

## ページ閲覧時間を考慮した Web アクセスパターンの抽出とその分析\*

三原宏一朗

東北大学工学部情報工学科<sup>†</sup>

寺邊正大

東北大学大学院情報科学研究科<sup>‡</sup>

橋本和夫

東北大学大学院情報科学研究科<sup>§</sup>

## 1 はじめに

近年、大規模化・複雑化が進む Web サイトの運営・管理を支援するため、ユーザのページリクエストを時系列順に記録したデータである Web アクセスログの利用が検討されてきた。Web アクセスログの解析はユーザのサイト利用状況を理解するための手がかりとなるため、サイト運営における有益性が報告されている [1]。

最近では Web アクセスログ解析にデータマイニング技術を応用する Web 利用マイニングの研究が進められている。Web 利用マイニングで用いられる手法には様々ある [2]。例えば、アイテム集合マイニングや順序パターンマイニング、グラフマイニングなどである。特にグラフマイニングは、ユーザがサイト内のページを遷移していく様子を、複数のページを同時に閲覧している状況も含めてパターンとして抽出することができる。タブ機能を有するブラウザの普及も始まり、今後ユーザが複数ページを同時に閲覧するような利用状況は増えると予想されるため、ユーザのサイト内での行動解析をより詳細に行うにはグラフマイニングが有用である。

一方、Web アクセスログ解析においては、ページ遷移だけでなくページ閲覧時間もユーザのサイト利用状況を理解する上で重要な要素とされる。同じページでもユーザ毎に、あるいは経路毎にそのページの持つ重要度・重みは違っており、その違いは閲覧時間から推定することができる。ここでいう重みとはあるページがユーザに有益な情報を含むかどうかを表し、セッション中の閲覧時間が長いほどそのページはユーザにとって重要なものであると考える。ところが、既存研究では閲覧時間によるページの重みの違いまで考慮したパターン抽出は報告されていない。

本稿では、サイト管理者が Web サイトの構成やコンテンツ、デザインの変更を行う際に有用な情報を提供することを目的とし、グラフマイニングを用いて、閲覧時間によるページの重みの違いを考慮した Web アクセスパターンを抽出する Web アクセスログ解析システムを提案する。

## 2 関連研究

## 2.1 Web 利用マイニング

Web 利用マイニングは Web ログマイニングとも呼ばれ、Web アクセスログなどのページリクエスト履歴に基づき、データマイニング技術を応用してユーザのサイトへのアクセスパターン解析や行動予測を行うものである。例えばあるサイトにおいて、ページ a を閲覧した後 b に移動し、c を別ウィンドウで開いた上で b から d に移動して、c と d を同時閲覧するという利用のされ方が多かった場合を考える。この場合、Web 利用マイニングで抽出される Web アクセスパターンは、大まかに表 1 のようにまとめられる。近年急増している EC サイトなどでは複数の商品ページを同時に開いて比較・検討するといった利用方法も考えられ、またタブ機能付ブラウザの普及も進みつつある。このため、今後ユーザのページ遷移に分岐遷移が含まれる状況が増加すると予想されることから、提案システムではこの分岐遷移をグラフ構造で抽出できるグラフマイニングを採用する。

## 2.2 ページ閲覧時間を考慮したパターン抽出

平手ら [3] は、順序パターンマイニングに関して、ページリクエストの時刻差により入口ページから各ページに到達するまでの時間を計算し、パターン抽出の際に考慮する手法を提案している。この手法は時間要素をページ間の距離として扱っており、マーケティングにおけるコンバージョンの達成度を評価するのに適している。一方本稿は、セッション中の各ページがユーザにとって重要かどうかという観点から、サイトのリンク構造やコンテンツの変更を行う際の参考となるようなパターンの抽出を行うことを目的としている。そこで本稿では、時間要素をページ間の距離ではなく、各ページの重みとして扱う。



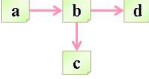
\* Mining and Analysis of Web Access Patterns Considering Browsing Time

<sup>†</sup> Koichiro Mihara, Department of Information Engineering, School of Engineering, TOHOKU University

<sup>‡</sup> Masahiro Terabe, Graduate School of Information Sciences, TOHOKU University

<sup>§</sup> Kazuo Hashimoto, Graduate School of Information Sciences, TOHOKU University

表 1 Web 利用マイニングで応用される手法の例

パターン	抽出手法	特徴
	アイテム集合マイニング	相関ルールとして扱うこともでき、各ページ間の関連性の理解に有効。
	順序パターンマイニング	リクエスト順序から得られるページ遷移のみを対象としているため、複数ページを同時に閲覧している場合などの分岐遷移は考慮できない。
	グラフマイニング	分岐遷移も考慮し、サイトのリンク構造に近い形でパターンが抽出できる。サイト内でよく利用される経路の特定がし易く、サイトのリンク構造の再構成などに有効。

### 3 提案システム

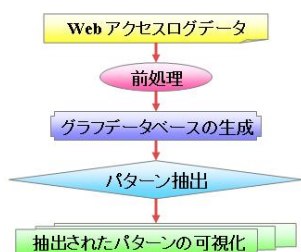


図 1 提案システムの処理の流れ

処理の流れは図 1 のようになる。前処理では、不要な情報を除去し、ユーザセッションの識別を行う [4]。グラフデータベース生成の際は、各ページの閲覧時間と重みを求め、セッション毎にグラフ構造に落とし込む。

あるページ P の閲覧時間は、P をリファアラ (Ref, あるページをリクエストしたときユーザが閲覧していたページ) として含むリクエストのリクエスト時刻 (req-time) と、P をリクエストページ (Req) として含むリクエストの req-time の差として求まる。分岐遷移の場合は、P を Ref として含むリクエストのうち、より新しいリクエストの req-time から差を取る。

重みの付加には重み付け関数を導入し、各ページに 0, 1, 2 のいずれかの重みを付ける。重み 0 は閲覧時間が非常に短く、ユーザにとってあまり有益でないと思われるページに付ける。重み 1 は閲覧時間が長く、ユーザに有益な情報が含まれると思われるページに付ける。重み 2 はセッション内で Ref として存在しないページ (離脱ペー

ジ) に付ける。離脱ページに関しては閲覧時間が計算できないため、このような扱いにする。また、重み付け関数はサイト管理者が任意に設定でき、重み 0, 1 の区別を行う際の基準を与えるものとする。

各セッションをグラフ化する際は、各節点をページと重みの組とし、同じページかつ同じ重みの場合は同一節点とする。また、辺は Ref から Req の向きにとり、同じ節点を同じ向きに結ぶリクエストが複数存在する場合は一本の辺にまとめて扱う。従って、得られるグラフデータベースは循環有向グラフデータを含むことになる。

こうして生成されたグラフデータベースに既存のグラフマイニングアルゴリズムを適用して頻出部分グラフを抽出し、結果を可視化して表示する。

2.1 節の例において、提案システムにより得られる結果は例えば図 2 のようになる。四角、二重四角、丸はそれぞれ重み 0, 1, 2 を意味する。提案システムでは、抽出されたパターンの一部に含まれるページとその部分グラフ構造が同じでも、重視されているページの違いによって異なるパターンとして抽出できる。そのため、ユーザが重視するページとページ遷移の変化を比較でき、より詳細なサイト利用状況の解析が可能となる。

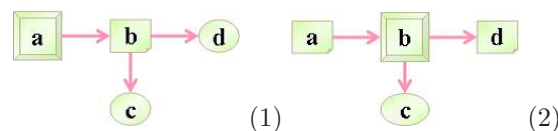


図 2 提案システムによって抽出されるパターン例

### 4 まとめ

本稿では、Web アクセスログ解析にページ閲覧時間に基づく各ページの重みを考慮したグラフマイニングを適用するシステムを提案した。今後は、このシステムを実装し、実際の Web アクセスログデータに適用する実験を通じ、抽出されるパターンや処理時間について閲覧時間を考慮しない場合との比較・検証を行う。

### 参考文献

- [1] 石井研二, “ホームページ アクセスログ解析の教科書”, 翔泳社 (2004).
- [2] R.Iváncsy and I.Vajk, “Frequent Pattern Mining in Web Log Data”, *Acta Polytechnica Hungarica, Journal of Applied Science at Budapest Tech Hungary, Special Issue on Computational Intelligence*, Vol.3, No.1, pp.77-90 (2006).
- [3] 平手勇宇, 山名早人. “時間情報を含むシーケンシャルパターンマイニングの一般化”, *DEWS2006*, (2006).
- [4] R.Cooley, B.Mobasher, and J.Srivastava, “Data Preparation for Mining World Wide Web Browsing Patterns”, *Knowledge and Information Systems*, Vol.1, No.1, pp.5-32 (1999).