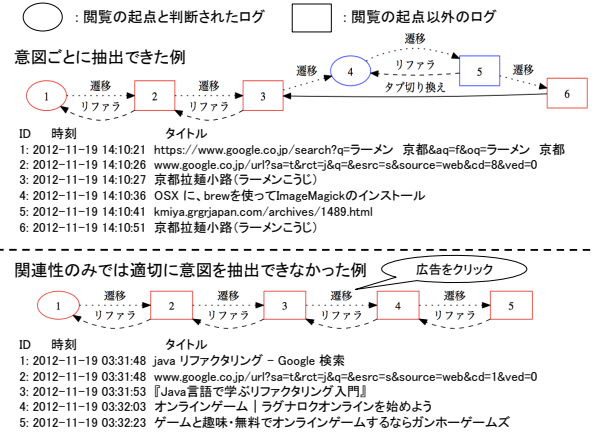


図3 提案手法によって連結されたログ群

のように限定的な分析の結果、本手法がユーザ意図の推移を捉えることに有用であることを確認した。

一方で、図3下部のようにページの内容ではなく広告に興味を持ちクリックをした場合、閲覧の関連性のみでは適切に意図を抽出することができないことがわかった。

4 今後

本手法ではブラウザ拡張で機械的に認識できるユーザアクションのみで関連性を抽出しているため、たとえば本文ではなく同一ページに掲載された広告に興味を引かれてクリックした場合には、実際にはユーザ意図の推移が発生しているはずだが、関連性で繋がっていると判断してしまっている。そのため今後は、ページの内容まで考慮したユーザ意図の推移候補の抽出に取り組む。

参考文献

- [1] 大槻一博, 服部元, 松本一則ほか: パーソナル・オンラインニュース配信システムの実証実験, 日本データベース学会論文誌, Vol.7, no.1, pp.43-48, 2008.
- [2] 土方嘉徳: 情報推薦・情報フィルタリングのためのユーザプロファイリング技術, 人工知能学会誌, 19(3), pp.365-372, 2004.
- [3] 丹羽智史, 土肥拓生, 本位田真一: Folksonomy マイニングに基づく Web ページ推薦システム, 情報処理学会論文誌 47(5), pp.1382-1392, 2006.
- [4] 杉山一成, 波多野賢治, 吉川正俊ほか: ユーザからの負担なく構築したプロファイルに基づく適応的 Web 情報検索, 電子情報通信学会論文誌 D-I, Vol.J87-D-I(11), pp.975-990, 2004.
- [5] 長野翔一, 市川裕介, 小林透: 短期的な興味プロファイル構築に向けたウェブ閲覧履歴のクラスタリング方式の提案, 電子情報通信学会論文誌 D-I, Vol.J95-D(4), pp.734-746, 2012.
- [6] 長野翔一, 高橋寛之, 中川哲也: ユーザの要求変化に着目したウェブ閲覧履歴の分類方式, 情報処理学会研究報告, 自然言語処理研究会報告 Vol.2008, no.90, pp.65-70, 2008.