平成 20 年度　電気情報工学科卒業研究概要

関係型データマイニングを利用した Web ページビュー構造のマイニング

犬塚研究室　—
No. 14117619　伊藤 弘和

1　はじめに

ユーザが Web を利用したときに生成されるデータ（アクセスログデータ）に対しマイニングをおこなうことを Web 利用マイニングという．Web 利用マイニングでは，ユーザがブラウザを通じて見る画面全体をページビューという単位で扱う．ページビューは一つの Web ページを表現する html ソース，画像，音楽ファイルなどのオブジェクトの集合である [1]．また，関係型データマイニングとは，一階述語論理を用いて事例の性質を表現し，性質の中で頻出なものを枚挙するマイニング手法である．本研究では，関係型データマイニングに適用できるようなページビューの構成を提案し，関係型データマイニングで扱うのに適した構造であるか検討する．

2　Web 利用構造マイニングのデータ表現

ページビューの表現法に求められることは，まず複雑な構造のページを表現できることである．ページビューには多重なフレーム構造が含まれることがあり，さらに各フレームに表示される html ソースも内部に構造をもつ．そのような構造を表せるデータ表現にする必要がある．次に，ユーザのページ遷移を正確に表現できなければならない．ページビューは html ソースに示された表現だけでは決まらない．なぜなら，フレーム構造によって与えられた複数のフレーム情報は，フレーム構造が解除されない限り部分的な変化しかもたらさず，影響を受けない部分が存在するからである．つまり，ページビューはユーザの行動の順を追うことによって正しく得られる動的な構造であることを意識したデータ表現にする必要がある．

3　ページビュー構成アルゴリズム

本研究では上記条件を満たすように，大きなページビューに小さなページビューが含まれるという入れ子構造を用いることにした．また，ユーザの行動を追うために，アクセスログを時間順に見るようにした．遷移は前のページビューを単位として構成される．ログデータをたどりページビューを再構成するアルゴリズムを表 1 に示す．このアルゴリズムを実装し，テストデータでは想定のページビューのデータ表現が得られることが確認できた．

4　実験と結果

提案したデータ表現を MAPIX という関係型データマイニングアルゴリズムによりマイニングした．その際に用いた背景知識を表 2 に示す．目標述語とは，着目したい対象を示し，経路述語は無意味な述語の連結を防ぎ，判定述語は経路述語をたどった結果として

表 1: ページビュー再構成アルゴリズム

Input: ログデータ
　　ログにアクセスして得る情報
Output: テーブル群

1. ログがなくなるまで **repeat**
2. ログを一つ読み込み，URL にアクセスして 必要な情報を取得
3. url，has_part_in_frame に情報を追加
4. 新規ページビューなら view に情報追加
5. 表示されなくなったページビューを DEAD にする
6. 親ページビューがあれば新たな親ページビューを view に追加
7. 古 親ビューの情報をすべてコピーして has_part_in_view に追加
8. 5〜7 を再帰的に実行

表 2: 背景知識として用意した述語

目標述語	説明
target	top ページビュー
経路述語	
has_part	ページビューの親子関係
linked_to	top 同士のつながり
判定述語	
terminal	シーケンスの最後の top
branch	分岐している
html	frameset 以外の html ソース
image	画像

目標述語 (または目標述語に連なる述語) の性質を表す．マイニングの結果例を以下に示す．

top(A), html(A).
top(A), has_part(A, B), html(B).
top(A), linked_to(A, B), terminal(B).

上段は top のページビューが html というもっとも簡単な構造である．中段は top のページビューがフレームセットであり，フレームは html ソースであることを示す．下段は top(A) がページビュー (B) に遷移したとき，そのページビューを見て Web 閲覧を終了したことを示す．

5　まとめ・今後の課題

本研究では，ページビューの構造について提案し，テスト用データにマイニングが実行できることを確認した．動的に生成されるページビューとその遷移をマイニングの対象とする枠組みを与えることができた．

今後の課題としては，実際の Web データをマイニングして，用意した構造が関係型データマイニングの視点からどの程度有用であるかの検証をおこなうことである．

参考文献

[1] B.mobasher. Web Usage Mining,in B.Liu(ed.) Web Data Mining, pp.449-483 ,Springer, 2007.