INTELLIGENT WEB-HISTORY BASED ON A HYBRID CLUSTERING ALGORITHM FOR FUTURE-INTERNET SYSTEM

J.PRABAVATHY. and S.K.SUDHASARANYA

1 Department of Computer Science & Engineering, Pallavan College of Engineering, Kanchipuram, Tamil Nadu, India saranyacse.1987@gmail.com

2 Department of Computer Science & Engineering, Pallavan college of Engineering, Kanchipuram, Tamil Nadu, India

Abstract-
Recommendation systems can take advantage of semantic reasoning-capabilities to overcome common limitations of current systems and improve the recommendations’ quality. In this paper, present a personalized-recommendation system, a system that makes use of representations of items and user-profiles based on ontology in order to provide semantic applications with personalized services. The recommender uses domain ontologies 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; on the other hand, 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. Web Usage Mining plays an important role in recommender systems and web personalization.

In this project, we propose an effective recommender system based on ontology and Web Usage Mining. The first step of the approach is extracting features from web documents and constructing relevant concepts. Then build ontology for the web site use the concepts and significant terms extracted from documents. According to the semantic similarity of web documents to cluster them into different semantic themes, the different themes imply different preferences. The proposed approach integrates semantic knowledge into Web Usage Mining and personalization processes.

INTRODUCTION

WEB-PAGE recommendation has become increasingly popular, and is shown as links to related stories, related books, or most viewed pages at websites. When a user browses a website, a sequence of visited Web-pages during a session (the period from starting, to existing the browser by the user) can be generated. There are a number of issues in developing an effective Web-page recommender system, such as how to effectively learn from available historical data and discover useful knowledge of the domain and Web-page navigation patterns, how to model
and use the discovered knowledge, and how to make effective Web-page recommendations based on the discovered knowledge. A great deal of research has been devoted to resolve these issues over the past decade. It has been reported that the approaches based on tree structures and probabilistic models can efficiently represent Web access sequences (WAS) in the Web usage data. This project presents a novel method to provide better Webpage recommendation based on Web usage and domain knowledge, which is supported by three new knowledge representation models and a set of Web-page recommendation strategies. The first model is an ontology-based model that represents the domain knowledge of a website. The construction of this model is semi-automated so that the development efforts from developers can be reduced. The second model is a semantic network that represents domain knowledge, whose construction can be fully automated. This model can be easily incorporated into a Web-page recommendation process because of this fully automated feature. The third model is a conceptual prediction model, which is a navigatin network of domain terms based on the frequently viewed Web-pages and represents the integrated Web usage and domain knowledge for supporting Web-page prediction.

II PRELIMINARIES

Useful knowledge discovery from Web usage data and satisfactory knowledge representation for effective Web-page recommendations are crucial and challenging. Existing system provide method to efficiently provide better Web-page recommendation through semantic enhancement by integrating the domain and Web usage knowledge of a website. Two new models are proposed to represent the domain knowledge. The first model uses ontology to represent the domain knowledge. The second model uses one automatically generated semantic network to represent domain terms, Web-pages and the relations between them.

Another new model, the conceptual prediction model, is proposed to automatically generate a semantic network of the semantic Web usage knowledge, which is the integration of domain knowledge and Web usage knowledge. A number of queries have been developed to query about these knowledge bases. Based on these queries, a set of recommendation strategies have been proposed to generate Web-page candidates. The recommendation results have been compared with the results obtained from an advanced existing Web Usage Mining (WUM) method.

I11 PROPOSED APPROACH

In proposed system present a personalized-recommendation system, a system that makes use of representations of items and user-profiles based on ontology 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 proposed is domain-independent, is implemented as a Web service, and uses both explicit and implicit feedback-collection methods to obtain information on user’s interests. Proposed recommender system based on ontology and Web Usage Mining. The first step of the approach is extracting features from web documents and constructing relevant concepts. Then build ontology for the web site use the concepts and significant terms extracted from documents. According to the semantic similarity of web documents to cluster them into different semantic themes, the different themes imply different preferences.

**IV IMPLEMENTATION**

A. **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 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.
B. 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 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.

C. 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 leading to clicks on similar URLs. Query reformulation graph and the query click graph from search logs, and how to use them to determine relevance between queries or query groups within a user’s history.

C. HISTORY GROUPING

Second, it involves a high-computational cost, since we would have to repeat a large number of query group similarity computations for every new query. query groups is to 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. First, 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.

V RELATED WORK

Web users may not always be successful in using a representative vocabulary when locating objects in a system. Therefore, query expansion attempts to expand the terms of the history in Information Retrieval (IR) and web search. Among the various QE approaches presented in literature, some take advantage of implicit relevance feedback, some use external sources, and some implement semantic QE. These techniques are generally non-releuser focused. There are also user-focused QE methods. For example, methods implicitly select terms from the query logs and/or their associated clicked documents, and methods requiring the user to explicitly provide relevance feedback or perform interactive query expansion.

VI CONCLUSION

In this project, a novel approach has been proposed to infer user search goals for a query by clustering its feedback sessions represented by pseudo-documents.

First, we introduce feedback sessions to be analyzed to infer user search goals rather than search results or clicked URLs. Both the clicked URLs and the unclicked ones before the last...
click are considered as user implicit feedbacks and taken into account to construct feedback sessions. Therefore, feedback sessions can reflect user information needs more efficiently. Second, we map feedback sessions to pseudo document approximate goal texts in user minds. The pseudo-documents can enrich the URLs with addition a textual contents including the titles and snippets. Based on these pseudo-documents, user search goals can then be discovered and depicted with some keywords. Finally, anew criterion CAP is formulated to evaluate the performance of user search goal inference. Experimental resultson user click-through logs from a commercial search engine demonstrate the effectiveness of our proposed methods.

VII FUTURE ENHANCEMENT

Integrating domain knowledge with Web usage knowledge enhances the performance of recommender systems using ontology-based Web mining techniques. The construction of this model is semi-automated so that the development efforts from developers can be reduced. The user-profile learning algorithm, responsible for expanding and maintaining up-to-date the long-term user’s interests, employs a domain-based inference method in combination with other relevance feedback methods to populate more quickly the user profile and therefore reduce the typical cold-start problem. The filtering algorithm, which follows a stemming approach, makes use of a semantic similarity method based on the hierarchical structure of the ontology to refine the item-user matching score calculation.

VIII REFERENCES