IJETST- Volume||01||Issue||05||Pages 665-670||July||ISSN 2348-9480 2014

Open access Journal International journal of Emerging Trends in Science and Technology

Personalized Recommendations in e-Learning System using Web Mining

Authors

Mr. Namdev Anwat[1], Mrs. Varsha Patil[2]

[1]Department of Computer Engineering,
Matoshri College of Engineering and Research Center, Nashik
Near odha gaon, Eklahare, Nashik, Maharashtra - 422105, India
University of Pune, Pune
Email: *namdev.anwat@gmail.com*
[2]Department of Computer Engineering,
Matoshri College of Engineering and Research Center, Nashik
Near odha gaon, Eklahare, Nashik, Maharashtra - 422105, India
University of Pune, Pune
Email: *varshapatil@gmail.com*

Abstract
Personalization is a subclass of information filtering system that seek to predict the 'ratings' or 'preferences' that a user would give to an items, they had not yet considered, using a model built from the characteristics of an item (content-based approaches or collaborative filtering approaches). Web mining is an emerging field of data mining used to provide personalization on the web. It consist three major categories i.e. Web Content Mining, Web Usage Mining, and Web Structure Mining. In this paper, a simplified Personalized Recommendation System using web usage mining and content mining is proposed to provide personalized search results to learners. The web usage mining and content mining technologies are used for system implementation aim to identify personalized recommendations. Basically, It consist four major steps to construct personalized recommendations are Data Collection, Data Processing, Data Analysis, and Output Generation. The proposed architecture in this paper is implemented using .Net technology to achieve system goals.
Keywords: *Recommendations, Web Usage Mining, Association Rule Mining, Clustering, Lingo..*

1. Introduction

Personalization means persons would get the things or results according to their interests and expectations without giving much more input. Personalization systems are a subclass of information filtering system that seek to predict the 'ratings' or 'preferences' that a user would give to an items, they had not yet considered, using a model built from the characteristics of an item (content-based approaches or collaborative filtering approaches). Personalization systems analyzes the individual characteristics and habits without expecting much more input from user and constructs an automated responses to fulfill individual needs. Personalization systems have become very common in recent years. These systems are more flexible, reliable, and dynamic to provide personalized results [1].

Web mining is an emerging field of data mining that automatically extracts useful information and patterns from web documents and constructs personalized results. Web mining techniques are commonly used by various kinds of organizations to extract useful information and patterns on the basis of interest and habits of customers/users. This extracted information is used to promote business, understanding market dynamics, new promotions floating on the internet, personalized advertisements etc [2].

So, in this paper our motivation is to provide personalized recommendations in e-learning system using web mining technology. The rest of the paper is organized as follows: Chapter (2) introduces the related work done on personalization by different researchers using web mining. Chapter (3) briefly introduces web mining taxonomy and proposed recommendation system. Chapter (4) gives system implementation and algorithmic details and Finally, Chapter (5) concludes the proposed system and lists future scope.

2. Related Work

Now a day, Personalization systems becoming more popular because of it analyzes individual differences of each user and provides personalized responses according to each individual's needs, interests and preferences. It performs modifications concerning the contents or even the structure of web site dynamically. So, number of researchers has done much on in this area. Some related interesting research efforts are discussed below.

Nasraoui and Krishnapuram et al. [2000] discovered the user session files and formulated groups on the basis of similar characteristics using fuzzy algorithms [3]. According to their research a user or a page can have more than one cluster. In their proposed approach, after preprocessing of usage data dissimilarity matrix of preprocessed data is

created. This is used by fuzzy algorithms in order to cluster typical user session.

Mobasher et. al. [2000] proposed most advanced system, "WebPersonalizer" [4]. It is a powerful framework for mining web log files to extract the useful information for the purpose of recommendations based on the browsing similarities of current user to previous user. After collecting and cleaning of usage data (creating various abstractions of collected data), data mining techniques such as association rule mining, sequential pattern discovery, clustering, and classification are applied in order to discover interesting usage patterns.

The most important contribution of Berendt [2001] in the area of web usage mining is STRATDYN (Strategic and Dynamic) [5] add-on module. It determines the differences between navigational patterns of user and then it exploits the site semantics in the visualization of the results. In this approach, web pages are grouped together on the basis of concept hierarchies. He focused on "interval based coarsening" technique for usage data at different levels of abstraction. For this purpose he used basic and coarsened stratograms for visualization of the results.

Magdalini Eirinaki et. al. [2003] focused on web usage mining. This process relies on the application of statistical and data mining methods to the web log data, resulting in a set of useful patterns that indicate user's navigational behavior [1]. In this approach various data mining algorithms are applied to find navigational patterns.

Yuewu Dong et. al.[2010], presented a simplified architecture of distance education system based on web usage mining and content mining to realize personalization [6]. They have constructed a basic user interface based on database of BUPT-SK and external data mining tools to realize personalized environment in distance education system.

All above researchers have mostly focused on web usage mining and user profiling for personalized recommendations using search queries and their approach is to provide general personalized environment for all kinds of web users. In the research of Yuewu Dong et. al., they have created a basic learning environment using existing web mining tools on experimental basis. So, this paper focuses on implementation of personalized recommendation system using web mining strategies and algorithms to provide personalized recommendations.

3. Personalized Recommendation System

In this article we are interested in web mining that uses web data sources in order to discover hidden knowledge about users and their behavior on the Web. Such knowledge, if taken advantage of, brings to organization nothing than benefits and leads directly to profit increase. Site modification, business intelligence, system improvement, personalization and usage characterization are the areas in which the potentials of web usage mining and content mining have been recognized and extensively used. Following section gives an overview of web mining technology.

3.1. Web Mining Technology

Web mining technology is emerging field of data mining for WWW based information and resources. The basic focus of web mining is to use data mining techniques and algorithms to extract useful and hidden patterns from unstructured and huge web data or resources. Web mining taxonomy is divided into three categories according to sources of web data [2]. These categories are Web content mining, Web usage mining, and Web structure mining shown in Figure 1. As you can see from this Figure;

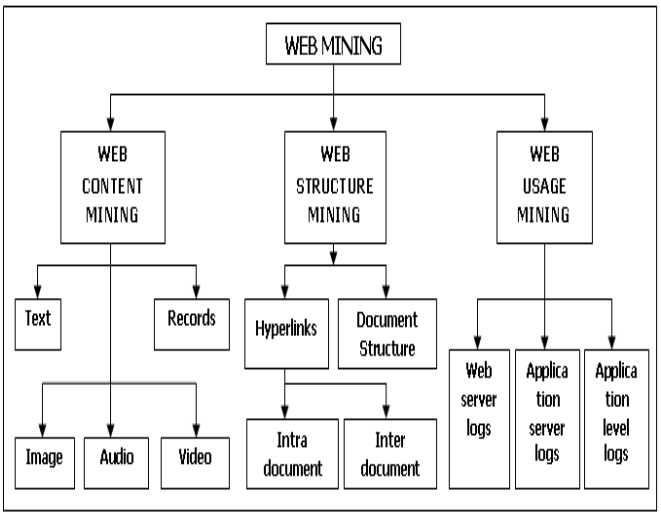

Figure 1: Web mining overview.

(i) **Web content mining** means the extraction of useful information and web knowledge from web sources or web contents such as text, image, audio, video, and structured records [2].

(ii) **Web usage mining** is the application of data mining techniques to find out interesting patterns from web usage data. It mainly tries to extract useful and interesting patterns from usage data such as server logs, client browser logs, proxy server logs, cookies, user sessions, registration data, mouse clicks, user queries, bookmarks etc. and any other data as the results of user interactions [6].

(iii) **Web structure mining** tries to identify the structure of hyperlink in html documents and deduce knowledge [2].

Personalization is a typical application of Web Mining, which can be used to improve web site usage by customizing the contents of a web site with respect to the visitor's need. The personalized web content can take the form of recommended links or items, targeted advertisements, or adjusted text and graphics. The purpose is to customize the interactions on a web site depending on the user's explicit and/or implicit interests, habits, and desires. Web mining technology helps service providers to improve their services by gaining general knowledge about the different user groups. More importantly it can help to provide personalized interfaces and services according to individual characteristics of each user [1]. Figure 2 shows typical web mining process for web personalization.

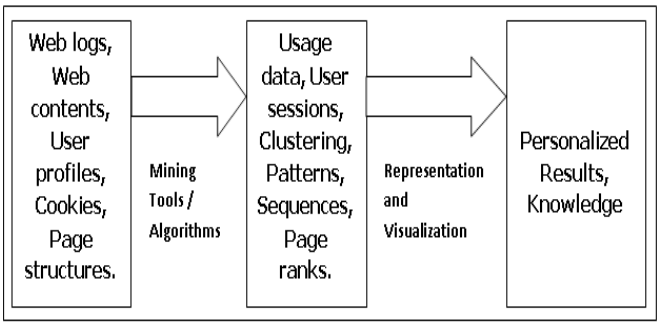

Figure 2: Typical web mining process.

3.2. Recommendation System Overview

Web based learning systems are generally designed to provide interactive interface to access the performance of students, to deliver course materials and other contents, to interact with students, track the academic records of the students, and to provide administrative functions such as management of student profiles [6]. In this paper our focus is to provide personalized recommendations with user functions by analyzing user interests, preferences and habits using web mining technology.

3.2.1. Framework of Proposed System

Figure 3 provides a simplified framework of Personalized Recommendation System. It consists of learners, their interaction with system, and different web mining algorithms to fulfill their needs. The web mining technology used in this context focuses on following aspects of the system.

i. It provides a personalized environment to the learners. It recommends adaptive results to learners according to their individual characteristics such as knowledge structure, learning objectives, learning style and preferences.

Figure 3: Framework of proposed system.

The main focus of this paper is to provide services according to above mentioned points. This paper focuses on how to provide learners with personalized recommendations.

3.2.2. Modules of Proposed System

Personalized recommendation system provides online learning recommendations with the support of web usage mining and web content mining to provide flexible learning platform for learners. The modules of proposed system are shown in Figure 4. The steps of recommendation process consist of following modules: (1) Data collection (2) Data processing (3) Data analysis (4) Output Generation.

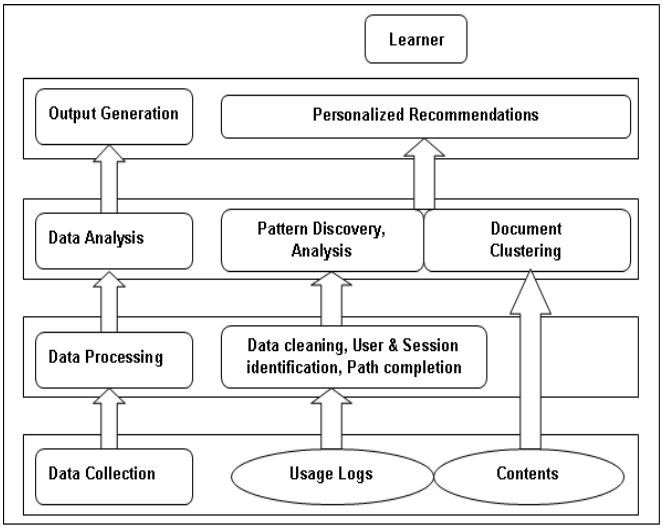

Figure 4: Modules of proposed system.

a) Data Collection

This module collects data from log files and stored contents. It collects two types of data i.e. contents and usage logs. Contents are nothing but real academic data. Usage log means access pattern of web pages. Content data can be collected from coursewares or notes. Usage logs are collected from server logs or browser logs.

b) Data Processing

The input to data processing is usage data collected by collection module. Usage data such as web log files are preprocessed for data cleaning, user and session identification. Then extracted information can be transferred into normalized form by applying web mining [10]. In this module web usage mining process is applied to identify user's navigational patterns. Figure 5 shows the web usage mining process of proposed system.

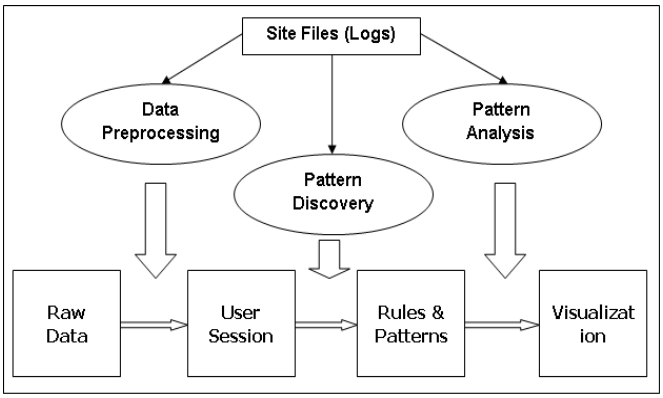

Figure 5: Web usage mining process.

The most important task of the web usage mining process is data preparation. Data preparation task includes the major steps such as data cleaning, user identification, session identification, path completion, and data summarization [10]. Figure 6 shows the complete process of data preprocessing of proposed system. The data cleaning process involves removing the irrelevant data from the database log. This data can be in the form of requests from a non-analyzed source, data with missing attributes or the attributes that are not needed for the system goal. Next step is user identification where it identifies all unique users of the system from data logs by following certain heuristics. In session identification,

it defines the number of times the user has accessed a web page. Here we use time out mechanism to identify the access time of the user for a respective web page. Path completion process makes certain, where the request came from and what all pages are involved in the path from the start till the end. The referrer plays an important role in determining the path for a particular request. Data summarization is one of the advanced data preprocessing tasks that are performed after all of the above processes. In this process, the data is inserted into a relational database system for further generalization and computations [10].

c) Data Analysis

Data analysis module uses the different web mining strategies and data mining algorithms for in depth examination of web information in order to identify useful patterns and sequences [7]. The input to this module is cleaned log records for particular user's session or classified web contents. Here data mining algorithms are applied for pattern discovery and clustering. In proposed system, Web usage mining is applied on usage data in order to extract useful patterns concerning the learner's navigational behavior. Web content mining is applied for classifying the contents in semantic categories in order to make information retrieval.

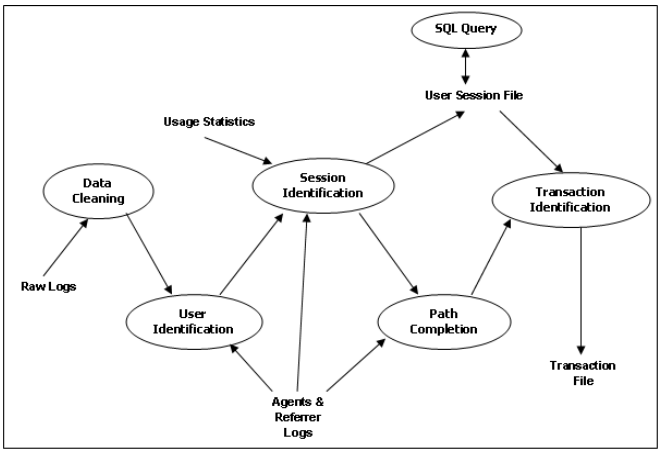

Figure 6: Data pre-processing task.

d) Output Generation

Based on the results of data analysis phase, personalized information is generated and appropriate personalized results will be provided to the user. Thus, personalized recommendations are constructed and provided to the learners

4. System Implementation and Results

This section of paper focuses on implementation algorithms and results of proposed system. Personalized recommendations are constructed using web mining strategies and algorithms. The techniques of data mining such as association rule mining is applied for pattern discovery and Lingo clustering is applied on learning contents to realize personalized recommendations.

4.1. Algorithmic Details

Association rules technique is generally applied to a database of transactions consisting of a set of items. This rule implies some kind of association between the transactions in the database. It is important to discover the associations and correlations between these set of transactions [8]. In the web data set, the transaction consists of the number of URL visits by the client, to the web site. It is very important to define the parameter support, while performing the association rule technique on the transactions. This helps in reducing the unnecessary transactions from the database. Support defines the number of occurrences of user transactions within the transaction log. The discovery of such rules from the access log can be of tremendous help in reorganizing the structure of the web site. The frequently accessed web pages should be organized in their order of importance and be easily accessible to the users.

The clustering and discovery rules allow grouping the items with similar attributes together. Therefore, when new data is added to the database, it can be clustered on the basis of its attributes [12]. In the web transaction data set, the clustering can result in forming learners with similar interests or learners that visit the specific web page based on their demographic information and access patterns. The clustering of learners into specific groups can help in forming more personalized recommendations for the future.

4.1.1. Pattern Discovery Algorithm

Association rules are a data mining technique that searches for relationships between attributes in large data sets. The best known algorithms for mining association rules are Apriori, AprioriTID, STEM, DIC, Partition Algorithm, Elcat, FP-grow, etc. In proposed system we use Apriori algorithm to find useful navigational patterns.

4.1.2. Clustering Algorithm

The general idea behind LINGO is to first find meaningful descriptions of clusters, and then, based on the descriptions, determine their content. To assign documents to the already labeled groups LINGO could use the Latent Semantic Indexing in the setting for which it was originally designed for given a query, retrieves the best matching documents. When a cluster label is fed into the LSI as a query, as a result contents of the cluster will be returned. This approach should take advantage of the LSI's ability to capture high-order semantic dependencies in the input collection.

4.2. Results and Discussion

To realize personalization and to provide personalized recommendations proposed system in this paper is implemented using .net technology. Personalized recommendation system is based on web mining strategies and data mining algorithms. To evaluate a results and performance of system following performance measures are considered.

4.2.1. Performance Measures

Typically the measures require a collection of documents and a query, whereas every document is either relevant or non relevant to a particular query. For evaluating proposed system we have used some performance measures such as **accuracy, precision, recall, f-measure, and fall-out** in the context of relevant documents and irrelevant documents [12]. Using this we have evaluated the performance of system with respect to accuracy of clusters, overall performance of system. Also, we have compared existing learning system personalization such as customized strategy with our

proposed system. Following are the some bar charts which show the evaluation results of proposed system:

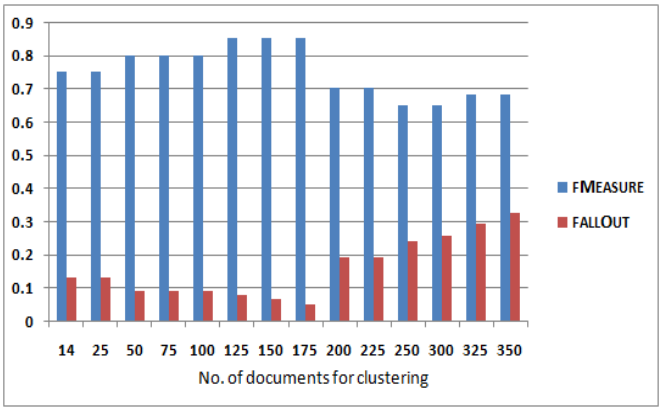

Figure 7: Cluster analysis.

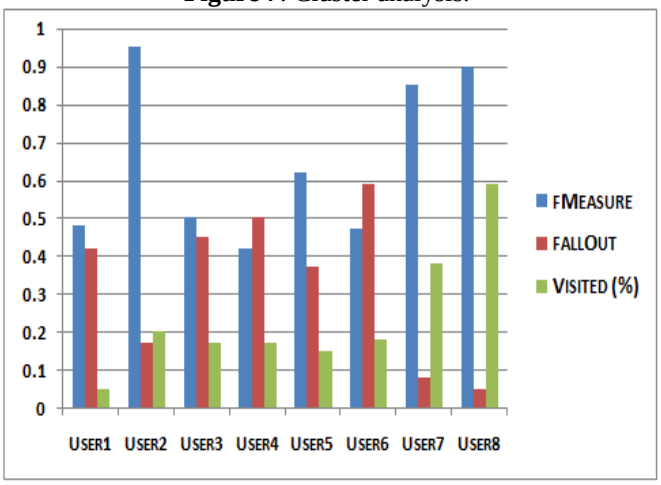

Figure 8: Overall system performance.

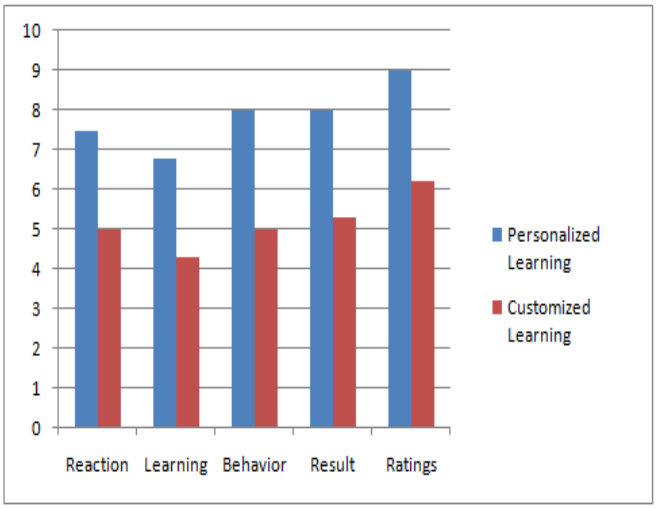

Figure 9: Proposed system vs. customized system.

On the basis of above evaluations we have concluded that, data mining techniques are applied to extract statistical information and discover usage patterns, cluster the users into groups: (1) Analyzing the relationship of learners' interests with particular clusters using clustering method; (2) According to pattern discovery algorithm, highly accessed items are given to the user; (3) Finding out the sequence relationship in their learning process by using sequence pattern mining, and suggestions about the next learning step will be given also on the basis of their learning status.

5. Conclusion

In this paper, a simplified architecture of personalized recommendation system using web mining is implemented to provide simple, flexible, and modularized recommendations to users of learning system. For conceptual clarity this paper focuses on process of personalization using mining algorithms. With the help of SRS, design and modeling concepts personalized recommendation system is implemented using .net technology. Based on the results of data collection, pretreatment, and analysis phases, personalized recommendations are constructed to realize personalization. This environment has characteristics such as Personalized User Interface, Personalized Recommendations, and Optimized course structure etc.

Our future work focuses on, (a) Mining of courses on the basis of learners records in order to make categorized design (b) Implementation of Recommendation Agents using web structure mining to give dynamic interest links from external search engines.

References

[1] Magdalini Eirinaki and Michalis Vazirgiannis, "Web mining for web personalization", ACM Transactions on Internet Technology, 03(01):1-27, February 2003.

[2] Raymond Kosala and Hendrik Blockeel, "Web mining research: A survey", SIGKDD Explorations, pages 95-104, July 2000.

[3] Nasraoui O., Frigui H., Krishnapuram R., and Joshi A, "Extracting web user profiles using relational competitive fuzzy clustering", IJAI Knowledge Discovery, 09(04):8-14, April 2000.

[4] Mobasher B., Cooley R., and Srivastava J, "Automatic personalization based on web usage mining", ACM Communication, 43(08):142-151, August 2000.

[5] Berendt B, "Understanding web usage at different levels of abstraction: Coarsening and visualizing sequences", ACM SIGKDD Knowledge discovery & Data mining, 04(07):104-108, August 2001.

[6] Yuewu Dong and Jiangtao Li., "Personalized distance education system based on web mining", IEEE Education and Information Technology, 02(05):187-191, August 2010.

[7] Michael Azmy, "Web content mining research: A survey", ACM SIGMOD Explorations, 01(01):203-212, November 2005.

[8] Jaideep Srivastava, Robert Cooley, Mukund Deshpande, and Pangning Tan, "Web usage mining: Discovery and applications of usage patterns from web data", ACM SIGKDD Explorations, 01(03):187-192, January 2000.

[9] Robert Cooley, Bamshad Mobasher, and Jaideep Srivastava, "Data preparation for mining world wide web browsing patterns", Knowledge and Information Systems, 01(01):84-89, February 1999.

[10] Sasa Bosnjak, Mirjana Maric, Zita Bosnjak, "The role of web usage mining in web application evaluation", Management Information Systems, Vol. 05(01):31-36, 2010.

[11] Borges J. and Levene M, "Data mining of user navigation patterns", Springer-Verlag, 1836(08):92-111, April 1999.

[12] Maurice D. Mulvenna, Sarabjot S Anand, and Alex G. Buchner, "Personalization on the net using web mining", ACM Communication, 43(08):122-128, August 2000.

[13] Margaret H. Dunham, "Data Mining: Introductory and Advanced Topics", Pearson, 01 edition, April 2006.

[14] Jiawei Han, Micheline Kamber, and Jian Pei, "Data Mining: Concepts and Techniques", Morgan Kaufmann, 03 edition, May 2007.

Author Profile

Namdev Anwat received the B.E. degree in Information Technology from Sanjivani Rural Education Society's College of Engineering in 2007. He is pursuing his M.E. from Matoshri College of Engineering and Research Center, Nashik. He now works with Government Polytechnic Nashik as a visiting faculty.