AUSTRALIAN JOURNAL OF BASIC AND APPLIED SCIENCES

ISSN:1991-8178 EISSN: 2309-8414
Journal home page: www.ajbasweb.com

Role Of Agent Technology In Web Usage Mining: Prediction Of Popularity Of Web pages Using JADE

[1]V.P. Kannan and [2]Dr. K. L. Shunmuganathan

[1]*Research Scholar, Department of Computer Science and Engineering, Bharathiyar University, Coimbatore, India*
[2]*Professor, Department of Computer Science and Engineering, R.M.K Engineering college, Chennai, India*

Address For Correspondence:
V.P. Kannan, Research Scholar, Department of Computer Science and Engineering, Bharathiyar University, Coimbatore, India
E-mail: vpkannan6@yahoo.com, klsnathan@gmail.com

ARTICLE INFO

Article history:
Received 28 May 2016
Accepted 29 July 2016
Published 17 August 2016

Keywords:
Web Usage Mining, Agent, pattern discovery, pre processing.

ABSTRACT

Recently, keen interest has been shown in studying the goal behind a user's Web query so that this goal can be used to improve the quality of the results of a search engine that, in turn, improves the popularity of web pages. Advertisement fees can be decided based on this factor. Metering process must be efficient and secure against the fraudulent attempts by servers. Web Usage Mining (WUM) can be used for successful organization of Web sites, generation of adaptive Web sites, generation of business, maintenance of services, personalization, and examination of network traffic flow and so on. The WUM consists of three steps, viz., pre-processing, and pattern discovery and pattern analysis. WUM has become an active area of research in field of data mining due to its vital importance. This work elaborates web usage mining process and highlights the research challenges. Here we use FIPA-OS (Foundation for Intelligent Physical Agents) toolkit enabling rapid development interacting agent based system in the JADE (Java Agent Development Environment) platform. It simplifies the implementation of multi-agent systems implementations are made easy using a middle-ware that satisfies the FIPA specifications.This paper further highlights agent technology and explores applicability of agents in web usage mining.

INTRODUCTION

Today World Wide Web has become a popular medium to search for information. Whenever some information is desired from the web, we come across a huge amount of information, out of which sorting the relevant information is left for the user. This situation is termed as information overload which leads to difficulty in finding relevant information.

Web mining is the extraction of exciting and constructive facts and inherent information from artefacts or actions related to the WWW. Web Mining is used as a tool to remove the problem of information overload while searching information over the WWW. Various organizations are also employing the web mining process in order to make their website more user-friendly and to improve the web surfing experience of users. Internet is a pool of diverse information sources such as text, images, hyperlinks, audio and video (Maruthu Pandi J, *et al*., 2016). Depending upon the type of data being mined, web mining (Dong, D., 2009; Federico Michele Facca and Pier Luca Lanzi, 2003) is categorized as

A. Web content mining:

We discover useful information from the contents of web site which may include text, hyperlinks, metadata, images, videos, and audios. Search engines and web spiders are used to gather data for content mining.

To Cite This Article: V.P. Kannan and Dr. K. L. Shunmuganathan., Role Of Agent Technology In Web Usage Mining: Prediction Of Popularity Of Web pages Using JADE. ***Aust. J. Basic & Appl. Sci.,*** *10(12): 392-406, 2016*

B. Web structure mining:

It is the application of mining techniques on the hyperlink structure of the web. It is useful in measuring the ranking of a web page and the information about a page's ranking; and it provides the relationship and similarity between different websites.

C. Web usage mining or web log mining:

Users' behaviour or interests are revealed by applying data mining techniques on web log file. It generates the secondary data such as user interests and behaviour as a result of these interactions.

This work explores Web Usage Mining in detail and throws light on its various aspects. Web usage mining system should be able to gather useful usage data thoroughly, filter out irrelevant usage data, establish the actual usage data, discover interesting navigation patterns, display the navigation patterns clearly, analyse and interpret the navigation patterns correctly, and apply the mining results effectively (Jayasowbaghyam, E. and K. Nithya, 2016).

Main Motivation - Metering the Popularity of Web Sites: There are variety of approaches for making money on the Internet. Advertising is currently the main source for revenues, and the current figures estimate that advertising on the Internet will be a multi-billion dollar industry in the year 2000. For advertising to be effective, the advertisers must have a way to measure the exposure of their advertisements. This measurement considerably influences the fees demanded for advertising, and is common to all forms of traditional advertising (like TV or newspapers). Measurement of popularity of web services should be impartial, accurate and efficient.

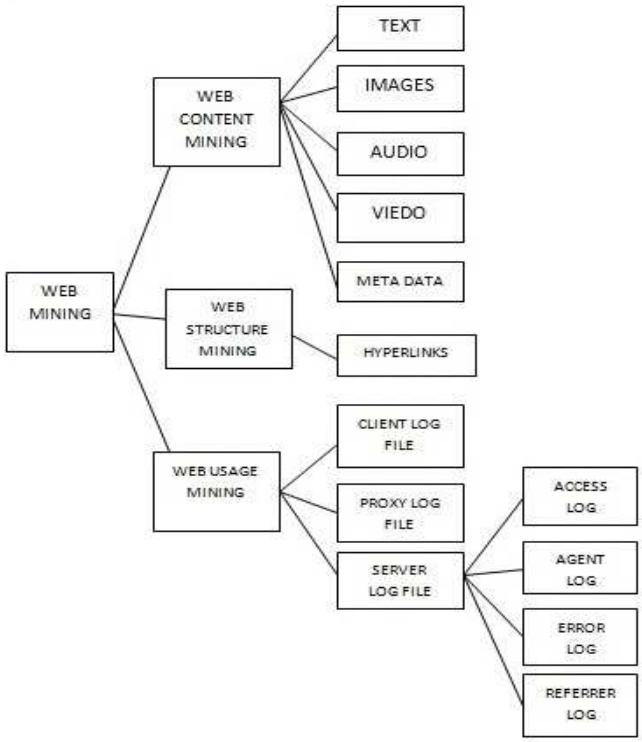

Fig. 1: Classification of Web Mining Based On Mined Data

Overview Of Web Usage Mining:

Accurate Web usage data could assist to draw the attention of new customers, maintain present customers, enhance cross marketing / sales, assist in devising effective promotional campaigns, track leaving customers and identify efficient logical structure for their web space. User profiles could be constructed by merging user's navigation paths with other data characteristics like page viewing time, hyperlink structure, and page content. Conversely, as the size and complexity of the data escalates, the statistics suggested by conventional web log examination techniques may prove insufficient and highly intelligent. Then, mining methods will be required to assist researchers in developing a better strategy for Web usage mining.

Web Usage Mining is a division of Web Mining, which is sequentially a component of Data Mining. Web Usage Mining is basically used to study the user's browsing behaviour to find out the areas of interests and nature of contents in which users are more interested. It analyzes the data such as websites or the links most visited by the user and the topics of interest of user, etc. The results obtained after such analysis are helpful in

many ways such as optimizing future searches and restructuring of websites (Genesereth, M.R. and S.P. Ketchpel, 1994) to provide better experience to the users. These results are also helpful in making page recommendations for the users. Data collected in web server logs is not directly suitable for applying web mining techniques, it has to be pre-processed to make it suitable for mining.

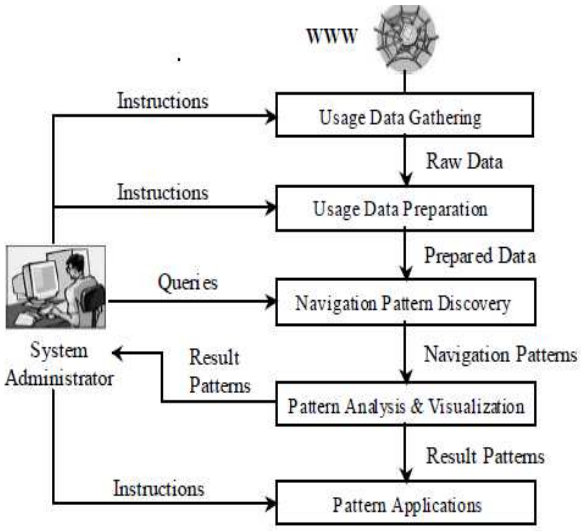

Fig. 2: Web Usage Mining System Structure

There are four stages in web usage mining.

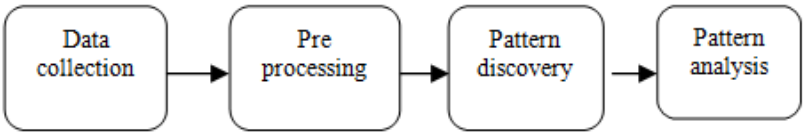

Fig. 3: Stages involved with web usage mining

A. Data Collection:

Data Collection is the first step in web usage mining process. It consists of gathering the relevant web data. Data source can be collected at the server-side, client-side, proxy servers, or obtained from an organization's database which contains business data or consolidated Web data.

1) Server Level Collection:

It collects client requests stored in the server as web logs. Web server logs are plain text that is independent from server platform. Most of the web servers follow common log format as

"ip-address username password date/timestamp URL version status-codebytes-sent"

Some servers follow Extended log format along with referrer and user agent. Referrer is the referring link URL and user agent is the string describing the type and version of browser software used. Web cache and the IP address misinterpretation are the two drawbacks in the server log. Web cache keeps track of web pages that requests and saves a copy of these pages for a certain period. If there is a request for same page, the cache page is in use instead of making new request to the server. Therefore, these requests are not recorded into the log files.

2) Cookies:

Unique ID is generated by the web server for individual client browsers and it automatically tracks the site visitors. When the user visits next time, the request is sent back to the web server along with ID. However, if the user wishes for privacy and security, they can disable the browser option for accepting cookies.

3) Explicit User Input:

Data is collected through registration forms and provides important personal and demographic information and preferences. However, this data is not reliable.

4) Proxy level collection:

Data is collected from intermediate server between browsers and web servers. Proxy caching is used to reduce the loading time of a Web page experienced by users as well as the network traffic load at the server and client sides. Access log from proxy servers are of same format as web server log and it records the web page request and response for the server. Proxy traces may reveal the actual HTTP requests from multiple clients to multiple Web servers. This may serve as a data source for characterizing the browsing behaviour of a group of anonymous users sharing a common proxy server.

B. Pre-processing:

Performs a series of processing of web log file covering data cleaning, user identification, session identification, path completion and transaction identification. The information available in the Web is heterogeneous and unstructured. Therefore, data pre-processing is predominantly significant phase in WUM (Fürnkranz, J., 2005). The goal of pre-processing is to transform the collected raw data into a set of user profiles.

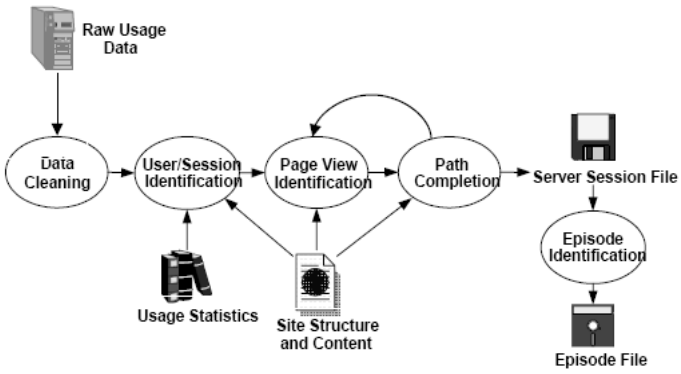

Fig. 4: Pre-processing

1) Data Cleaning:

- remove irrelevant references and fields in server logs (Federico Michele Facca and Pier Luca Lanzi, 2003; Kolari, P. and A. Joshi, 2004)
- remove references due to spider navigation
- remove erroneous references
- add missing references due to caching

2) User Identification:

User Identification means identifying Unique users considering their IP address. Following heuristics are used to identify unique users:

(1) If there is a new IP address, then there is a new user.

(2) For more logs, if the IP address is the same, but the operating system or browsing software is different, a reasonable assumption is that each different agent type for an IP address represents a different user (Griss, M., 2001).

Different IP values in the IP address field represent different users:

- Even if the IP values are same, but agent field values are not same, then they represent different users.
- If the IP values as well as the agent field values are same, then the data surfed by the user is taken into account. If there is any link between the page last requested and earlier requested pages then they represent same users otherwise different. If there is no link between data accessed or there is large interval between the accessing times, then it is identified as a new user session.
- Same user may visit web more than once at different point of time, so a time heuristic (Zaïane, O.R., 2001) is used to divide those intervals into different user sessions.

3) Session identification:

A threshold value of accessing time is setup (say 30 Minutes). When the access time of the same user exceeds 30 minutes then it is considered as different user session.

- Various sessions are then labelled with session IDs and values are assigned to them.

- To calculate the time difference, four fields i.e. day, time hh, time mm and times indicating days, hours, minutes and seconds are added to Page view.

4) Page view :

It is an aggregate representation of a collection of Web objects contributing to the display on a user's browser resulting from a single user action (such as a click-through). Conceptually, each page view can be viewed as a collection of Web objects or resources representing a specific "user event," e.g., reading an article, viewing a product page, or adding a product to the shopping cart.

5) Path completion:

Many of important page accesses are missed in the Web log file due to the existence of local cache and proxy server. The task of path completion is to fill in these missing page references and makes certain, where the request came from and what all pages are involved in the path from the start till the end. So, missing pages are added as follows: The page request is checked whether it is directly linked to the last page or not. If there is no link with last page check the recent history. If the log record is available in recent history then it is clear that "back" button is used for caching until the page has been reached. If the referrer log is not clear, the site topology can be used for the same effect. If many pages are linked to the requested page, the closest page is the source of new request and so that page is added to the session. There are three approaches in this regard.

- Reference Length approach
- Maximal Forward Reference
- Time Window

C. Pattern Discovery:

Once the user and sessions are identified, various data mining techniques (Jennings, N.R. and M. Wooldridge, 1996; Li, Y. and N. Zhong, 2004) such as clustering, classification & association rules are applied to find out the various sets of matching patterns

1) Classification:

Classification involves grouping the data into several pre-defined classes. The data is segregated by matching with the best described features or properties of a class or category.

2) Clustering:

Clustering means to group similar patterns together. Clusters can be usage clusters and page clusters. Clustering of users establishes a set of users with similar browsing patterns. Classes in clustering are not predefined rather they emerge as a result of clustering. Clustering of pages highlights and clubs pages with similar or related content.

3) Association rules:

Association rules discover those pages which are associated or related with each other in some way. The pages which are accessed together are discovered using association rules. For example, a person searching for journals in any field may also be interested in conferences, seminars and workshops in the same domain.

D. Pattern Analysis:

Pattern analysis (Jennings, N.R. and M. Wooldridge, 1996) is the last step in the overall WUM process that has two fundamental goals. The first goal is to extract the interesting rules, patterns or statistics from the output of the pattern discovery process by filtering the irrelative rules or statistics . Another aim of this analysis is to obtain some information and offer valuable insights about users' navigational behaviour. For example we can understand the number of users that started from a page and proceeded through some certain pages and finally visited their goal page. Also, we can obtain some information about page popularity or some pages that contain the most information for a visitor. The exact analysis methodology is usually governed by the application for which Web mining is done (Kolari, P. and A. Joshi, 2004). The most common form of pattern analysis is combining WUM tools with a knowledge query mechanism such as SQL. Visualization techniques, such as graphing patterns or assigning colours to different values, can often highlight overall patterns or trends in the data. Content and structure information can be used to filter out patterns containing pages of a certain usage type, content type, or pages that match a certain hyperlink structure.

Web usage mining applications:

Web usage mining is used to discover interesting user navigation patterns and can be applied to many real-world problems such as improving Web sites/pages, making additional topic or product recommendations,

user/customer behaviour studies, Online prediction etc. (Zaïane, O.R., 2001) Users' behaviour is used in different applications such as Personalization, e-commerce, etc. to improve the system and to improve the system design as per their interest etc.

A. Personalization of Web Content:

Web personalization offers many functions such as simple user salutation to more complicated features such as content delivery as per users' interests. Content delivery is very important since non expert users are overwhelmed by the quantity of information available online. It is possible to anticipate the user behaviour by analyzing the current navigation patterns with patterns which were extracted from past web log. Web Usage Mining techniques can be used to provide personalized web user experience (Nina, S.P., *et al*. 2009). For instance, in real time, it is possible to predict the user behaviour by comparing the current navigation pattern with typical patterns which were extracted from past web log. In this area, recommendation systems are the most common application; their aim is to recommend interesting links to products which could be interesting to users.

B. E-commerce :

E-Commerce applications need customer details for Customer Relationship Management. Mining business intelligence from web usage data is dramatically important for e-commerce web-based companies (Bhuvaneshwari, C. *et al*., 2016). Web usage mining techniques can also be useful in Customer Relationship Management (CRM). The issues specific to business such as customer attraction, customer retention, cross sales, and customer departure are mainly in focus (Reddy, K.S. *et al*. 2012).

C. Pre-fetching and Caching:

The results produced by Web Usage Mining can be exploited to improve the performance of web servers and web-based applications. With the use of weblogs that store user's access history, one can predict future accesses. Typically, Web Usage Mining can be used to develop proper pre-fetching and caching strategies so as to reduce the server response time.

D. Support to the Design:

Usability is one of the major issues in the design and implementation of web sites. The results produced by Web Usage Mining techniques can provide guidelines for improving the design of web applications (Qiu, F. and J. Cho, 2000). Adaptive Web sites represent a further step. In this case, the content and the structure of the web site can be dynamically reorganized according to the data mined from the users' behaviour.

E. Development Of User Friendly Web:

User's browsing pattern is analysed through Web Usage Mining which gives information regarding user's preferences. This information may be useful in development of websites which are easy and interesting to use. Naive users feel uncomfortable while surfing the internet due to complex structure of the websites. Application of Web Usage Mining can change the scenario. Web Usage Mining also allows restructuring of existing websites leading to their better management.

F. Security:

Pattern analysis develops a pattern about user's preferences, way of surfing, areas of interests, etc. Every time a user visits the Internet, the same pattern is followed. If a user other than the routine user visits on the Internet, change in browsing pattern is immediately detected. Thus, this technique may be used for intrusion or unauthorized access detection.

G. Search Engine Optimization:

Browsing pattern analysis can help in developing optimized search engines which will produce only relevant and filtered information by understanding what exactly the user want to search. Such technique will save the user from information overload problem.

Usage mining techniques are very useful to focus on customer attraction, customer retention, cross sales and customer departure. System Improvement is done by understanding the web traffic behaviour by mining log data so that policies are developed for Web caching, load balancing, network transmission and data distribution. Patterns for detecting intrusion fraud, attempted break-ins are also provided by mining. Performance is improved to satisfy users. Site Modification is a process of modifying the web site and improving the quality of design and contents on knowing the interest of users. Pages are re-linked as per customer behaviour.

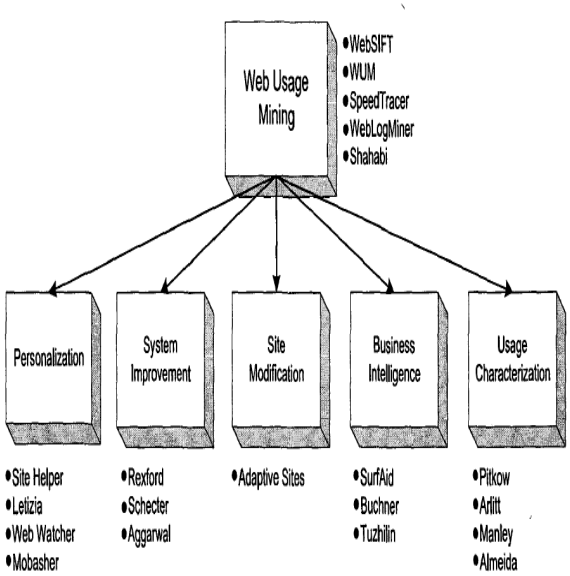

Fig. 5: Major Application Areas of Web Usage Mining

Agent-Based Approach:

Generally agent-based Web mining systems can be placed into the following three categories:

- Intelligent Search Agents:

Several intelligent Web agents have been developed that search for relevant information using domain characteristics and user profiles to organize and interpret the discovered information.

- Information Filtering/Categorization:

A number of Web agents use various information retrieval techniques and characteristics of open hypertext Web documents to automatically retrieve later and categorize.

- Personalized Web Agents:

This category of Web agents learn user preferences and discover Web information sources based on these preferences, and those of other individuals with similar interest.

Research Challenges:

We list below the current main web usage mining open problems: (Zhan, L., and L. Zhijing, 2003)

1. The quantity of data is continuously increasing.
2. The pre-processing step does not receive enough analysis efforts.
3. The Websites have no or little semantic definitions for their Web pages.
4. The sequential pattern mining techniques for WUM are not appropriate for dealing with the specifics of Web usage data, mainly with its huge quantity.
5. The sequential pattern mining techniques often provide short and uninteresting results.
6. The three steps of the WUM process are not coordinated to create a coherent and unique process.

The main areas of research in this domain are Web log data pre- processing and identification of useful patterns from this pre-processed data using mining techniques.

Web Usage Mining has provided a better way of web utilization but still there are some areas which need attention and more research need to be performed in these areas.

A. Development Of Adaptive Websites:

In the field of personalization of web, adaptive websites (Dong, D., 2009) can improve the search scenario. Adaptive websites are the websites which can change their organization & presentation according to the preferences of the users. Use of agents can make it simpler and user friendly. Agents are employed to provide recommendations for the users. Although work in this direction has started but it is in initial stage and requires more research efforts.

B. Data Sampling:

In case of on-line social networking sites, data sampling to reduce the size of the data is the main concern. Here data can be email, image, text, video etc. and a combination of these. Thus to reduce the size of the data in Web Usage Mining is a big challenge.

C. Privacy Concern:

In order to personalize the web and improve the browsing experience, it is necessary for the site administrator to identify the users uniquely. But, most users do not want them to get monitored every time they visit on the web. Some rules and regulations need to be applied so that site administrator can perform analysis of data without affecting the user's identity. That is, there should be strict rules that user's identity will not be disclosed and usage data will not be misused.

Scope Of Web Usage Mining:

Web Usage Mining provides a vast and precious amount of information which, if used properly, can be used in many applications of commercial importance. Also, it can improve the web interaction drastically.

It has three main goals (Kosala, R. and H. Blockeel, 2000) for any application.

- To optimize user's web surfing experience.
- To improve the performance of a website.
- To improve the design of a website.

Web mining is an upcoming area of research which is gaining attention from research fraternity. Information overload is making commercial and personal web usage cumbersome for the users. Web mining is the only solution for filtering web contents, selecting and providing appropriate contents to end users. Considering the large volumes of data already available on the web and the rate of new information upload on the web, there is dire need of a technology which can automate the processes involved in web mining, specifically in usage mining. Agent technology seems to be a promising solution in this case. Agents are already been employed in many web based research domains such as semantic web, wireless sensor networks, web services, etc. and had been proved beneficial.

Literature review highlights that researchers have not paid much attention towards employing agent technology in web usage mining in past, however some efforts are being made in this direction recently (Maryam Jafari, *et al*., 2013). Kosala et.al in (Federico Michele Facca and Pier Luca Lanzi, 2003) indicated that two basic approaches used by agents in mining process are:

- ***Content Based Approach:***
  an analysis of the content explored by any user is performed.

- ***Collaborative Approach:***
  users having similar searching behaviour are found and then recommendations are made based on their interests.

Nowadays, agents are an integral part of web search engines in the form of web crawlers. Once employed in web usage mining, they can help facing the existing challenges in this domain. Agent technology has become prime solution for problems focusing on distributed, dynamic and heterogeneous environments such as WWW. Since web usage mining possesses the above said attributes, we would like to explore applicability of agent technology in this domain.

Major research systems and projects concerning Web usage mining.

No.	Title	URL	Major Method/Application
1	Adaptive Web Sites	http://www.cs.washington.edu/research/adaptive/	Pattern application
2	GroupLens	http://www.cs.umn.edu/Research/GroupLens/	Recommender systems
3	MiDAS	–	Sequence discovery
4	WebQuilt	http://guir.berkeley.edu/projects/webquilt/	Proxy logging
5	WebLogMiner	http://www.dbminer.com/	OLAP application
6	WebSift	http://www.cs.umn.edu/Research/webshift/	Data mining
7	WUM	http://wum.wiwi.hu-berlin.de/	Sequence discovery

Web usage mining can be applied in the following areas (It probably deals with the existing practical applications, the on-going research areas and the scope of agent technology in the research areas.)

A. Adaptive Web Sites:

Personalization is one of the most widely researched areas in Web usage mining. An early effort in this direction was the adaptive Web site challenge posed by Oren Etzioni and colleagues. Adaptive sites automatically change their organization and presentation according to the preferences of the user accessing them. Other contemporary research seeks to build agent-based systems that give user recommendations. For instance, Web-watcher22 uses content- and structure-mining techniques to give guided tours to users browsing a page. Popular Web sites like Amazon.com use similar techniques for "recommended links" provided to users. All these approaches primarily use association rules and clustering mechanisms on log data and Web pages.

B. Robust Fuzzy Clustering:

Anupam Joshi and colleagues use fuzzy techniques for Web page clustering and usage mining, and they use the mined knowledge to create adaptive Web sites. They argue that given the inherent ambiguity and complexity of the underlying data, clustering results should not be clearly demarcated sets but rather fuzzy sets, that is, overlapping clusters. For instance, a user can belong to multiple user interest groups because at different times he or she accesses the Web for different information or merchandise. Insisting that each user fit only a single group is clearly inconsistent with this reality. Moreover, given the noise expected in the data despite cleaning attempts, the clustering process must be robust in the statistical sense. Raghu Krishnapuram and colleagues discuss fuzzy clustering and its application to Web-log analysis and present a fast linear clustering algorithm that can handle significant data noise. They use this algorithm to cluster Web access logs and use the traversal patterns identified for specific groups to automatically adapt the Web site to those groups.

C. Association Rules:

Early systems used collaborative filtering for user recommendation and personalization. Bamshad Mobasher and colleagues used association-rule mining based on frequent item sets and introduced a data structure to store the item sets. They split Web logs into user sessions and then mined these sessions using their suggested association rule algorithm. They argue that other techniques based on association rules for usage data do not satisfy the real-time constraints of recommender systems because they consider all association rules prior to making a recommendation. Ming-Syan Chen and colleagues proposed a somewhat similar approach that uses a different frequent item set counting algorithm.

D. Recommender Systems:

J. Ben Schafer and colleagues note that recommender systems have enhanced e-business by converting browsers to buyers,

- increasing cross-sell by identifying related products, and
- building loyalty.

These systems primarily use association rule mining for pattern detection. In an e-business scenario, a recommender system uses customers' Web baskets and shopping carts as data sources. Amazon.com has the most prominent application: "Customers who bought product A also bought product B."

E. Web Site Evaluation:

Myra Spiliopoulou suggests applying Web usage mining to Web site evaluation to determine the needed modifications primarily to the site's design of page content and link structure between pages. Such evaluation is one of the earliest steps in Web usage analysis conducted by Web sites and is necessary for repeat visitors. Evaluation is important because all subsequent Web usage mining techniques are effective only in the presence of large amounts of data created by repeat visitors. The main technique for evaluating data is to model user navigation patterns and compare them to site designers' expected patterns. The Web utilization miner (WUM) analysis tool, for example, incorporates evaluation. Hamlet: To Buy or Not to Buy Etzioni and colleagues applied Web mining to airline ticket purchasing. Airlines use sophisticated techniques to manage yield, varying ticket prices according to time and capacity. Etzioni's approach mined airline prices available on the Web and price changes over time to produce recommendations regarding the best time to buy tickets. Many more innovative areas are yet to be explored.

F. Privacy Issues:

Recent data-mining privacy violations have caused concern, specifically when data mining has involved vertically partitioned data, that is, data about the same entity or individual from multiple sources. One example is Terrorist Information Awareness DARPA-initiated program that aims to aggregate information from disparate sources to detect patterns that might indicate a terrorist. This program has led to serious public debate about whether such a system, even if technologically possible, should be used. DoubleClick's online advertising is an instance of tracking user behaviour across multiple sites. If a user's transactions at every Web site are identified

through uniquely identifiable information collected by Web logs, they could create a far more complete profile of the user's shopping habits. The current Web privacy architecture provided by the Platform for Privacy Preferences Protocol and AP3P Preference Exchange Language (APPEL) lets users control this kind of usage by explicitly agreeing or disagreeing to such tracking. However, Web sites and popular Web browsers offer limited support for such tools. Web mining research should accommodate this preference set and enforce it across organizations and databases. Lorrie Cranor surveys possible research direction in this area.

Web usage mining tools:

- SpeedTracer
- Suggest 3.0
- Collaborative filtering
- Web Personalizer
- Classifying User Navigation Patterns Using Longest Common Subsequence Algorithm

TOOLS	FEATURES
Data Pre-processing Tools	
Data Preparator	Performs cleaning, extraction and transformation of data before pattern discovery
Sumatra TT	Platform independent data transformation tool. Based on Sumatra script and support Rapid application Development.
Lisp Miner	Performs data pre-processing by analysing the click stream and data collected.
SpeedTracer	Mines web server logs and reconstruct the user navigational path for session identification
Pattern Discovery Tools	
SEWEBAR-CMS	Provides interaction between data analyst and domain expert to perform discovery of patterns. Helps in selection of rules among various rule association rule mining.
i-Miner	Discover data cluster by using fuzzy clustering algorithm and fuzzy inference system for pattern discovery and analysis.
Argunaut	Develop the patterns of useful data by using sequence of various rules.
MiDas (Mining In-ternet Data for Associative Sequences)	Discover marketing based navigational pattern from log files. It applies more features to traditional sequential method.

Pattern Analysis Tools

Webalizer	GNU GPL license based and produces web pages after analysing patterns.
Naviz	Visualization tool that combines 2-D graph of visitor access and grouping of related pages. It describes the pattern of user navigation on the web.
WebViz	Analyse the patterns and provides them in the form of graphical patterns.
WebMiner	Mines the useful patterns and provides the user specific information.
Stratdyn	Enhances WUM and provides visualization of patterns.

Agent Technology:

Agents are software entities which perform a specific task on behalf of their user. These reflect agents in the real world which do something on user's behalf, and any technology that makes use of agents to perform a function is termed as agent technology. Once deployed an agent reside in an environment, senses its own input as and when it occurs and acts on it to achieve its predefined goal autonomously. Agents can also be categorized depending upon their characteristics. These are

A. Personal Agents:

Those agents who interact directly with the user, monitor user's activities & user's preferences are personal agents. All the three classes of agents discussed earlier are the kinds of personal agents. Personal Agents can be classified in following categories based on the nature of their task performance (Srivastava, J., *et al*, 2000)

1) Gopher Agents:

These agents, on meeting a specified condition, perform the task and give the result to the user. For example, inform the user when email from any particular id is received.

2) Service Performing Agents:

The agent performs a task only when the user requests them to do so. For example, book an air ticket to U.K next month.

3) Predictive Agents:

The agent which voluntarily provides information and services to the user, without being asked for, are known as predictive agents. For example, an agent may inform a user that a heavy discount is being offered by a particular brand of user's interest.

4) Mobile Agents:

Those agents who visit the remote sites to collect the information and perform tasks are known as mobile agents. These agents return back to their source and provide results to the user.

5) Collaborative Agents:

Agents which collaborate with other agents to collect information or work in groups to perform some task are known as collaborative agents

Some examples of Agent Applications:

- User Interface Agents
  - Microsoft Office Assistant.
- Business process Agents
  - Data-driven workflow management.
- Information Management Agents
  - Email filtering agents.
  - Web browsing assistant
  - Notification agents.
  - Resource Discovery agents.
  - Data Mining agents

Agents are preferred for various applications because they can reduce human work, handle information overload, provide automated help to the vast hordes of untrained users, provide a new, more powerful methodology to develop complex software systems, asynchronous / autonomous task execution, reduction of network traffic by transferring the agents to remote places to carry out a task rather than having to remotely monitor the task execution, negotiation capabilities, learning capabilities – Proactive.

Some of the existing Agent Platform implementations:

- Telescript (General Magic).
- Aglets Workbench (IBM).
- Concordia (Mitsubishi).
- JADE (Java Agent Development Framework, Sun Oracle University).
- Grasshopper (IKV++).
- Odyssey (General Magic).
- Voyager (Object Space).
- KQML
- ACL
- FIPA-OS

Agent Host:

It keeps track of all the agents executing in the system. It will interact with other Agent Hosts to transfer an agent from one system to the other.

Agent:

The Agent class defines the agent. An instance of this class exists for each agent executing on a given agent host.

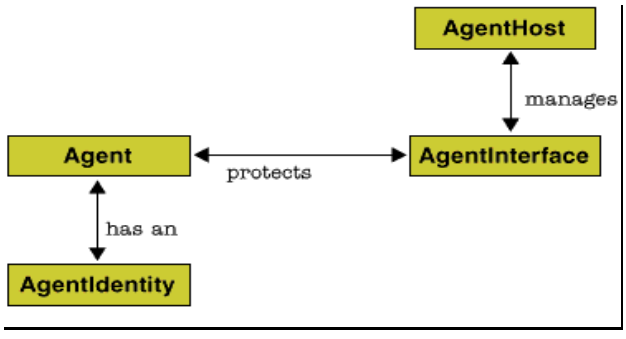

Fig. 6: Agent architecture

Agent Interface:

An instance of this class envelopes an agent and provides access to it via a well-defined interface. It is also the primary conduit for communication between agents. An Agent Interface instance is the only handle an agent gets to the other agents executing on a given host.

Agent Identity:

An instance of this class uniquely identifies an agent. Agents use this information to identify the agents with whom they are interested in collaborating.

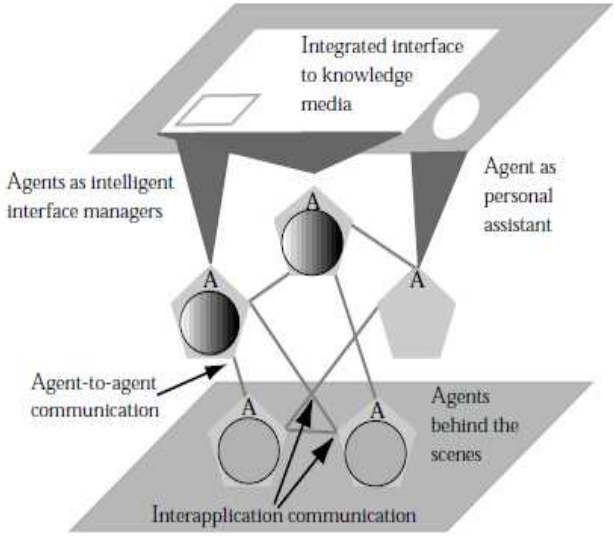

Fig. 7: Agent Enabled System Architecture

Fipa-Os Agent Development:

FIPA (Foundation for Intelligent Physical Agents) to set computer software standards and FIPA-OS toolkit enabling rapid development of FIPA interacting agent based system. JADE is a Java Agent Development Framework implemented in Java language which is a software framework. It simplifies the implementation of multi-agent systems. Implementations are made easy using a middle-ware that satisfies the FIPA specifications. It supports debugging and deployment phase through a set of tools. The various features of JADE are:

- Distributed agent platform.
- Support to the execution of multiple, parallel and concurrent agent activities.
- Efficient transport of ACL messages.
- Ready to use interaction protocols.

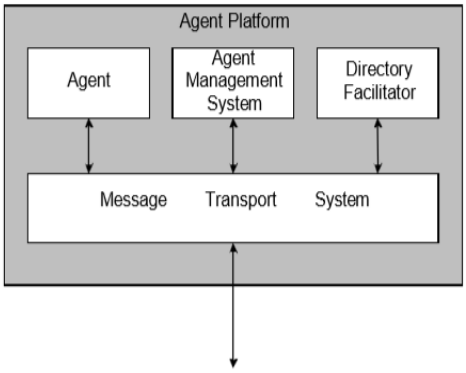

Fig. 8: Reference architecture of a FIPA Agent Platform

The framework contains:

1. *Agent Management System:* It will act as a supervisor to control over the access and use of Agent platform.
2. *Directory Facilitator:* It provides default yellow page services in the platform.
3. *Message Transport System:* It is a software component that will monitor the message exchanges to/from the platform, also called as Agent Communication Channel (ACC).

All the ACL messages are placed in a private queue for each agent. Agent will access the messages using these modes: *blocking, polling, timeout and pattern matching based.*

Agent platform may contain several hosts. Each host will contain one Java Application, and hence one Java Virtual Machine (JVM). An environment to execute an agent and several agents concurrently is provided by JVM which will act as a basic container of agents. Main Container contains AMS and DF. All the other containers on same or different systems are called as *secondary containers* and provide a complete runtime environment.

Scope Of Agents In Web Usage Mining:

Web mining is an upcoming area of research which is gaining attention from research community. Information overload is making commercial and personal web usage cumbersome for the users. Web mining is the only solution for filtering web contents, selecting and providing appropriate contents to end users. Considering the large volumes of data already available on the web and the rate of new information upload on the web, there is a need of a technology which can automate the processes involved in web mining, specifically in usage mining [9]. Agent technology seems to be a promising solution in this case. Agents are already been employed in many web based research domains such as semantic web, wireless sensor networks, web services etc. and had been proved beneficial.

Conclusions:

This work explored the area of Web Usage Mining emphasizing on its application areas and research challenges. The role and applicability of software agents in context of web usage mining has also been explored. When employed, agents can help overcome challenges existing in this domain, and intelligent mechanisms may be designed to address problems of web usage mining.

REFERENCES

Bhuvaneshwari, C. *et al*., 2016. Analyzing the Risk of Diabetes Mellitus using Patient EMR by Association Rule Mining, International Journal ADVANCES in NATURAL and APPLIED SCIENCES, 10(5): 102-106.

Dong, D., 2009. Exploration on Web Usage Mining and its Application, in the Proceedings of the IEEE International Workshop on Intelligent Systems and Applications, ISA 2009, pp: 1-4.

Federico Michele Facca and Pier Luca Lanzi, 2003. Recent Developments in Web Usage Mining Research, in the Proceedings of 5th International Conference on Data Warehousing and Knowledge Discovery, pp: 140-150.

Fürnkranz, J., 2005. Web mining. In The Data Mining and Knowledge Discovery Handbook, O. Maimon and L. Rokach, Eds. Springer, Berlin, Germany, pp: 899-920.

Genesereth, M.R. and S.P. Ketchpel, 1994. Software Agents, Communications of the ACM, 37(7): 48-53.

Griss, M., 2001. Software agents as next generation software components, In Component-Based Software Engineering: Putting the Pieces together, Edited by G.T. Heineman and W.T.Councill, Addison-Wesley, pp: 641-657.

Jayasowbaghyam, E. and K. Nithya, 2016. Adaptive Modeling On Satellite Image Processing Information Extraction Using Knowledge Fusion Mining, International Journal ADVANCES in NATURAL and APPLIED SCIENCES, 10(5): 59-67.

Jennings, N.R. and M. Wooldridge, 1996. Software Agents, IEEE Review, pp: 17-20.

Kolari, P. and A. Joshi, 2004. Web Mining: Research and Practice, Computing in Science and Engineering, 6(4): 49-53.

Kosala, R. and H. Blockeel, 2000. Web Mining Research: A Survey, ACM SIGKDD Explorations Newsletter, 2(1): 1-15.

Li, Y. and N. Zhong, 2004. Web Mining Model and Its Applications on Information Gathering, Knowledge-Based Systems, 17: 207- 217.

Maruthu Pandi J, *et al*., 2016. Efficient Feature Extraction for Text Mining, International Journal ADVANCES in NATURAL and APPLIED SCIENCES, 10(4): 64-73.

Maryam Jafari, *et al*., 2013. Extracting Users'Navigational Behavior from Web Log Data: a Survey. Journal of Computer Sciences and Applications, 1(3): 39-45.

Nina, S.P., *et al*. 2009, Pattern Discovery of Web Usage Mining. In the proceedings of International Conference on Computer Technology and Development., pp: 499-503.

Qiu, F. and J. Cho, 2000. Automatic identification of user interest for personalized search, in the Proceedings of the 15th International Conference on World Wide Web, pp: 727-736.

Reddy, K.S. *et al*. 2012. Understanding the Scope of Web Usage Mining & Applications of Web Data Usage Patterns, in the Proceedings of the International Conference on Computing, Communication and Applications, ICCCA – 2012, pp: 1-5.

Srivastava, J., *et al*, 2000. Web usage mining: Discovery and applications of usage patterns from web data,

ACM SIGKDD Explorations Newsletter, 1(2): 12-23.

Zaïane, O.R., 2001. Web usage mining for a better web-based learning environment, In the Proceedings of Conference on Advanced Technology for Education, pp: 60-64.

Zhan, L., and L. Zhijing, 2003. Web Mining Based On Multi-Agents, Proceedings of the 5th International Conference on Computational Intelligence and Multimedia Applications, pp: 90.