Middle-East Journal of Scientific Research 24 (11): 3475-3478, 2016
ISSN 1990-9233

DOI: 10.5829/idosi.mejsr.2016.3475.3478

Personalized Recommender System for Improving Web Search

M. Geetha, Rittu Bhakuni, T. Sangavi and B. Tulasi

Department of Computer Science and Engineering,
Rajalakshmi Institute of Technology, Chennai, India

Abstract: Recommendation systems can be developed by taking the advantage of semantic reasoning capabilities to overcome common limitations of current systems and improve the recommendations' quality. In this paper, we present a personalized-recommendation system, a system a system that produces exact results based on previous search results and user interest. The proposed system uses domain ontology to enhance the personalization, on the one hand, user's interests are modelled in a more effective and accurate way by applying a domain-based inference method. The stemmer algorithm used by our content-based filtering approach, which provides a measure of the affinity between an item and a user, is enhanced by applying a semantic similarity method.

Key words: Query clustering · Search history · History grouping

INTRODUCTION

Web mining is the use of data mining techniques to automatically discover and extract information from Web documents and services. Web Usage Mining plays an important role in recommender systems and web personalization. Web navigation refers to the process of navigating a network of information resources in World Wide Web, which is organized as hypertext and hypermedia. Finding particular relevant information from a large collection of data is not an easy task, since many users are frequently faced with a huge number of alternative links that can be followed. Hence, it would be useful and helpful for the users if they were provided with a tool that can assist them in performing the Web navigation processes [1]. Human errors and human time wastage during Web navigation are the two major challenges that have not yet been fully solved. Certain Web tasks may be easily automated if a Website oriented client application is programmed to achieve the user's aims, emulating the user in the navigation process of the Web site. But this may not fulfill the user requirement exactly [2]. Chasing hyperlinks to find relevant information may be daunting .A learning system, cognizant of a user's interests, can be employed to automatically search for and retrieve relevant information by following appropriate hyperlinks. The design of such a learning system for automated Web navigation is done using adaptive dynamic programming methods. The availability of knowledge base and the use of less powerful approximator such as linear mapping becomes an issue [3]. Many mathematical models such as markov model have been widely used to represent and analyze user web navigation data. These model is not feasible to incorporate within an adaptive Web site platform in order to assess it as an online tool to provide link suggestions for users navigating within a site [4]. Predicting users' future requests in the World Wide Web can be applied effectively in many important applications from web search to personalization systems. These applications has traditional tradeoffs between modeling complexity and prediction accuracy [5]. The above proposed methods are either time consuming or has many complexities in implementation techniques. The functional convenience can be improved by higher task completion rate and less time for given tasks.

Existing System: In the existing system there are three basic steps followed. The first step is the actual usage pattern extraction. To extract actual user behavior web server logs and client log data are used. Web server log data are our data source in which each entry contains the IP address of the originating host, timestamp, requested web page and other data. The conversion of raw data to

Corresponding Author: M. Geetha, Department of Computer Science and Engineering, Rajalakshmi Institute of Technology, Chennai, India.

usage pattern involves the following step. The first and foremost step is the data preparation and preprocessing. The next step is user session identification in which activity record of each user is segmented into sessions, with each representing a single visit to a site. The final step is path. A tool is developed to automate the above mentioned steps expect path completion. The next step is the anticipated pattern extraction.

To capture anticipated user behavior we use cognitive user models. ACT-R cognitive user model provides detailed and sophisticated process models of human performance in interactive tasks with complex interfaces. Due to the complexity of ACT-R model development and low level rule based programming language in which it relies on a new cognitive architecture called Ideal User Interactive Path model was constructed.

IUIP model consist of number of states and transitions. Particular goal can be achieved by sequence of related operation rules specified by a series of transitions. This model need to be constructed individually for novices and experts by cognitive experts by utilizing their domain expertise and their knowledge of different users interactive behavior.

The final step is the usability problem identification. The extracted actual user navigation pattern is compared with corresponding IUIP model. This comparison yields a set of deviation result. The common usability problems can be found by focusing on deviations that occur frequently. The deviation can be classified into two types logical deviation and temporal deviation .

The performance of existing system depends on the sizes of training datasets. The bigger the training dataset size is, predicted pages are limited within the discovered Web access sequences. So we propose a new methodology to overcome these problems.

Proposed System: In proposed system, we present a personalized-recommendation system as shown in Fig. 1, a system that makes use of representations of items and user-profiles based on ontology's in order to provide semantic applications with personalized services. The semantics method achieved by using two different methods. A domain-based method makes inferences about user's interests and a taxonomy-based similarity method is used to refine the item-user matching algorithm, improving overall results. The recommender system proposed is domain-independent, which is implemented as a Web service and uses both explicit and implicit feedback-collection methods to obtain information on user's interests.

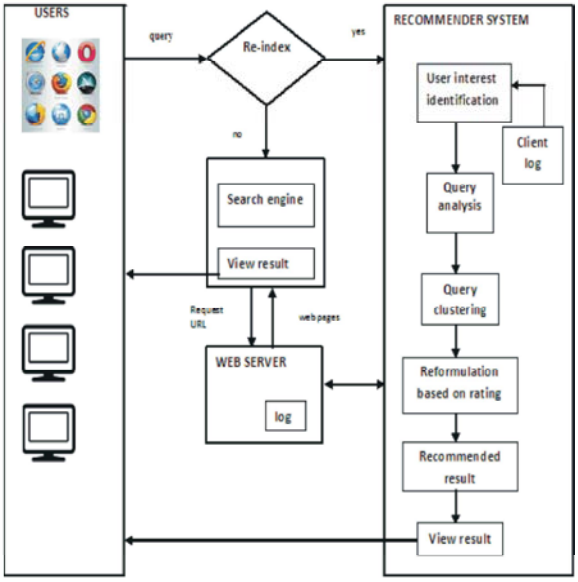

Fig. 1: Personalized Recommender System

Proposed recommender system is based on ontology and Web Usage Mining. The first step of the approach is extracting features from web documents and constructing relevant concepts. The next step is to build ontology for the web site use, which includes the concepts and significant terms extracted from documents. According to the semantic similarity of web documents we cluster them into different semantic themes, the different themes imply different preferences. The system can be divided into four major modules.

Creating Search History: Any personal documents such as browsing history and emails on a user's computer could be the data source for user profiles. This focus on frequent terms and limits the dimensionality of the document set, which further provides a clear description of users' interest. This module allows the search engine to better understand a user's session and potentially tailor that user's search experience according to her needs. Once query groups have been identified, search engines can have a good representation of the search context behind the current query, using queries and clicks in the corresponding query group.

Query Clustering: User's queries can be classified into different query clusters. Concept-based user profiles are employed in the clustering process to achieve personalization effect. The most similar pair of concept nodes and then, merge the most similar pair of query nodes and so on. Each individual query submitted by

Table 1: Comparison of traditional search engine with personalized recommender system

Parameter	Traditional Search Engine	Personalized Recommender System
Mining technique	Web structure mining	Web usage mining
Accuracy of result	Retrieves all related web pages	Displays the exact results
Domain dependency	Domain dependent	Domain independent
Working procedure	Assigns large value to more important pages	Gives values in a sorted order to the web pages
Input parameters	Back links are used as input parameters in this algorithm	Mined word is used as input parameter in this algorithm
Algorithm used	Page ranking algorithm	Stemming algorithm and k means clustering algorithm
Results Quality	Quality of results produced by this algorithm is medium	Higher Quality of results are produced than page rank algorithm

each user is treated as an individual node and each query with a user identifier. We perform the grouping in a similar dynamic fashion, whereby we first place the current query and clicks into a query group.

Query Reformulation: To ensure that each query group contains closely related and relevant queries and clicks, it is important to have a suitable relevance between the current query groups. We assume that users generally issue very similar queries and clicks within a short period of time. The search history of a large number of users contains signals regarding query relevance, such as which queries tend to be issued closely together. This captures the relationship between queries frequently used leading to clicks on similar URLs. Query reformulation graph and the query click graph are developed from search logs and how to use them to determine relevance between queries or query groups within a user's history is given.

History Grouping: Query groups will first treat every query in a user's history as a query group and then merge these query groups in an iterative fashion (in a k-means). However, this is impractical in our scenario for two reasons. It may have the undesirable effect of changing a user's existing query groups, potentially undoing the user's own manual efforts in organizing her history. The next reason is it involves a high-computational cost, since we would have to repeat a large number of query group similarity computations for every new query.

Comparison Study of Proposed System with the Existing System: Our model is developed by a technique called web usage mining which improves the human interaction with the system by identifying the browsing patterns which is done by analyzing the navigational behavior of user whereas the traditional search engine uses web structure mining which focuses more on discovering the model of link structure of the web pages and to generate structured summary about the website and web page. Once the query from the user is submitted, the efficient search engine has to retrieve the exact information that the user wants which is called as accuracy of results. In the existing system accuracy of results is still a challenge.

The proposed model displays the exact information by extracting the usage pattern and analyzing it. The systems that are already existing are domain dependent whereas the proposed model is domain independent because it uses both explicit and implicit feedback-collection methods to obtain information on user's interests and it is implemented as a web service. If we observe the working procedure of a search engine, it assigns large value to more important pages instead of dividing the rank value of a page evenly among its outlink pages. Rather our model works by giving values in a sorted order to the web pages returned by a search engine as a numerical value in response to a user query. The traditional method works by page rank algorithm which is purely based on number of inlinks and outlinks. But our method uses stemming algorithm which focuses mainly on improving the usability of web systems.

CONCLUSION

We have developed a new system as a solution and improvement of navigation-related Web usability problems by retrieving extracted usage patterns and potential user interest. As demonstrated, our system can help improve the usability of Web systems. Once a query is searched then the next time the query is reposted the task of retrieving the specific URL becomes simpler. Our system can contribute significantly to the energy saving which means the total data used will be reduced. The time taken for retrieving large number of results in traditional browsers is huge. But our recommender system which gives out the exact web pages of users' interest it significantly reduces time complexity. The major advantages of our system can be quantified by the progressively better effectiveness (higher task completion rate) and efficiency (less time for given tasks). Our method is not intended to and cannot replace heuristic

usability evaluation .With automated tool support for a significant part of the activities involved, our method is cost-effective. It would be particularly valuable in the two common situations, where an adequate number of actual users cannot be involved in testing and cognitive experts are in short supply. Client logs in our method represent real users' operations in natural working conditions.

For the future work, a key information extraction algorithm will be developed to compare with the term extraction method in this work and we will perform intense comparisons with the existing semantic Web-page recommendation systems.

REFERENCES

1. Agarwal, A. and M. Prabaker, 2009. "Building on the usability study: Two explorations on how to better understand an interface," in Human-Computer Interaction. New Trends, J. Jacko, Ed. New York, NY, USA: Springer, 2009, pp: 385-394.
2. Anderson, J.R., D. Bothell, M.D. Byrne, S. Douglass, C. Lebiere and Y. Qin, 2004. "An integrated theory of the mind," Psychol. Rev., 111: 1036-1060.
3. Arce, T., P.E. Rom´an, J.D. Vel´asquez and V. Parada, 2014. "Identifying web sessions with simulated annealing," Expert Syst. Appl., 41(4): 1593-1600.
4. Arlitt, M.F. and C.L. Williamson, 1997. "Internet Web servers: Workload characterization and performance implications," IEEE/ACMTrans. Netw., 5(5): 631-645.
5. Barnum, C.M. and S. Dragga, 2001. Usability Testing and Research. White Plains, NY, USA: Longman.
6. Liu, B., B. Mobasher and O. Nasraoui, 2011. "Web Usage Mining," in Web Data Mining: Exploring Hyperlinks, Contents and Usage *Data*, B. Liu, Ed.: Springer-Verlag Berlin Heidelberg, pp: 527-603.
7. Mobasher, B., 2007. "Data Mining for Web Personalization," in The Adaptive Web. vol. 4321, P. Brusilovsky, A. Kobsa and W. Nejdl, Eds.: Springer-Verlag Berlin, Heidelberg, pp: 90-135.
8. Stumme, G., A. Hotho and B. Berendt, 2004. "Usage Mining for and on the Semantic Web, " AAAI/MIT Press, pp: 461-480.
9. Dai, H. and B. Mobasher, 2005. "Integrating Semantic Knowledge with Web Usage Mining for Personalization," in Web Mining: Applications and Techniques, A. Scime, Ed. Hershey, PA, USA:IGI Global, pp: 276-306.
10. Rios, S.A. and J.D. Velasquez, 2008. "Semantic Web Usage Mining by a Concept-Based Approach for Off-line Web Site Enhancements," in Web Intelligence and Intelligent Agent Technology, 2008. WI-IAT '08. IEEE/WIC/ACM International Conference on, pp: 234-241.
11. Salin, S. and P. Senkul, 2009. "Using Semantic Information for Web Usage Mining based Recommendation," in 24th International Symposium on Computer and Information Sciences, pp: 236-241.
12. Bose, A., K. Beemanapalli, J. Srivastava and S. Sahar, 2007. "Incorporating Concept Hierarchies into Usage Mining Based Recommendations," in Proceedings of the 8th Knowledge discovery on the web international conference on Advances in web mining and web usage analysis Philadelphia, PA, USA: Springer-Verlag, pp: 110-126.
13. Mabroukeh, N.R. and C.I. Ezeife, 2009. "Semantic-Rich Markov Models for Web Prefetching, " in 2009 IEEE International Conference on Data Mining Workshops Miami, Florida, USA, pp: 465-470.