

Eleventh International Multi-Conference on Information Processing-2015 (IMCIP-2015)

Neuro-Fuzzy Based Hybrid Model for Web Usage Mining

G. Shivaprasad*, N. V. Subba Reddy, U. Dinesh Acharya and Prakash K. Aithal

Department of CSE, Manipal Institute of Technology, Manipal University, Manipal 576 104, India

Abstract

Web Usage mining consists of three main steps: Pre-processing, Knowledge Discovery and Pattern Analysis. The information gained from the analysis can then be used by the website administrators for efficient administration and personalization of their websites and thus the specific needs of specific communities of users can be fulfilled and profit can be increased. Also, Web Usage Mining uncovers the hidden patterns underlying the Web Log Data. These patterns represent user browsing behaviours which can be employed in detecting deviations in user browsing behaviour in web based banking and other applications where data privacy and security is of utmost importance. Proposed work pre-process, discovers and analyses the Web Log Data of Dr. T.M.A.PAI polytechnic website. A neuro-fuzzy based hybrid model is employed for Knowledge Discovery from web logs.

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of organizing committee of the Eleventh International Multi-Conference on Information Processing-2015 (IMCIP-2015)

Keywords: Feed forward neural network; Fuzzy C means clustering; User identification; User session identification; Web log data.

1. Introduction

The increasing number of web based applications has resulted in collection of massive amount of data in web server logs. This has resulted in Knowledge Discovery by the application of Web Usage Mining techniques. In Web Usage Mining browsing patterns of users are analyzed to extract useful information. Business communities make use of the discovered knowledge to increase the profit by personalizing the web sites for the customer thereby improved customer satisfaction. With the goal of understanding the user behaviors and preferences and thereby increasing the profit of the web based applications, Web Usage Mining discovers interesting usage patterns hidden in web data¹. Usually, Web Usage Mining consists of Pre-Processing, Knowledge Discovery and Pattern Analysis.

Web Server Logs serves as the input to the Web Usage Mining process. The Web Log Data is unstructured and noisy and ambiguous. In order to extract useful patterns from Web Logs, it must be preprocessed to remove noisy and irrelevant data, thereby reducing the bulk of data to be processed. Data mining algorithms can then be applied to the pre-processed web logs to extract useful patterns.

The uncertain and fuzzy characteristic of the browsing behavior of user is very difficult to model. Generally, a Web site attracts many different groups of users. For example, a college web site visitor group may include prospective students, parents, bankers, civil contractors, book publishers and book shop owners etc along with other anti social

*Corresponding author. Tel.: +91-0820-2924514.

E-mail address: shiva.prasad@manipal.edu

elements. Each group can have some specific need or goal. Further, a user within a group can have different intention during his visit to the web site. For example, a banker may visit a college web site and access links pertaining to admission dates so as to campaign his bank for educational loans. Same banker may visit the college web site to know about faculty details to campaign their Car or House Loan Scheme. Hard Clustering algorithms fail to capture such overlapping behaviors or interests of users as these algorithms push every object exclusively to a single cluster. Hence, fuzzy clustering algorithms are more suitable for Web Usage Mining domain. Hence, in the earlier work, the Fuzzy C-Means Clustering algorithm was employed for clustering the web user sessions.

In this work, a hybrid model based on neuro – fuzzy clustering is implemented to efficiently cluster the users of polytechnic website based on similar browsing patterns. The Web Log was preprocessed using Dimensionality reduction techniques and combined methodologies.

The rest of the paper is organized as follows. Section 2 presents the related work; Section 3 presents the proposed method; Section 4 deals with the experimental set up and result analysis and finally, Section 5 concludes the work.

2. Related Work

A. Bhargav, *et al.*² proposes a framework for Web Usage Mining consisting of Pre-processing, Pattern Discovery and Users classification. This framework classifies the users based on country, site entry and access time. M. A. Eltahir, *et al.*³ and Sanjay Kumar Malik, *et al.*⁴ explores and discusses Information extraction from user navigation history using Web Usage Mining. A detailed survey on data collection and pre-processing stage of web usage mining is discussed by Varnagar C. R.⁵. K. Sudheer Reddy, *et al.*⁶ proposes several data preparation techniques of access stream to identify the unique sessions and unique users. To analyze the learners' behaviour to help in learning evaluation and to enhance the structure of a given course, Educational data mining techniques are employed⁷. B. U. Maheswari, *et al.*⁸ proposes a new algorithm for pre-processing and clustering of web log.

Data Mining and Analysis techniques based on regular expressions on the data generated by University HTTP Server Logs has been proposed by Adamov A.⁹. M. Joshi, *et al.*¹⁰ implements and analyzes the performance of different soft – computing algorithms over an educational site. CLIQUE (CLUstering in QUEst) algorithm for clustering web sessions for web personalization has been adopted by K. Santhisree, *et al.*¹¹. S. Nadi, *et al.*¹² proposes a model for dynamic recommendation based on fuzzy clustering techniques, applicable to currently on-line users. The fuzzy clustering approach, in this study, provides the possibility of capturing the uncertainty among Web user's behaviours. A fuzzy clustering around medoids approach is adopted by Pierpaolo D'Urso *et al.*¹³ to classify ordered sequences (paths) representing patterns of individual behaviour in an actual or virtual space – time domain. Z. Ansari, *et al.*¹⁴ employs a Fuzzy Membership Function to assign weights to sessions based on the number of URLs accessed by the sessions followed by application of Fuzzy c-Mean Clustering algorithm to discover the clusters of user profiles. A neuro – fuzzy model for data clustering is introduced by Farhat Roohi¹⁵.

3. Methodology

3.1 Web log data collection

Each user request to the server will be recorded in a web server log. Each line in a log represents the request made to the server and usually in Extended Log Format contains the following fields:

- i. Host/The Remote IP address: Identifies who had visited the web site.
- ii. User Authentication: Username and password if the server requires user authentication (generally “-”).
- iii. Date and time of the request: Used to determine how long a visitor has spent on a given page.
- iv. The HTTP Request: The Method (GET, POST, HEAD, etc.) used for information transfer is noted along with the Requested Resource Name (an HTML page, an image file, or a script) and Protocol Version (HTTP protocol being used).
- v. The Request Status: HTTP status code returned to the client (200, 404 etc.).
- vi. The Page size: Content – length of the document transferred.

Download English Version:

<https://daneshyari.com/en/article/487465>

Download Persian Version:

<https://daneshyari.com/article/487465>

[Daneshyari.com](https://daneshyari.com)