

A Comparative Study of Web Page Personalization Techniques of Search Engines

Mohd Suaib
Department of Computer Science
Integral University, Lucknow
Email: suaib [AT] iul.ac.in

Mohammadi Akheela Khanum
Department of Computer Science
Integral University, Lucknow
Email: akheelakhanam [AT] iul.ac.in

Abstract— We have been witness to the fact that in the recent year there is an abysmal increase in the volume of digital data and internet users. As the size of internet based information grows exponentially, the voluminous raw data need to be fetched smartly in an efficient manner. There are several approaches and schemes to obtain, tailor and use digital data as per our requirement.. It is increasingly difficult for search engines to know what the web user is intended to obtain. This is an emerging and interesting area where the data mining techniques is being applied. Web mining is an special area of data mining, which works on web content mining, web usage mining and web structure mining. Ranking Web aims at providing the search results of WebPages the user look for by assigning the ranks over pages which are relevant and most concerned to the query. Different metrics have been used in various search works to rank web pages according to their qualitative measures. Through web usage analysis we can effectively improve the ranks of web pages in line with the user's preferences and need. In this paper, various techniques in search engine personalization have been used and concluded. Finally the findings on web page ranking approaches in the current scenario are analyzed & elaborated and what is expected on future work is discussed thereafter.

Keywords—search engine, web personalization, Web mining, Data Mining, Page Ranking

I. INTRODUCTION

A number of studies have shown that a vast majority of queries to search engines are short and under-specified and users may have completely different intentions for the same query. For example, a real-estate agent may issue the query “office” to look for a vacant office space, while an IT specialist may issue the same query to look for popular Microsoft productivity software. To address these differences among the users, there has been a recent surge of interest in personalized search to customize search results based on a user's interest. Given

the large and growing importance of search engines, personalized search has the potential to significantly improve user experience.

Unfortunately, studies have also shown that the vast majority of users are reluctant to provide any explicit feedback on search results and their interest. Therefore, a personalized search engine intended for a large audience has to learn the user's preference automatically without any explicit input by the users.

II. WEB MINING

A. An Overview

Web Mining is the use of data mining techniques which extracts interesting unknown patterns from the web. Nowadays everything is digitized. In this digital era, people mostly depend on the internet for their information need. Web mining is one of the emerging research fields due to the vast growth of the web as well as users of the Web. One of the important goals for the research community is, providing the best service to the users.

B. Web Personalization

Web usage patterns are extracted in any one of the sources or combination of one or more such as server web log files, client side cookies and ISP or transaction database. Most of the researches make use of server web log files for pattern extraction. [3] Provides web personalization, in which data are extracted from server log files. After completing the preprocessing task, with the help of user transactions and server URL cluster is created using multi-variant k-means method. In a web site, when a user requests a URL in an active session, the given URL is compared with URL Clusters. Accordingly a set of recommended URL's are sent along with the requested URL.

The author's [4] Nicolaas Matthijs and Filip Radlinski proposed personalization approach by analyzing user's long term browsing behavior. The user's profile which is used to re-rank web search results according to their preferences. In this work, they are considering entire visited web page instead of web snippets for extracting concepts using C/NC method and

viterbi algorithm. While re-ranking the search results, the terms in the profile are compared with terms in the web-snippets according to that, scores are assigned for the snippets. TF, TF-IDF and personalized B25 weighting approaches were used for computing weighting for each term.

Research paper [5] proposed mobile user personalization for information retrieval based on the factors such as time, location interest of the user. Graph-based ontological representation is used to create user profile from web usage pattern of a user at a particular time and location. The past user profile is compared with new submitted query to match the relevant user profile. Based on that, research results from reputed search engine will be re-ranked.

Research paper [6] based on web user personalization with content, location and time preference. In which, user submitted query is analyzed to know how much content and location concepts associated at a particular time then ontology will be created for content concept and location concept. Concepts are extracted based on the frequency of a keyword that exists in a web snippet. From the user clicked results click content entropy and location entropy will be created to show how much the user interested in particular content or location concepts. User profile will be created based on ontology and entropy. This research makes use of Ranking SVM for learning user preference.

Research [7] Mobile user's needs are depend on two factors such as time and experience. Interests of users are not static, it is changing from time to time along with needs. To create dynamic profile with the consideration of time-zones and user experiences XML seems to be the best because of extensibility as well as a way to standardize the profile through the incorporation of XML schemas.

Research paper [8] focuses on search engine personalization. This work is based on both positive and negative preferences of user. Concepts are extracted from user's click through web snippets to build concept-based user profiles automatically. To achieve this, preference mining rules are applied. The user profiling strategies were evaluated and compared with the personalized query clustering method.

Research paper [9] Ontologies and the Semantic Web are two important research fields that are beginning to receive great attention. Semantically analyzing the web content as well as the user's query is more important in the personalization domain. Language Modeling and Question Answering are two important Natural Language Processing (NLP) research areas that could lead to break-through in the development of personalized search systems. New search engines based on these technologies may be able to understand the users' intention through the analysis of user-supplied natural language questions. They may be able to better understand keywords in the queries by recognizing various sentence types, analyze syntax, and disambiguate word senses in context. As a result, search results will be more accurate, satisfactory, and reliable.

Research paper [10] is too focus for personalizing web search by extracting concepts for user given query and for user clicked result from web snippets. By this method, personalized results were displayed. Concepts are extracted only from web-snippets not from web pages due to the reasons such as huge volume of information and time consuming.

Research paper [11] focused on personalized travel recommendation. Freely available community-contributed photos are the primary source to achieve this task. From which user specific profiles or attributes (e.g., gender, age, race) as well as travel group types (e.g., family, friends, couples) are automatically retrieved. User profile and context information were detected from mobile sensors. This work makes use of probabilistic Bayesian learning framework to recommend the next travel location from his/her current location or even delivering context-related advertisements or services.

III. APPLICATION OF WEB USAGE MINING

Research work [12] recommends web page for users in a web site with the help of domain knowledge and user preference. Ontologies are constructed for domain and user preferences. Research paper [13] Users interest may change over time. Their interests are dynamic. This paper focuses on mining users preferences to build user profile based on temporal pattern analysis using temporal rule mining. This paper proposed a new algorithm called Fuzzy-Temporal Association Rule Mining Algorithm (FTARM). Research paper [14] Mining user preferences and query associated information in terms of concepts. It consists of two types of concepts such as content and location concepts. Personalization is performed using ontology based approach. Research paper [15] uses Ranking SVM for Document retrieval in Information Retrieval System. In IR, queries are used to retrieve documents. This research focuses on how to intensify training for top-ranked documents and training performed with queries are less relevant with documents using RSVM with Hinge loss function and two parameters such as gradient descent and quadratic programming for obtaining maximum-margin classification. Research paper [16] proposed mining model consists of sentence-based concept analysis, document-based concept analysis, corpus-based concept-analysis, and concept-based similarity measure. This work identifies terms which are highly convey the semantics of sentence, document and corpus levels rather than the traditional analysis of the document only. The similarity between documents is calculated based on the extracted concepts using a new concept-based similarity measure (ctf). Documents are clustered based on that similarity measure.

Research paper [17] represented text-categorization, using distributional features in which a novel values are assigned to a word is called distributional features. Distributional features in the sense, compactness of the appearance of the word and the position of the first appearance of the word. In this research, tfidf-style equation is constructed and ensemble learning

technique is used. It considers frequency of a word i.e., how many times a word exists in a document and where the word appears at the very first time and compactness of words. In this research, just combines existing frequency with the distributional features to improve performance with little additional cost. This paper [18] proposes a new interactive post processing approach, ARIPSO (Association Rule Interactive postProcessing using Schemas and Ontologies) to prune and filter discovered rules. They proposed and used Domain Ontologies and Rule Schema formalism. With the integration of domain knowledge, user knowledge and rule schema the number of rules can be reduced and it leads to set of effective rules.

This [19] work did personalization based on analyzing user's unvisited pages that contain relevant information to the user. It constructs domain ontology and creates user profile based on PERSONALISED PAGE VIEW (PPV) GRAPH from the concepts extracted using tf-idf. It recommends user interested/preferred information in the shortest path.

The work [20] is focused how to extract information and identify new attributes in the new unknown sites using the already defined information extraction knowledge such as wrapper using Bayesian learning and expectation-maximization (EM) techniques. In [21] this paper, proposed an unsupervised, page-level data extraction approach to figure out the structure description and templates for each individual Deep Website, which contains either singleton or multiple data records in one Webpage. FiVaTech record-level extraction system applies to tree matching, tree alignment and mining techniques to achieve the challenging task.

Research [22] is trying to provide the knowledge based mining instead of data mining using ontological approach. Through ontology, they are trying to create semantic web in which data are arranged in structured manner which is helpful to acquire most relevant and accurate results. But the construction of ontologies needs manual effort. Research paper [23] gives an overall view of semantic web and application of Data mining in semantic web and Ontologies.

Research paper [24] discuss about how to analyze web logs and how to use web analytics tools. It also provides a case-study of a SDSS sky server and the normalized data set of sky server web log. Hengshu Zhu [25] proposed a novel method for automatically classifying mobile apps with the help of web knowledge and user contextual information.

In research paper [27] authors proposed an approach based on GA, where a collection of webpage is used to represent chromosome in web usage mining problem.

Research paper [28] discusses a new material recommender system framework based on sequential pattern mining and multidimensional attribute-based collaborative filtering (CF) is proposed. In this paper modified Apriori and PrefixSpan algorithms are implemented to discover latent patterns in accessing of materials and use them for recommendation. Leaner Preference Tree (LPT) is introduced to take into account multidimensional-attribute of materials, and learners' rating and model dynamic and multi-preference of learners in the

multidimensional attribute-based CF approach. In research paper [29] the authors have proposed a method consists of two steps: (1) discovering content-related item sets using item-based collaborative filtering (CF), and (2) applying the item sets to sequential pattern mining (SPM) algorithm to filter items according to common learning sequences.

IV. DISCUSSIONS

Information retrieval got a great attention in research communities due to this digital era. In real scenario, users are not aware of the exact keyword for their information retrieval process. Due to this, they are getting both relevant and irrelevant results. In some cases without knowing the exact keyword users are getting irrelevant results. Today's search engines are more powerful. They analyze semantic relationship between words and use various glossaries trying to provide the best results to their users. Even though, there is gap in between the human and computer. In the sense that, analyzing and understanding differs from human to machine. The machine should analyze and understand the information need of users and provide the best service to the users. From this study we conclude that the integration of personalization, knowledge based mining and semantic web are bring the traditional web scenario into web 3.0. Current and ongoing researches are making use of ontologies. Table 1 show that, most of the authors have preferred ontologies for personalization. With the help of ontologies, researchers are trying to make the machine to think and analyze like a human being in the field of web usage mining. Concepts are very important to construct ontologies because it clearly explains the contents of a web documents. Exact concepts are used to increase the accuracy of the information retrieval process.

From this survey, we conclude that to personalize the website and to extract the most relevant concepts from analyzing the entire documents using NLP. In offline, when we analyze the entire documents and constructing domain knowledge ontologies with extracted concepts is not affecting the response time of an Information Retrieval system. By incorporating personalization with this approach, we can provide the required and effective result to the user. In Information retrieval, providing the relevant and required information to the user is most important. To do this, the machine should know the following such as what the user actually means and where that occurs, how it occurs. Ontologies and Natural Language Processing (NLP) techniques are very much powerful and helpful to achieve the above mentioned task. These are the findings with the study of twenty five research papers.

V. CONCLUSION & FUTUREWORK

The importance of web personalization in E-Government services has been described by the authors in [26] and they have also emphasized on how the users will react to acceptance of information technology in E-Government domain.

From the study, we conclude that incorporating the information about the usage history of the users can play an

important role in providing the efficient ranking of web pages based on the user's preference. In order to give the better personalization result to the user, machines should know what about the web documents and pages deals and what, the users are requesting. These two facts should be understood by the machine. For this, machine needs to think, understand and decide like human beings. For achieving this, ontologies and natural language processing techniques are supportive.

TABLE1. Various techniques for web user personalization

S.No	Approach used	Goal	Author(s)
1	Multi-variant k-means methods	Personalized URL Recommendation	BamshadMobasher
2	CNC, TF, TF-IDF and personalized B25 weighting	Personalized search results from search Engine	NicolaasMatthijis
3	Conceptual Term Frequency	Document Clustering	OurdiaBouidghaghen et.al
4	Graph based Ontological similarity Measure	Personalization for Mobile Users	D.Dhanalakshmi et.al
5	Ontology representation and RSVM	Personalization for Mobile users	Christoforos Panayiotou et.al
6	XML Schema	Personalization for Mobile users	Kenneth Wai-Ting Leung and DikLun Lee
7	Natural Language Processing, Verb argument structure, Conceptual term frequency	Document Clustering	Alessandro Micarelli et.al
8	Bipartite graph, Personalized Agglomerative Clustering	Concept based user profiles	Jie Yu and Fangfang Liu
9	Bayesian learning	Recommendation of next travel location	Yan-Ying Chen et.al
10	Association Rule Interactive post-Processing, Schemasing Ontologies	Interesting Association Rules	Thi Thanh Sang Nguyen
11	Ontology, Personalized page view (PPV) graph	Personalization (Preferred information in the shortest path)	VeeramalaiSankaradass and Kannan Arputharaj
12	Fuzzy-Temporal Association Rule Mining	Dynamic Personalization	Yunbo Cao, Jun Xu et.al
13	Bayesian learning and expectation-maximization (EM) techniques	Extracting Information from unseen websites	Shady Shehata , FakhriKarray
14	frequent patterns	Personalized search results from search Engine	Xiao-Bing Xue and Zhi-Hua Zhou
15	tfidf+ distributional features	Document Clustering	Claudia Marinica and Fabrice Guillet
16	tree templates and schema	Extracting data from web pages	S.Sendhilkumar and T. V. Geetha
17	ontology-based, multi-facet (OMF) profile, RSVM	Personalized search results from search Engine	Tak-Lam Wong and Wai Lam

18	Ontology, semantic mining	Knowledge based information	Mohammed Kayed and Chia-Hui Chang
19	Ontology	Recommendation of web page, Personalization	Konstantin Todorov
20	Normalization	Web log Analysis	M. Jordan Raddick, Ani R. Thakar et.al
21	VSM, LDA model, Gibbs Sampling, BP-Growth Algorithm ,Maximum Entropy and Limited- Memory BFGS	classification	Hengshu Zhu, Enhong Chen et.al
22	Genetic Algorithm	Prioritizing Web Links Based on Web Usage andContent Data	K. Chaudhary and S. Gupta
23	Sequential Pattern Mining and Attribute Based Collaborative Filtering	Personalized Recommendation of Learning Material	M Salehi, IN Kamalabadi and MBG Ghoushchi
24	collaborative filtering (CF), sequential pattern mining (SPM)	Recommend learningitems in users' learning processes.	W Chen, Z Niu, X Zhao and Y Li

REFERENCES

- [1] 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.
- [2] Raymond Kosala and Hendrik Blockeel," Web Mining Survey" SIGKDD Explorations, Copyright@2000 ACM SIGKDD, July 2000 Vol.2, Issue 1-Page.
- [3] BamshadMobasher,"Web Personalizer:A Server Side Recommender System Based on Web usage Mining," downloaded from internet.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] Christoforos Panayiotou, Maria Andreou, George Samaras and Andreas Pitsillides, "Time Based Personalization for the Moving User," Proceedings of the International Conference on Mobile Business (ICMB'05)0-7695-2367- 6/05,IEEE.

- [8] Kenneth Wai-Ting Leung and DikLun Lee, "Deriving Concept-Based User Profiles from Search Engine Logs," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, No. 7, July 2010.
- [9] Alessandro Micarelli, Fabio Gasparrini, Filippo Sciarone and Susan Gauch, "Personalized Search on the World Wide Web", P. Brusilovsky, A. Kobsa, and W. Nejdl (Eds.): *The Adaptive Web*, LNCS 4321, pp. 195–230, 2007, © Springer-Verlag Berlin Heidelberg 2007.
- [10] Jie Yu and Fangfang Liu, "Mining user context based on interactive computing for personalized web search," *2nd International Conference on Computer Engineering and Technology*, 2010, Vol. 2, pp. 209-214.
- [11] Yan-Ying Chen, An-Jung Cheng and Winston H.Hsu, "Travel recommendation by mining people attributes and travel group types from community-contributed photos," *IEEE Transactions on Multimedia*, vol. 15, no. 6, October 2013.
- [12] 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.
- [13] Veeramalai Sankaradass 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.
- [14] Kenneth Wai-Ting Leung, DikLun Lee, Wang-Chien Lee, "Personalized Web Search with Location Preferences," *ICDE Conference* 2010.
- [15] Yunbo Cao, Jun Xu, Tie-Yan Liu, Hang Li, Yalou Huang and Hsiao-Wuen Hon, "Adapting Ranking SVM to Document Retrieval", *SIGIR '06*, August 6–11, 2006, Seattle, Washington, U.S.A.
- [16] 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.
- [17] 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.
- [18] 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.
- [19] S. Sendhil Kumar and T. V. Geetha, "Concept based personalized Web Search," *Advances in Semantic Computing* (Eds. Joshi, Boley & Akerkar), TMR e-Book, Chapter 5, vol. 2, pp 79- 102, 2010.
- [20] 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.
- [21] 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.
- [22] Sumaiya Kabir, Shamim Ripon, Mamunur Rahman and Tanjim Rahman, "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.
- [23] Konstantin Todorov, "Data Mining, Ontologies and the Semantic Web", April 2013, Online tutorial.
- [24] M. Jordan Raddick, Ani R. Thakar, and Alexander S. Szalay and Rafael D.C. Santos, "Ten Years of SkyServer I: Tracking Web and SQL e-Science Usage," July/August 2014, www.computer.org/cise, Co-published by the IEEE CS and the AIP.
- [25] Hengshu Zhu, Enhong Chen, Hui Xiong, Huanhuan Cao and Jilei Tian, "Mobile app classification with enriched contextual information," *IEEE Transactions on Mobile Computing*, vol. 13, no. 7, July 2014.
- [26] V. Krishnaraju, S. Mathew, V. Sugumaran, "Web Personalization for User Acceptance of Technology: An Empirical Investigation of E-government Services", *Information Systems Frontiers* pp. 1-17, 2016.
- [27] K. Chaudhary and S. Gupta "Prioritizing Web Links based on Web Usage and Content Data." In *Issues and Challenges in Intelligent Computing Techniques (ICICT)*, International Conference on IEEE, pp. 546-551, 2014.
- [28] M Salehi, IN Kamalabadi and MBG Ghouschi "Personalized Recommendation of Learning Material using sequential pattern mining and attribute based Collaborative Filtering." *Education and Information Technologies* 19 Springer , no. 4 pp. 713-735, 2014.
- [29] W Chen, Z Niu, X Zhao and Y Li "A hybrid Recommendation Algorithm adapted in e-learning environments." *World Wide Web* 17, no Springer. pp 2 271-284, 2014.
- [30] J Lu, D Wu, M Mao, W Wang and G Zhang. "Recommender System Application Developments: a Survey." *Decision Support Systems Elsevier* 74 pp. 12-32 , 2015.
- [31] C.J. Carmona, S. Ramirez-Gallego, F. Torres, E. Bernal, M.J. del Jesus, S. Garcia "Web Usage Mining to improve the design of an E-Commerce Website: OrOliveSur.com." *Expert Systems with Applications* 39 Elsevier, pp. 12 11243-11249, 2012.
- [32] K. Chaudhary and S. Gupta "Web Usage Mining Tools & Techniques: A Survey." *International Journal of Scientific & Engineering Research* 4, no. 6, 2013.
- [33] J Xu, K. Waharmann and S Li "The development, Status and Trends of Recommender Systems: A Comprehensive and Critical Literature Review." In *Proceedings of International Conference Mathematics and Computers in Science and Industry*, pp. 117-122, 2014.