Knowledge Discovery from Web Data for Web Personalization

Sowbhagya M P, Ganavi K R, Yogish H K

Abstract: Because of the large and rapid increase in web data size and number of users, web users now face the problems of overloading and drowning information. As a result, Recovery of internet-based data and web applications, providing web users with more accurate information becomes a critical issue. In this study, by analyzing web data features, we aim to improve the performance of web information retrieval and web presentation through web data mining processes that discover the knowledge (intrinsic relationships) between web data expressed as textual, linkage or usability information. We concentrate on discovering web usage patterns through web usage mining, and then using the discovered usage knowledge along with profile information to provide web users with more personalized web content. Personalization is an engaging service for website visitors, based on their characteristics and deliberate behaviors to facilitate conversion and long-term commitment expectations. The purpose of this work is to extract the knowledge from web data and use this knowledge to create a web personalization system that allows users to access the content of their need from the website without specifically specifying it. The knowledge could be the navigational actions of the user as exposed by web access log analysis, as well as the characteristics and preferences of the user reflected by user profiles. Such knowledge is further analyzed to improve system performance, retention of users and/or modification of the site. This paper provides a comprehensive survey of the different approaches suggested by Web Personalization researchers and list out some of the issues that need to be tackled soon.

Keywords : Web Personalization, User Profile, Ontology, Information Retrieval, Semantic Web.

I. INTRODUCTION

The personalization of a website's services is avital step towards alleviating the abundance of data, making the internet a more comfortable atmosphere for specific user and thus building truthful relationships among the website and the user. Web Personalization is simply defined in as adapting information systems to individual users ' needs and interests. A custom website usually identifies its users, collects information about their interests, and adapts its products to suit the needs of the users. User personalization improves a visitor's web experience by delivering the details that
in the right way and at the right time, the user wants to see.

Revised Manuscript Received on December 16, 2019.

*Sowbhagya M P, Department of Computer Science & Engineering, affiliated to VTU Belagavi, Bengalure, India.
Email: sowbhagya.mp@gmail.com

Ganavi K R, Department of Information Science & Engineering, affiliated to VTU Belagavi, Bengalure, India.
Email:ganavi.kr@gmail.com

Dr.Yogish H K, Department of MCA, Ramaiah Institute of Technology, Bengaluru, India. Email: yogishhk@gmail.com

Web personalization also provides ways to learn more about customer requirements in the e-business, predict future trends and ultimately improve customer trustworthiness to the service provided. One way to increase the flexibility of the Internet is to simplify the application of Web-based services to their users. Machine learning approaches have a strong implementation record for specific tasks. Furthermore, integrating machine learning into broader application models like Knowledge Discovery Database (KDD) will afford a comprehensive explanation to the adaptation issue. KDD was used to analyze and extract useful information from Web-based data collected. This effort has been called Web Mining (Etzioni, 1996) and one of its branches is to analyze usage data. This process is known as an application for internet mining. In the early days of internet mining, its architecture was not commonly used. The emphasis was on expanding the skills of decision support for human experts to interpret and use as concise data models. This form of expertise can be provided directly to users without the involvement of any human expert to improve their website experience. The use of mining is now widely recognized as a valued source of site personalization perceptions and solutions. In this paper, for the purpose of personalization research, present a review of current work. Beginning with a review of the theory of personalization of Web and its relationship with usage mining, the focus will later be on the methods used in usage mining, the different solutions suggested in the literature, and how these methods can be applied to Web personalization systems.

II. PERSONALIZATION OF THE WEB

The word personalization covers methods and techniques for providing website visitors with a value-added browsing experience. This quality is accomplished through a variety of functions that a web personalization program can provide, making interacting with the website simpler, saving time for users, and thus achieving one of the website's main goals. In this section, we analyze the Configuration functions that can be presented with a set of parameters for the site configuration framework. A personalization of Web framework may offer a change of features ranging from basic user greeting to more complex features like delivery of custom content. Kobsa et al. suggests a category of Web configuration features, which here is generalized to a generic classification scheme. There are four basic configuration functions that we distinguish: memorization, guidance, customization and support for task performance.

Retrieval Number: B12271292S19/2019©BEIESP
DOI: 10.35940/ijitee.B1227.1292S19

733

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

i. *Memorization*: Here, in its ' memory ' archives, the program records and stores user data including name and browsing history. Examples:

a. *User Salutation*: The Web configuration system identifies the user returning and shows the customer's name along with a welcome message. Although this function is simple, as users, websites that identify them as individuals and not as regular visitors are relaxed, this is first step towards increasing visitor loyalty.

b. *Bookmarking*: In the past, the program stores the web pages of a website and shows them for that site to the client through a custom bookmarking scheme.

c. *Custom access rights*: A website may use custom access rights to distinct authorized users from common users. For different types of data, such as reports or even for running web applications.

ii. Guidance: Guidance as a customization tool refers to the customization system's attempt to help the user get the information the user is searching for in a page quickly, as well as to provide alternate search options for the user. Not only does this configuration feature improve customer satisfaction, it also greatly alleviates the problem of overloading data that a large website user can face. Example: Recommendation of hyperlinks

iii. Customization: Customization feature refers to the Web page's content, structure and design adjustment to match the user's expertise and interests. The main objective is to handle the load of data by promoting the communication between the client and the platform.

a. *Personalized layout.* It is an inherited function from Adaptive UIs, where a web page is modified its user profile based on its design, color, or local data.

b. *Content Customization*. To adapt to the experience, desires and preferences of the user, the content of the website displayed to a user may be changed.

c. *hyperlinks Customization*. You can customize hyperlinks within a current page. Here the site will be updated by inserting or removing hyperlinks within a page. This can help improve the whole website architecture by removing unusable links and modifying the website topology to make it more accessible.

iv. *Task Performance Support*: Supporting task success is an ability requiring direct action on a client's behalf. This is the most sophisticated configuration feature inherited from the Adaptive Systems development class known as Personal Assistants.

III. REQUIREMENTS OF WEB PERSONALIZATION SYSTEM

The above-mentioned roles place a number of criteria on the development of a personalization framework to build a reliable and scalable system.

i. *Domain Specification*

Domain-sensitive is the feature of a Web personalization framework. The same platform that operates in different domains, such as digital library, website may offer diverse functions for customization. It is therefore necessary to specify and explain thoroughly the framework under which the personalization program must function.

ii. *User Identification*

To differentiate returning visitors from first-time visitors,

it is important to identify the user who accesses a website. However, this is not always possible due to concerns about privacy and anonymity on the Web. Nonetheless, what may be necessary is to describe the target, aspirations or encouragement of the user who accesses the page.

iii. *Efficient user data acquisition*

A configuration program could be capable of gathering all the necessary customization user data. The quantity of data and type required be subject to on the results selected for optimization functions. User data collection is an ongoing process, as it should be possible for the system accommodate large data volumes and increased Variability of information due to internet existence.

iv. *Flexible data processing*

To distinguish noise from relevant data, the data collected should be processed, analyzed and evaluated, and eventually formatted in order to be prepared for customization. Data processing is a domain-dependent process and it requires a high degree of flexibility to adapt the design approach to various domains and configuration functions.

v. *Efficient Creation of Models*

Web configuration system can create and maintain user models efficiently and reliably, i.e. Data collected by the system about desires, expertise, goals and preferences of the users. Application models can be constructed manually or through a machine learning system.

vi. *Considerations of practical and legal nature*

Usability of Web personalization can be tailored to the needs and goals of the user and should be compatible with various important functional constraints, such as response time. In addition, personal information of the client should always be protected, and the user should be aware of how this information is collected and used.

IV. APPOROACHES TO WEB PERSONALIZATION

Personalization was a variety of online optimization strategies have been introduced in recent years, known as a solution for the problem of overloading data and increasing the trustworthiness of visitors to the website during development of web. Although discussing these methods in depth is not within the scope of the study, a brief summary of the most common approaches is given below. Mobasher et al. (2000a) classify Web personalization techniques into three generic approaches:

a. *Manual decision rule systems.* According to this approach, a web-based service is customized through its designer's manual intervention and usually with the user's cooperation. Static user models are typically obtained through a user registration procedure and a number of rules are specified manually for the web content provided to users with different models. Two examples of this approach from a wide range of products are Yahoo!'s Personalization Engine (Manber et al., 2000) and Websphere Personalization (IBM).

b. *Content-based filtering systems.* To discover a user's personal interests, this group of techniques applies machine learning methods to web content, mainly text. NewsWeeder (Lang,1995) is a software which adopts this approach, which is able to adapt user models based on the browsing activity of a user based on the similarities between web documents that contain news items. Such models can be used according to the specifications of each client to process news items.

c. *Social or collaborative filtering systems*. The aim of this approach is to customize a product without needing web content analysis. Personalization is accomplished by searching for common features in different user preferences, usually expressed directly by them in the form of item ratings, and registered by the program. Examples of products that also use this method are the Recommendation Engine (Net Perceptions) and WebSphere Personalization (IBM), while its most famous application is in the electronic shop of amazon.com.

V. THE ROLE OF KNOWLEDGE IN PERSONALIZATION

For all approaches applying data mining to Web data over the past decade, scientists have proposed a new unifying field called Web Mining. Data mining tools are aimed at extracting information from the Internet rather than extracting data. Web mining work is generally divided into 3 categories: Web content mining, Mining Network Use and Web Mining Layout.The aim of web content mining is to use data mining to extract useful information from web pages content. Web use mining aims to find interesting usage patterns by analyzing data on web use. Finally, web structure mining is a new area where data mining is applied to the framework of the web graph.Initially, the goal of Web Usage Mining was to help the human decision-making process andtherefore, the process usually results in a series of data models disclosing implicit knowledge of information objects, such as web pages on a website. These models are estimated and manipulated by experts, such as Business Intelligence Market Expert or the website director, who wants to improve the design of the website and improve user experience. While most of work in Web Use Mining is not about configuration, it is straightforward to connect to automated customization tools. Web use mining research can be a source of web customization ideas and approaches. The communication between the user and the website in question is the usage information. Web use mining offers an approach to gathering and pre-processing this information and creates models that represent the user's interests and behaviors Such templates could be used automatically by a configuration program. That is to say, to carry out the requisite personalization functions without the involvement of any human expert. This form of information, i.e. the application models, is a practical understanding of web personalization. Web application mining methods may be used by a web personalization program to achieve the necessary robustness and versatility.The main reason for this research is the close relationship between Web application mining and Web customization. Image. Fig. 1 Summarizes the stages described graphically.
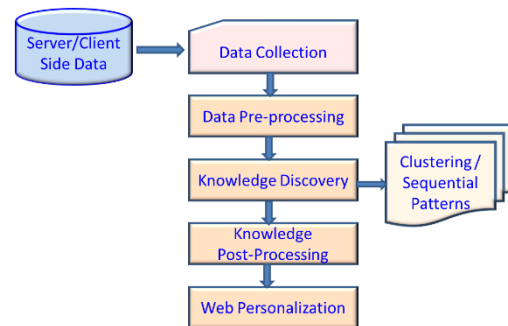
Knowledge Discovery from Web Data Process:

Fig. 1. Knowledge Discovery Stages from Web Data

Data Collection. During this stage, usage data from various sources such as Web servers, clients, or from intermediary sources such as proxy servers and packet sniffers is gathered.
Data Pre-processing. This is the stage where data are cleaned from noise, their differences are resolved and merged and combined to be used as input to the next Pattern Discovery stage. It mainly includes sorting information, identifying users and user sessions. The methods used here can provide effective data processing.
Knowledge Discovery. Here, knowledge is discovered from data by applying machine learning and statistical techniques such as clustering, grouping, discovery of associations, and sequential pattern discovery. The patterns required for Web customization are consistent with users ' behavior and interests. This is the stage where learning methods are used to automate user models construction.
Knowledge Post-Processing. In this last stage, the extracted knowledge is evaluated and usually presented in a form that is understandable to humans, e.g. using reports, or visualization techniques. For Web personalization the extracted knowledge is incorporated in a Personalization module in order to facilitate the personalization functions. Fig. 2 shows the proposed architecture of knowledge discovery from web data for web personalization.
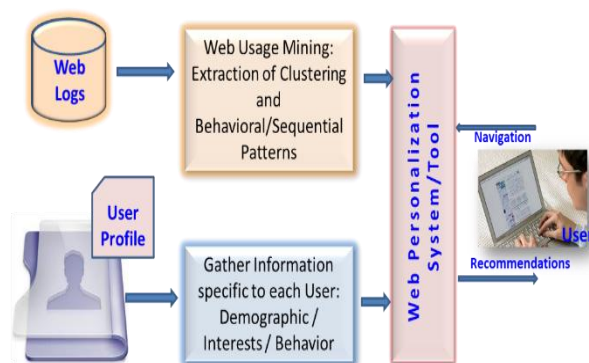

Fig. 2. Knowledge Discovery from Web Data

A. DataCollectionforPersonalization

The first step in the process of personalization is to collect appropriate Web data, which will be analyzed to provide useful *information* on the actions of the users. Web personalization has two main data sources: data on the web server side and data on the client side. The collection of web usage data is the first step towards realizing the functions of web personalization. Registration data, for instance, can help to realize all functions, as these data provide a clear view of the user profile in terms of desires, interests and expertise if they are sufficiently accurate. Different methods of data collection are needed for different functions of memorization. User greeting requires direct user input to obtain the user's name via registration information. Either in a local database or in a cookie file, the user name can be stored. Usually, guidance roles require information about what the user is searching for on a website, as well as information about the knowledge level of the user. Using server-side and client-side data as well as intermediate sources such as packet sniffers, this information can be compiled. Customization functions mainly require information on the desires and expectations of the users. Through capturing the browsing history of users, this data is gained. For this purpose, both server-side and client-side data are appropriate, although server log files are typically used as the main source of information.

B. Preprocessing for Web Personalization

In the first stage of data mining, site data collected is typically complex and voluminous. To be used to uncover trends, these data must be arranged in a clear, structured and detailed view. As with most data mining applications, pre-processing of information involves removing and sorting redundant and obsolete data, predicting and replacing missing values, eliminating noise, transforming and encoding data, and resolving any inconsistencies. For the success of data mining, the role of data transformation and encoding is especially important. This stage involves the identification of users and user sessions in web application mining, which are to be used as the basic building blocks for pattern discovery. The accurate identification of users and user sessions is particularly important for site configuration, since individual user models are focused on user behavior encoded in user sessions and correctly associated with the relevant users.

Removing noise and irrelevant data is the first step towards improved Web personalization as information not related to the user's browsing behavior that misguide the pattern discovery process. In addition, user identification is one of the web personalization system's most critical parameters. As a result, most of the configuration functions require both data filtering and user identification in some way. User identification is the only step required for memorization functions to pre-process information. Using user session recognition methods is the first step towards better user feedback and is also critical in defining the purpose of the user to perform a certain task in order to realize the class of configuration functions.

C. Knowledge Discovery

Use of machine learning and statistical methods to derive data from preprocessed data. These methods reflect the four most common approaches used in data mining literature: clustering, grouping, connection discoveryand sequential

pattern discovery. Difficulty of defining large amount of web data for supervised learning has resulted in clustering, unattended methods being introduced.

A. *Clustering pattern discovery*: Many approaches used to discover patterns from Web data are methods of clustering. The aim of clustering is to divide a set of data into groups that are very different from each other and whose members are very similar. Han and Kamber (2001), suggest a general taxonomy of methods of clustering that includes the following categories:

- *Partioning methods* that build k groups of a given set of data where each group is a cluster.
- *Hierarchical methods* which decompose the data set and create a hierarchical cluster structure.
- *Model-based methods* that find the best fit between a set of data and a model.

Clustering has been used to group specific users of browsing activity and related content Web pages (Srivastava et al.,2000). The requirement that clusters should be permitted to overlap is a major restriction placed on the option of the clustering method by Web application mining. This is important for web personalization as more than one group may belong to a user or a web page. Description of clustering algorithms is presented in Table-I. The table describes Web Use Mining applications using different algorithms, the clustering approach they follow, and an description of each method's strengths and weaknesses in the Web configuration context.

Table-I. Summary of clustering algorithms

Algorithm	Application	Clustering type	Merits	Limitations
Leader (Hartigan,1975)	Clustering user sessions	Partitioning	Incremental, Qualitative Clusters	Order dependent
EM (Dempster et al., 1977)	Clustering user sessions represented by Markov chains	Partitioning	Memory Efficient, Sound mathematical background	Slow linear Convergence, Computationally expensive
Fuzzy c clustering (Bezdek,1981) Nasraoui et al. (1999)	Clustering user sessions	Partitioning Fuzzy Clustering	Overlapping clusters	Design of membership function
COBWEB (Fisher,1987)	Clustering user sessions	Model-based	Incremental, Cluster characterization	Order dependent
Autoclass (Hanson et al.,1991)	Clustering user sessions	Model-based	Strong Mathematical Foundation. High quality clusters	Computationally Expensive, Requires prior assumptions
BIRCH (Zhang et al.,1996)	Clustering user sessions	Hierarchical	Incremental, Efficient for high-dimension data	Order dependent Creation of 'unnatural' clusters
Self-Organizing Maps (Kohonen, 1997)	Clustering user sessions	Model-based	Visualization of high-dimensional data	Preprocessing and Normalization of the data, Prior specification of the number of clusters
ITERATE (Biswas et al., 1998)	Clustering user sessions	Model-based	Order Independent	Not incremental, Not scalable
PageGather (Perkowitz and Etzioni,1998	Index Page Synthesis	Partitioning	Overlapping clusters, Each cluster is a direct behavioral pattern	Computationally expensive

B. Sequential pattern discovery introduces the time element in the process of discovery. The aim is to recognize temporal patterns (event sequences) that frequently occur in the data. This approach is particularly useful in recognizing site application information navigation trends. Two types of methods were used to discover sequential patterns: deterministic techniques, tracking the user's navigational actions, and stochastic methods using the sequence of visited web pages to predict future visits. Table-II presents an overview of the sequential pattern discovery methods. Also presented is the Web implementation mining software which uses each algorithm, as well as the sequential method of exploration and the advantages and disadvantages of each approach.

Table-II. Summary of sequential algorithms for pattern discovery

Algorithm	Application	Approach	Merits	Limitations
Borges and Levene (1999)	Extraction of navigation patterns from user sessions	Stochastic	Automatic generation of navigation path	Requires heuristics for Output refinement
Spiliopoulou et al. (1999a)	Extraction of sequence rules	Deterministic	Scalable, Meaningful patterns	Semi-automated procedure
Paliouras et al. (2000b)	Clustering of Navigational patterns,	Deterministic	Scalable, Meaningful patterns	Limited first-order Modeling patterns
Markov Models (Zhu,2001),	Link Prediction	Stochastic	Sound mathematical Background, Automatic generation of navigation path	Hard-to-interpret Models, Computationally expensive

D. Discovering personalization operational knowledge

The trends observed were processed and evaluated to help human experts ' decision-making system, for example. Managers of the company, responsible for taking appropriate action. Incorporating internet mining into the customization phase by explicitly feeding the collected information into the web-based system behavior configuration module is a more interesting approach to customization. This functional expertise is then passed on to users through one or more of the configuration functions discussed in this paper. A web configuration program's usability is largely based on the installation policy of a website., i.e. how data is transmitted to the end user.

VI. EXISTING SYSTEM

By means of various functions, personalization is accomplished by enforcing a certain automation rule. Two common types can be classified into the systems examined: single-function and multi-function systems. Multi-function systems can be designed to provide a diversity of customization features, although single-function systems are designed to offer a single configuration feature.The framework analysis presented here focuses on the customization solution that they provide in terms of policy support for this customization solution and functionality as well as the methodology they use for internet mining. Web mining provides a new and exciting approach to customization. The software systems have already used some of the methods used in Internet usage mining. Nevertheless, the software solution offered by these systems and the use of Web Mining to accomplish the task still have several open issues. Tables-III outline these systems.

Table-III. Summary of Existing Personalization Systems

System	Patterns Used	Personalized Functionality	Personalization Policy
SETA (Ardissono and Torasso, 2000)	Associations	Recommendation of Hyperlinks, User Salutation, Content customization	Single User, Dynamic
TELLIM (inTELLIgent Multimedia)	Classification	Personalized layout, Content customization	Multi User, Dynamic
Oracle9iAS	Classification and Associations	Recommendation of Hyperlinks, User Salutation	Multi User, Dynamic
NETMIND	Clustering and Classification	Recommendation of Hyperlinks, Personalized layout, Content customization	Multi User, Dynamic
SiteHelper	Classification	Recommendation of Hyperlinks	Multi User, Static
WUM	Sequential Patterns	Customization of Hyperlinks	Multi User, Static

A. Web Content based Search

Web content mininginvolves the use of data mining techniques. Due to the steady growth of web content and substantial economic gain from such mining, activities in the web content mining area have expanded rapidly. Yet automated discovery of targeted or unexpected information still faces some research challenges due to the complexity and lack ofweb datastructure. Table-IV summarizes different Web content-based search approaches [55].

B. User Interest based Search

Mining for web usage can be used for web-based search. Web usage mining is an important criterion for considerate the activities of users on the Internet and has recently been a favorite field for many researchers.

Table -IV. Summary of several approaches for Content based Web search

Approach	Factors	Solution	Limitations	Reference
Keyword based	Query content, Query semantic	Query based clusters	Query similarity estimation	[56]
Ontology based	Keywords, User logs	Keyword based clusters	Limitations of keywords	[57]
Heuristic based	Synonyms, Click stream, Queries	Fuzzy matching of Web queries	Distance estimation between queries	[58]
	Documents, Queries	Hybridization based user's similarity measures	Tuning of a member function	[59]
	User interest, favorites	Personalized information filtering	Straight forward to build reasoning system	[60]

Therefore, collected user access trends can be used in different applications, such as keeping track of a user's previously accessed websites. You can use these pages to identify the user's typical behavior and make predictions. Table-V summarizes the various Web search approaches.

Table-V. Summary of several approaches for Usage based Web search

Approach	Factors	Solution	Limitations	Reference
Log file based	Page view time	Page thresholding based session identification	Only one factor taken into account	[61]
Heuristic based	Session time, Links	Fuzzy clustering of user's session	Difficult to identify active user's session	[62]

C. Web Personalization and Profile of User

A profile of user is a list of a single user's personal information. A profile refers to the direct electronic representation of the features of an individual User profiles can be represented as a representation of a user model. The user profiles were generated for explanation of the user's context knowledge [47][48][49]. User profiles are the design models used by users when web data is collected.Users implicitly possess a concept model and are created from their knowledge of the context. Such expertise is used to collect relevant interest information and choices of a consumer. There are three types of user profiles: *interviewing, semi-interviewing, and non-interviewing.*

User profiles *interviewing* are known as ideal user profiles. These are obtained through the use of manual procedures like questionnaires, user interviews, etc. It is recommended that everyone read each document in these methods and give the document a positive or negative judgment on a particular topic.*Semi-interviewing* user profiles were created with limited user interaction through semi-automated techniques. For example, such methods usually give users a list of categories and ask users for interesting or not interesting categories. Users are not interested at all in *non-interviewing* methods, but instead explore client preferences. By observing user activity and behavior, they create user profiles and gain user background knowledge. In order to reflect user profiles, several models have been created. These models include regional or local knowledge base information. Market research uses the current knowledge base of the market and produces successful results. Common ontology like Word net, Thesaurus, Digital Libraries are the knowledge base that is widely used.

Store data that enables individual versatility is the common way to use a profile. This is called Content-based Filtering, which decides the relevance of the content when applied to a text document by comparing the keywords contained in a user profile with keywords extracted from the text [50]. User profiles are stored on the server to avoid network transmission of user profiles.Another powerful way to take advantage of user profiles is personal and collaborative filtering [51].This approach gathers a group of people's user profiles and creates similarity-based recommendations. To enforce cooperative filtering, all users' profiles must be linked, so centralizing them on the server is also the best storage location. Since all custom web applications must have access to the full set of profiles for a specific user, the user profile must be stored on the browser.

D. *Customized search based on Semantics*

The aim of Personalization is to define a subset of Web data that corresponds to a user or user group's interested profile. This can be done by recommending pages or filtering user-friendly web pages [52]. For example, this can be achieved, for instance, by evaluating user access to web data from historical data history, and Clustering user-relevant topics previously visited web pages based on content similarities. If a new website is found to be like one of the clusters, it can be sent to the user. Custom search defines content and user profiles using Semantic Web Standards (RDF and OWL).Web data access personalization based on semantics can be used effectively to improve accuracy and search recall, by re-ranking search results based on the past activities of the learner. Using Ontology is the cornerstone of the Web Personalization Semantic approach. Since web pages are annotated with entity tags, a user-accessed web pages will contribute to more active content recommendations. Table-VI provides the list of different web search approaches

Table-VI. Summary of various Web search concepts proposed in the literature

Concepts	Factors				Precision	Recall	Reference
	Query	Keyword	Link	User Logs			
Ontology	---	✓	---	✓	60.4%	70.2%	[57]
Density-based Clustering	✓	---	---	---	70.6%	78.3%	[56]
Fuzzy matching	✓	---	---	---	80.0%	83.2%	[58]
Hybrid Fuzzy Sets	✓	---	---	---	79.1%	81.4%	[59]
Neuro-fuzzy Systems	---	---	✓	✓	69.1%	87.0%	[62]
Fuzzy Neural Networks	✓	✓	---	---	70.2%	75.3%	[60]
Fuzzy classification	---	✓	✓	---	70.4%	78.7%	[63]
Statistical and Networks Theory	---	✓	✓	---	60.6%	68.3%	[64]
Neural Networks	---	---	---	✓	60.4%	73.2%	[65]
Page Thresholding	---	---	---	✓	36.4%	78.5%	[61]

VII. RECENT TECHNOLOGYICAL DEVELOPMENTS

According to Dr. Dave Chaffey, Digital Strategist and co-founder and Content Manager of Smart Insights marketing publisher and training network, existing personalization tools are divided under five categories: Personalization based analytics tools, personalization of retail, personalization of content or business management system, personalization of the B2B website as part of marketing automation and libraries and modules of personalization code. The five types of design options and suggestions are as follows.

Category 1. Personalization integrated with web analytics
Once internet personalization tools first emerged, they were part of expensive software systems that were only accessible to the largest companies, but today there are many services that are ideal for small businesses. These are two of the initial services available in a new form today and then a totally' new kid on the block.' Adobe Target, IBM product recommendations and Google Optimize are some of the examples.

Category 2. Personalization Software as Service (SaaS) for Ecommerce
The second category we are analyzing here is primarily for Ecommerce search and merchandising, which offers automated product placements using aggregated behavioral information and personal suggestions. Such solutions can be considered the top retail customization products for comparison if you are looking for a product separate from your ecommerce or content management system. Apptus, Barilliance - SaaS Personalisation for Ecommerce, Bunting Website Personalisation, Certona, ChoiceStream, Dynamic Yield, Evergage (includes B2B), Emarsys, Monetate Intelligent Personalization Engine, Omniconvert, Pure360 - Ecommerce personalization, Pureclarity - Ecommerce personalization, Qubit and Yusp are some of the example implementations.

Category 3. Personalization features available as part of Commerce Management Systems
According to Damien of Digicon's opinion, Episerver, Sitecore, Kentico and Adobe with Target have built-in tools to configure content based on different rules such as geo-location, search terms, links, lead scores and also provide more advanced user experience and profile-based configuration. Episerver has an interesting personalization approach which they describe as' Customization of no-rules: more knowledge, less work.' It applies machine learning and statistical analysis to user behavioral information to customize classification and product pages without the need for rules specification instead of conventional rules-based automation used by many of the specialist services.

Category 4. B2B, marketing automation, and publisher website personalization tools
Evergage is recommended as a tool that fits best in this category that has a wide range of options for serving personalization for content marketing in different placements on a site. Other alternatives for B2B personalization or non-retail consumer personalization are: BrightInfo (cost-effective), Idio (enterprise), Ion Interactive (enterprise), Sitespect (intermediate).

Category 5. Personalization code libraries and modules
As part of Amazon Web Services, Amazon made the Amazon Personalize service freely available at the end of 2018.

VIII. RESULTS AND DISCUSSIONS

Including the use of internet mining techniques that were not thoroughly explored in web customization, several other open issues pledge more directions for research. Some of these are referred to here.
- Most pattern discovery methods have a major problem with their difficulty in handling very large data

sizes, such as those usually collected on the Internet. The scalability of the pattern discovery approach can become essential in a web personalization system .Scalability tends to the level of time and memory increase criteria for a variety of factors like. Nevertheless, the computational performance of the customization unit is even more important, integrating the patterns discovered.

• Another requirement for learning methods for cloud personalization platform,their ability to change models incrementally.

• User behavior differs over time and the construction of models should be affected.

• In addition, there are a variety of open issues related to the use of configuration data derived.

• The need for clear representation of the data extracted, i.e. user models created by web-use mining tools, is a practical issue.

• Ultimately, an important issue is the lack of studies to assess their outputfor internet configuration systems. This is partly due to the difficulty of setting clear evaluation criteria.

IX. CONCLUSION

Web personalization is the process of modifying a website to the needs of specific users, using the knowledge gained from analyzing the user's navigational actions in aggregationwith other data collected in the Web context, namely structure, content and user profile data. The ability to analyze very large data sets efficiently using pattern discovery methods is essentialsince amount of facts collected in common web-based services exceeds.The introduction of internet use mining into the system of web personalization has presented many methodological and technical issues.Simultaneously, it has scarcely understood the strength of this synergy between the two systems. Therefore, it remains unexplored several interesting paths.

REFERENCES

1. G. Doddegowda B J, G T Raju "A Novel Approach for Web Personalization through Web Mining Techniques"An International Journal of Advanced Computer Technology,4(5), May-2015(Vol-IV, Issue-V).
2. M. Albanese, A. Picariello, C. Sansone, L. Sansone, "A Web Personalization System based on Web Usage Mining Techniques", in Proc. of WWW2004, May 2004, New York, USA.
3. B. Mobasher, H. Dai, T. Luo, Y. Sung, J. Zhu, Integrating web usage and content mining for more effective personalization", in Proc. of the International Conference on Ecommerce and Web Technologies (ECWeb2000), Greenwich, UK, September 2000.
4. Dhanalakshmi.D and Dr.J.Komala Lakshmi, " A survey on data mining research trends," International Journal of Engineering And Computer Science,ISSN:2319-7242, vol. 3, Issue 10 October, 2014,pp.8911-8919.
5. Raymond Kosala and Hendrik Blockeel, "Web mining Survey" SIGKDD Explorations, Copyright@2000.
6. Susan Gauch, Jason Chaffee, Alexander Pretschner. 2003. Ontology-Based User Profiles for Search and Browsing. The OBIWAN Project.
7. Feng Qiu, Junghoo Cho. 2006. Automatic identification of user interest for personalized search. WWW: 727-736. ACM SIGKDD, July 2000 Vol.2, Issue 1.
8. Maimon, Oded; Rokach, Lior (Eds.). 2008. Soft Computing for Knowledge Discovery and Data Mining. Hardcover Springer Science and Business Media, Inc. XIV, 434 p. 74 illus.
9. Koutrika, G., Ioannidis, Y.: "A Unified User Profile Framework for Query Disambiguation and Personalization","Workshop on New Technologies for Personalized Information Access", held in conjunction with the 10th International Conference on User Modeling, pp. 44–53 (2005).
10. Mladenic D., "Text-learning and Related Intelligent Agents: a Survey", IEEE Intelligent Systems, 14(4):44–54 (1999).
11. Goldberg, K., Roeder, T., Gupta, D., Perkins, C.: Eigentaste, "IT Constant Time Collaborative Filtering Algorithm. Information Retrieval", Journal, 4(2):133–151, (2001).
12. Paulakis, S., Lampos, C., Eirinaki, M., Vazirgiannis, M.: Sewep, "IT Web Mining System Supporting Semantic Personalization", in "15th European Conference on Machine Learning and 8th European Conference on Principles and Practice of Knowledge Discovery in Databases", LNCS, vol. 3202, pp. 552-554, Springer (2004).
13. Jones, G.J.F., Brown, P.J., "Context-Aware Retrieval for UbiquitousComputing Environments", Invited paper in Mobile and Ubiquitous Information Access, LNCS, vol. 2954, pp. 227–243. Springer (2004).
14. Yang, Y., Aufaure, M.A., Claramunt, C., "Towards a DL-Based Semantic User Model for Web Personalization", in "Third International Conference on Autonomic and Autonomous Systems", pp. 61-61. IEEE Computer Society (2007).
15. Kuhn, W., Handling Data Spatially, "Spatializing User Interfaces", in "7th International Symposium on Spatial Data Handling", Advances in GIS Research II, vol. 2, pp.13B.1-13B.23, IGU (1996).
16. Resnik, P., "Semantic Similarity in a Taxonomy: an Information-Based Measure and its Application to Problems of Ambiguity in Natural Language", Journal of Artificial Intelligence Research, 11:95–130 (1999).
17. Larson, R.R., Frontiera, P., "Spatial Ranking Methods for Geographic Information Retrieval in Digital Libraries", in Heery, R., Lyon, L., (eds.) ECDL. LNCS, vol. 3232, pp. 45–56. Springer (2004).
18. Maceachren, A.M., Kraak, M.J., "Research Challenges in Geovisualization, Cartography and Geographic", Information Science, 28(1):3–12 (2001).
19. BamshadMobasher,"WebPersonalizer:A Server Side Recommender System Based on Web usage Mining," downloaded from internet.
20. NicolaasMatthijis, Personalizing Web Search using Long Term Browsing History," WSDM'11, February 9-12,2011,Hong Kong, China, Copyright 2011 ACM 978-1-4503-0493-1/11/02.
21. OurdiaBouidghaghen, Lynda Tamine and MohandBoughanem, "Context-Aware User's Interests for Personalizing Mobile Search," 2011 12th IEEE International Conference on Mobile Data Management, 978-0-7695-4436-6/11, DOI 10.1109/MDM.2011.51.
22. .D.Dhanalakshmi, R.Kousalya,V.Saravanan, "Time based Web User Personalization and Search," International Journal of Computer Applications (0975 –8887) Vol. 46, No.23, May 2012.
23. Thi Thanh Sang Nguyen, Hai Yan Lu, and Jie Lu, "Web-Page Recommendation Based on Web Usage and Domain Knowledge, "IEEE Transactions on knowledge and data engineering, Vol. 26, No. 10,october 2014.
24. VeeramalaiSankaradass and Kannan Arputharaj,"An Intelligent Recommendation System for Web User Personalization with Fuzzy Temporal Association Rules," European Journal of Scientific Research Vol. .51 No.1, 2011, pp.88-96.
25. KennethWai-Ting Leung, DikLun Lee, Wang-Chien Lee, "Personalized Web Search with Location Preferences," ICDE Conference 2010.
26. Yunbo Cao, Jun Xu, Tie-Yan Liu, Hang li, Yalou Huang and Hsiao-WuenHon,"Adapting Ranking SVM toDocument Retrieval", SIGIR'06, August 6–11, 2006, Seattle, Washington, U.S.A.
27. Shady Shehata , Fakhri Karray and Mohamed S. Kamel, "An Efficient Concept-Based Mining Model for Enhancing Text Clustering," IEEE Transactions on knowledge and data engineering, vol. 22, no.10, October 2010.
28. Xiao-Bing Xue and Zhi-Hua Zhou, "Distributional Features for Text Categorization," IEEE Transactions on knowledge and data engineering,vol.21, no.3, March 2009.
29. Claudia Marinica and Fabrice Guillet,"Knowledge-Based Interactive Post mining of Association Rules Using Ontologies," IEEE Transactions on knowledge and data engineering, vol. 22, No. 6, June 2010.
30. S.Sendhilkumar and T. V. Geetha, "Concept based personalized Web Search,"Advances In Semantic Computing (Eds. Joshi, Boley&Akerkar), TMRF e-Book, Chapter 5, vol. 2, pp 9-102, 2010.
31. Tak-Lam Wong and Wai Lam,"Learning to Adapt Web Information Extraction Knowledge and Discovering New Attributes via a Bayesian Approach",IEEE Transactions on knowledge and data engineering, Vol. 22, No. 4, 2010.

32. Mohammed Kayed and Chia-Hui Chang," FiVaTech: Page-Level Web Data Extraction from Template Pages," IEEE Transactions on knowledge and data engineering, Vol. 22, No. 2, February 2010.

33. Sumaiya Kabir, Shamim Ripon, Mamunur Rahman and TanjimRahman,"Knowledge- Based Data Mining Using Semantic Web," 2013 International Conference on Applied Computing, Computer Science, and Computer Engineering, IERI Procedia 7, 2014, pp. 113 – 119.

34. Konstantin Todorov, "Data Mining, Ontologies and the Semantic Web", April 2013, Online tutorial.

35. M. Jordan Raddick, Ani R. Thakar, and Alexander S. Szalay and Rafael D.C. Santos, "Ten Years of SkyServerI:Tracking Web and SQL e-Science Usage," July/August 2014,www.computer.org/cise, Copublished by the IEEE CS and the AIP.

36. Pretschner, A. and Gauch, S. 1999. Ontology Based Personalized Search. Proc. 11th IEEE Int'l Conf. Tools with Artificial Intelligence (ICTAI '99), pp. 391-398.

37. Chirita, P.-A., Nejdl, W., Paiu, R. and Kohlschu C.¨ tter. 2005. Using ODP Metadata to Personalize Search. Proc. 28th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR '05), pp. 178-185.

38. Liu, F., Yu, C. and Meng, W. .2002. Personalized Web Search by Mapping User Queries to Categories. Proc. ACM Int'l Conf. Information and Knowledge Management (CIKM '02), pp. 558-565.

39. Liu, F., Yu, C. and Meng, W. Jan. 2004. Personalized Web Search for Improving Retrieval Effectiveness. IEEE Trans. Knowledge and Data Eng., vol. 16, no. 1, pp. 28-40.

40. Shen, X., Tan, B. and Zhai, C. 2005. Implicit User Modeling for Personalized Search Proc. ACM Int'l Conf. Information and Knowledge Management (CIKM '05), pp. 824-831.

41. Tan, B., Shen, X. and Zhai, C. 2006. Mining Long-Term Search History to Improve Search Accuracy. Proc. 12fth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD '06), pp. 718-723.

42. Teevan, J., Dumais, S.T., and Horvitz, E. 2005. Personalizing Search via Automated Analysis of Interests and Activities. Proc. 28th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR '05), pp. 449-456.

43. Sugiyama, K., Hatano, K. and Yoshikawa, M. 2004. Adaptive Web Search Based on User Profile Constructed without Any Effort from Users. Proc. 13th Int'l World Wide Web Conf. (WWW '04), pp. 675-684.

44. Chirita, P.A., Firan, C. and Nejdl, W. 2006. Summarizing Local Context to Personalize Global Web Search. Proc. ACM Int'l Conf. Information and Knowledge Management (CIKM).

45. L Brin, S., Motwani, R. and Winograd, T.1998 .The PageRank Citation Ranking: Bringing Order to the Web.Technical report, Computer Science Dept., Stanford Univ.

46. Sun, J.-T., Zeng, H.-J., Liu, H., Lu, Y. and Chen, Z. 2005. CubeSVD: A Novel Approach to Personalized Web Search. Proc. 14th Int'l World Wide Web Conf. (WWW '05), pp. 382-390.

47. S. Gauch, J. Chaffee, and A. Pretschner, "Ontology-Based Personalized Search and Browsing" Web Intelligence and Agent Systems, vol. 1, nos. 3/4, pp. 219-234, 2003.

48. Y. Li and N. Zhong, "Web Mining Model and Its Applications for Information Gathering " Knowledge-Based Systems, vol. 17, pp. 207-217, 2004.

49. Y. Li and N. Zhong, "Mining Ontology for Automatically Acquiring Web User Information Needs" IEEE Trans. Knowledge and Data Eng., vol. 18, no. 4, pp. 554-568, Apr. 2006.

50. Morita M., Shinoda, Y., "Information Filtering Based on User Behaviour Analysis and Best Match Retrieval", in Proceedings of the 17th International ACM-SIGIR Conference on Research and Development in Information Retrieval, 1994, pp. 272-281.

51. Shardanand U., Pattie M., "Social Information Filtering: Algorithms for Automating "Word of mouth", in Proceedings of the Human Factors in Computing System, Denver, May 1995, pp. 210-217.

52. Xiaohui Tao, Yuefeng Li, and NingZhong, "Senior Member, IEEE, "A Personalized Ontology Model for Web Information Gathering", IEEE Transactions On Knowledge and Data Engineering, VOL. 23, NO. 4, APRIL 2011.

53. Dhanalakshmi.D, Dr. J.Komala Lakshmi "A Survey on Web User Personalization Techniques" 978-1-4799-3975-6/14 ©2014 IEEE.

54. HimaniArya ,Jaytrilok Choudhary, Deepak Singh Tomar "A Survey on Techniques for Personalization of Web Search" International Journal of Computer Applications (0975 – 8887) Volume 94 – No. 18, May 2014.

55. Selvakumar K, S Sendhilkumar, "Challenges and Recent Trends in Personalized Web Search : A Survey ", 978-1-4673-0671-3/11 2011 IEEE.

56. C. Lucchesez, S. Orlandoy, R. Peregoz, F. ilvestriz, G. Tolomeizy,"Detecting Task-based Query Sessions Using Collaborative Knowledge", 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology 2010.

57. Nisha Pahal, Payal Gulati, Parul Gupta, "Ontology Driven Conjunctive Query Expansion based on Mining User Logs", International Conference on Methods and Models in Computer Science, 2009.

58. Tao Cheng, Hady W. Lauw, Stelios Paparizos, "Fuzzy Matching of Web Queries to Structured Data", ICDE Conference 2010.

59. Chunlai Chai, Biwei Li, "A Novel Association Rules Method Based on Genetic Algorithm and Fuzzy Set Strategy for Web Mining", JOURNAL OF COMPUTERS, VOL. 5, NO. 9, SEPTEMBER 2010.

60. Dai Xuewu, Vic Grout, Tang Haokun and Li Jianguo, "NEURAL NETWORKS-BASED MULTI-INTEREST INFORMATION FILTERING", International Conference on Web Intelligence and Intelligent Agent Technology 2010.

61. Fang Yuankangi, Huang Zhiqiui, "A Session Identification Algorithm Based on Frame Page and Pagethreshold", 2010.

62. G. Castellano, A.M. Fanelli and M.A. Torsello, "NEWER: A system for NEuro-fuzzy Web Recommendation", Journal of Applied Soft Computing vol-11 PP-793-806, 2011.

63. Debajyoti Mukhopadhyay, Pradipta Biswas, Young-Chon Kim, "A Syntactic Classification based Web Page Ranking Algorithm", 6th International Workshop on MSPT Proceedings MSPT 2006.

64. Yi Zeng, Erzhong Zhou, Yulin Qin, Ning Zhong, "Research Interests: Their Dynamics, Structures and Applications in Web Search Refinement", IEEE/WIC/ACM International Conference on Web 2010

65. Cheng Chih Chang, Pei-Ling Chen, Fei-Rung Chiu and Yan-Kwang Chen,"Application of neural networks and Kano's method to content recommendation in Web personalization", Journal of Expert systems with applications PP- 5310-5316 2009.

AUTHORS PROFILE

Mrs. Sowbhagya M P, Research Scholar pursuing PhD at VTU RRC, working as Assistant Professor, Department of CSE, City Engineering College, Bengaluru. She has completed M.E in Web Technology from UVCE, Bengaluru.She qualified in GATE-2008, Her research interests are Web Mining and Applications, Pattern Recognition.

Ms. Ganavi K R, SJB Institute of Technology, Bengaluru. She has done Internship on Mobile App Development, and Augmented Reality at UtthungaTechnologies, Bengaluru. Her interests include Data Science and Machine Learning.

Dr. Yogish H K is currently working as Professor and Head, Department of MCA, Ramaiah Institute of Technology, Bengaluru. He has published more than 50 research papers in prominent National and International Journals. His research areas are Big Data, Data mining and Data Warehousing and Software Engineering.

Retrieval Number: B12271292S19/2019©BEIESP
DOI: 10.35940/ijitee.B1227.1292S19

740

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication