ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

A Survey on Domain Relevant Web-Page Recommendation Using Ontology and Web Usage Knowledge

Shubhada S. Mali[1], Prof. S. S. Nandgaonkar[2]

[1]Student of ME-II, Dept. of Computer Engineering, VPCOE, Baramati, Savitribai Phule Pune University, Baramati, Maharashtra, India

[2]Assistant Professor, Dept. of Computer Engineering, VPCOE, Baramati, Savitribai Phule Pune University, Baramati, Maharashtra, India

ABSTRACT: The World Wide Web provides different kind of web recommendations which are made available to users every day that includes image, audio, video, query suggestion and web page. Web-page recommendations play an important role in intelligent web system. Web Usage mining is the process of retrieving useful knowledge such as sequential patterns from web usage log. The predicted navigational patterns from the user's web usage log are used to provide personalized web-page recommendations. This paper uses better web-page recommendation by integrating web usage and domain knowledge via the three new knowledge representation models and a set of web-page recommendation strategies. First model is an ontology-based model which is used to represent the domain knowledge of a website or captures the domain of interest of particular user. The second model is the semantic network, which is a kind of knowledge map which represents domain terms, web-pages and relationship between them. The third model is the Conceptual Prediction Model (CPM), which is the combination of domain knowledge and web-usage knowledge. The recommendation strategies make use of the domain knowledge and prediction model to predict and recommend the next web-pages.

KEYWORDS: Domain ontology, semantic knowledge representation, semantic network, web usage mining, web page recommendation.

I. INTRODUCTION

The rapid growth of information on World Wide Web (WWW) has become huge repository of information and it keeps growing exponentially as new information or web-page added. The large amount of information is available on the World Wide Web (WWW). This information lacks an integrated structure or schema. So it becomes more difficult for users to access relevant information effectively. Hence it became more challenging to provide highly relevant information to users with diverse needs. Web Usage Mining (WUM) is the process of retrieving useful knowledge such as sequential patterns from web usage log [3]. The core of that web-page recommendation is the prediction of user's navigational behavior and evaluating which web-pages that users would like to view in the future.

Web-page recommendation has become increasingly popular, which is shown as provides suggestions relevant to e-commerce product, news, advertisement, education, government or most viewed pages at websites. When a user browses a website, a sequence of visited web-pages during session (i.e. Web Access Sequences) can be generated. The objective of the web-page recommender system is to predict web-pages that will be visited from a given web-page of a website and to improve web-site usability. Web-page recommendations based on the Web Access Sequences (WAS) from web usage data [1]. WAS is nothing but the sequences of visited web-pages by user's during a session. By using this approach, given the current visited web-page and k-previously visited pages, the web-pages that will be visited in the next navigation step can be predicted.

A number of problems or challenges have been encountered in the development of web-page recommender systems. If a user is visiting a web-page that has not been accessed before or is not in the discovered Web Access Sequence, e.g. a newly added web-page, then user cannot get recommendation. This is often referred to as the "new-page problem" or "cold-start problem". This problem occurs due to:

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

1. The recommendations are generated based on the recommendation rules or strategies obtained from the frequent access patterns discovered from the web-usage dataset.
2. The new page is not included in the web usage dataset so it cannot appear in any discovered patterns.
3. The system do not have recommendation rule corresponding to the new page.

The semantic-enhanced web-page recommendation is effective to overcome the "new-page" problem [2].

The growth of the web has created a big challenge for directing the user to the web pages in their areas of interest. Web usage mining plays an important role in finding these areas of interest based on user's previous actions. The extracted patterns in web usage mining are useful in various applications such as recommendation. Classical web usage mining does not take semantic knowledge and content into pattern generations. Recently, ontology shows as background knowledge, can improve pattern's quality. The aim is to design a hybrid recommendation system based on integrating semantic information with web usage mining and web page clustering based on semantic similarity. The integration of semantic patterns and web page clustering creates a list of appropriate recommendations, which yields more accurate recommendations.

II. WEB USAGE MINING AND RECOMMENDATION TECHNIQUES

1. Generating dynamic higher-order markov model in web usage mining [5]:

J. Borges and M. Levene in 2005 proposed to model a collection of user web navigation sessions as a Hypertext Probabilistic Grammar (HPG). Markov models have been widely used for modeling user's web navigational behavior. Markov models are used to represents a collection of user web navigation sessions. A HPG corresponds to a first order Markov model, which makes use of the N-gram concept to achieve increased accuracy by increasing the order of the Markov chain. Most web mining systems use techniques such as clustering, association rule mining and sequential pattern mining to search for patterns in navigation records, but do not take into account the order the in which pages were accessed. This limitation has been tackled by building a sequence of higher-order Markov models. Thus, this model is able to capture a variable length history of pages, where different history lengths are needed to accurately model user navigation. This method can mine navigation patterns, taking into account the order of page views.

2. Mining Web log sequential patterns with position coded PLWAP-tree [6]:

C.I. Ezeife and Y. Lu [6] in 2005 proposed a more efficient approach for using the WAP-tree to mine frequent sequences, which totally eliminates the need to engage in numerous re-constructions of intermediate WAP-trees during mining. This method builds the frequent header node links of the original WAP-tree in a pre-order fashion and uses the position code of each node to identify the ancestor/descendant relationships between nodes of the tree. It then finds each frequent sequential pattern, through progressive prefix sequence search, starting with its first prefix subsequence event. Sequential pattern technique is useful for finding frequent web access patterns. But the problem of mining sequential patterns from web logs is based on Web Access Sequence Database (WASD). WASD is used to indicate the number of access sequences in the database. This uses PLWAP algorithm for efficiently mining sequential patterns from web log. In order to avoid recursively re-constructing intermediate WAP-trees, pre-order frequent header node linkages and position codes are proposed. The PLWAP algorithm adapts the WAP-tree structure for storing frequent sequential patterns to be mined.

3. Semantic Web personalization [11]:

M. Eirinaki, D. Mavroeidis, G. Tsatsaronis and M. Vazirgiannis in 2006 proposed a semantic web personalization framework, which enhances the recommendation process with content semantics. This method focus on word sense disambiguation techniques which can be used in order to semantically annotate the web site's content with ontology terms. The framework exploits the inherent semantic similarities between the ontology terms in order to group web documents together and semantically expand the recommendation set. Web personalization is the process of customizing a web site to the needs of each specific user or set of users. Personalization of a web site may be performed by the provision of recommendations to the users, highlighting/adding links, creation of index pages, etc. The web personalization systems are mainly based on the exploitation of the navigational patterns of the Website's visitors. When a personalization system relies solely on usage-based results, however, valuable information conceptually related to what is finally recommended may be missed. The exploitation of the web pages' semantics can considerably improve the results of web usage mining and personalization.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

4. Sequential access pattern mining for web recommendations [7]:

B. Zhou, S. C. Hui, and A. C. M. Fong in 2006 proposed an efficient sequential access pattern mining algorithm, known as CSB-mine (Conditional Sequence Base mining algorithm). Sequential access pattern mining discovers interesting and frequent user access patterns from web logs. The CSB-mine algorithm is based on directly the conditional sequence bases of each frequent event which eliminates the need for constructing WAP-trees. This can improve the efficiency of the mining process significantly compared with WAP-trees based mining algorithms, especially when the support threshold becomes smaller and the size of the database gets larger. This proposed a sequential access-based web recommender system that has incorporated the CSB-mine algorithm for web recommendation. The CSB-mine algorithm has been applied to a web-recommender system that provides personalized web services for accessing related web-pages more efficiently and effectively. In the recommender system, the CSB-mine algorithm is used to mine sequential access patterns from a given web access sequence database. The mined patterns are stored in the pattern-tree, which is then used for matching and generating web-links for online recommendations. The goal of recommender system is to determine which web-pages are more likely to be accessed next by the current user. These approaches suffer from a major drawback in which most users surf websites anonymously via a proxy, and their identities are hidden and difficult to get.

5. Semantic Web mining [13]:

G. Stumme, A. Hotho, and B. Berendt in 2006 proposed the combination of the two fast developing research areas Semantic Web and Web Mining. They make use of Web Mining techniques for building the Semantic Web. Most data on the Web is so unstructured that they can only be understood by humans, but the amount of data is so huge that they can only be processed efficiently by machines. The Semantic Web addresses the first part of this challenge by trying to make the data (also) machine-understandable, while Web Mining addresses the second part by (semi-) automatically extracting the useful knowledge hidden in these data, and making it available as an aggregation of manageable proportions.

6. Concept hierarchies into usage mining based recommendations [12]:

A. Bose, K. Beemanapalli, J. srivastava and s. Sahar in 2006 proposed a novel technique to incorporate the conceptual characteristics of a website into a usage-based recommendation model. We use a framework based on biological sequence alignment. Similarity scores play a crucial role in such a construction and introduce a scoring system that is generated from the website's concept hierarchy. These scores fit seamlessly with other quantities used in similarity calculation like browsing order and time spent on a page. Additionally they demonstrate a simple, extensible system for assimilating more domain knowledge. They have focused on using the extracted patterns to predict the next user request during an online session with the website. Such systems are called Recommender Systems and are useful tools to predict user requests. This predictive ability has application in areas like pre-fetching of pages, increase in overall usability of the website. Recommendation models based only on usage information are inherently incomplete because they neglect domain knowledge. Better predictions can be made by modeling and incorporating context dependent information: concept hierarchy, link structure and semantic classification allow us to do so. This study described a method to combine usage information and domain knowledge based on ideas from bioinformatics and information retrieval. Investigate similarity calculations that use information content values weighted by context that could provide better estimates for similarity. There is a substantial scope for improvement in the ranking of recommendations: domain knowledge can again be used along with the local sequence alignment.

7. Data mining for web personalization [2]:

B. Mobasher in 2007proposed the web personalization process viewed as an application of data mining which must therefore be supported during the various phases of a typical data mining cycle. The stages of data mining cycle including preprocessing and integration of data from multiple sources and pattern discovery techniques that are applied to this data. This presented a number of recommendation algorithms for combining the discovered knowledge with the current status of a user's activity in a website to provide personalized content to a user. In particular profiles commonly used today which lack in their ability to model user context and dynamics. Users access different items for different reasons and under different contexts. User interests and needs change with time. Identifying these changes and adapting to them is a key goal of personalization. The solution to this challenge is likely to lead to the creation of the next-generation of more effective and useful web-personalization and recommender systems that can be employed in

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

increasingly more complex web-based environments. This study focused on web personalization where the recommended objects come from a repository of web objects (items or pages) browseable through navigation of links between the objects.

8. Semantic-rich Markov models for Web prefetching [4]:

N. R. Mabroukeh and C. I. Ezeife in 2009 proposed a lower order markov model with intelligent accurate predictions and less complexity than higher order models, also solving the problem of contradicting prediction. Predicting user's next page request on the World Wide Web is a problem that affects web server's cache performance and latency. Different methods look at the user's sequence of page views, and predict what next page the user is likely to view so it can be prefetched. Domain ontology provides useful source semantic information that can be used in next page prediction systems. The integration of semantic information directly in the transition probability matrix of lower order Markov models, which is provided as a solution to this trade-off problem. This integration is resulting into the semantic-rich lower order Markov models. This integration also solves the problem of contradicting prediction. The contradicting prediction problem occurs whenever there is more than one result with equal probabilities. But in this case, a trained markov model can be used to predict the single next state, given a set of k previous states.

9. Efficient Web usage mining process for sequential patterns technique [9]:

S. T. T. Nguyen in 2009 proposed a new web usage mining process for finding sequential patterns in web usage data which can be used for predicting the possible next move in browsing sessions for web personalization. This process consists of three main stages: preprocessing web access sequences from web server log, mining pre-processed web log access sequences by tree-based algorithm, and predicting web access sequences by using a dynamic clustering-based model. It is designed based on the integration of the dynamic clustering-based markov model with the Pre-Order Linked WAP-tree mining (PLWAP) algorithm to enhance mining performance. Web logs contain not only simple web usage sessions, but also useful information which can be used to trace web usage patterns in relation to browsing behavior and recommending more relevant web-pages to users. This process is called Web Usage Mining (WUM), which aims to discover potential knowledge hidden in the web browsing behavior of users. There are two major methods of sequential pattern discovery: deterministic techniques (recording the navigational behavior of the user) and stochastic methods (the sequence of web pages that have been visited in order to predict subsequent visits). This method focuses on the advanced model-based techniques of the stochastic methods for predicting hidden sequential patterns. Predicting web access patterns using the Markov model is very interesting in web personalization because of its special features, such as modeling a collection of navigation records, modeling user web navigation behavior. In order to resolve this complexity problem, this proposes a new mining algorithm for WUM based on the Markov model from the stochastic approach. The new feature of this algorithm is that it predicts user's web navigation patterns using the frequent web access sequences extracted from the web logs rather than all web-pages. As a result, the complexity of the Markov model can significantly decrease because only frequently accessed web-pages are used, resulting into small number of states.

10. Fast incremental mining of Web sequential patterns [8]:

C. Ezeife, Y. Liu in 2009 proposed two incremental web sequential mining algorithms, RePL4UP (Revised PLWAP For UPdate), and PL4UP (PLWAP For UPdate) which use the PLWAP tree structure to incrementally update web sequential patterns efficiently without scanning the whole database even when previous small items become frequent. The RePL4UP stores position codes of small items in the database sequences in its metadata during tree construction. During mining, RePl4UP scans only the new additional database sequences, revises the old PLWAP tree to restore information on previous small items that have become frequent, while it deletes previous frequent items that have become small using the small item position codes. PL4UP initially builds a bigger PLWAP tree that includes all sequences in the database using a tolerance support, *t*, which is lower than the regular minimum support, S. The position code features of the PLWAP tree are used to efficiently mine these trees to extract current frequent patterns when the database is updated. These approaches more quickly update old frequent patterns without the need to re-scan the entire updated database. The RePL4UP and PL4UP algorithms are applied to the incremental web sequential mining problem. The goal of the incremental mining of web sequential patterns is to generate current frequent patterns for the updated database (consisting of both old and incremental data) using mostly only the incremental (or newly added) data and previously mined frequent patterns.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

Therefore when compared with existing systems, Thi Thanh Sang Nguyen, Hai Yan Lu, and Jie Lu [14] in 2014 proposed the better web-page recommendation system. This study proposed a framework to efficiently provide better web-page recommendation through semantic enhancement by integrating the domain and web usage knowledge. This introduced the Conceptual Prediction Model (CPM) which provides better web-page recommendation through prediction of concepts from Frequently Viewed Term Patterns (FVTP) and also predicting the next web-pages through constructing Term Navigational Network (TermNavNet) of web usage knowledge or web-pages. This study also addresses the problem in developing web-page recommendation system, such as "new-page" problem and manually construction of domain knowledge using ontology. This web-page recommender system provides better performance than existing Web Usage Mining (WUM) Method.

III. SUMMARY

Sr. No.	Authors	Advantages	Methods	Disadvantages
1	J. Borges[5]	Widely used for modeling user's web navigation behavior. It can mine navigation patterns, taking into account the order of page views.	PLWAP-mine, Markov model	For some web sites users navigate with only a short history of the pages previously visited.
2	C.I. Ezeife [6]	WAP-tree to mine frequent sequences, which totally eliminates the need to engage in numerous reconstructions of intermediate WAP-trees during mining.	PLWAP-mine	The procedure for transforming the web log to database is still time-consuming. For mining sequential patterns in transaction databases, there is need to handle concurrency of events.
3	M. Eirinaki[11]	A semantic web personalization is used to enhance the recommendation process with content semantics.	Ontology	Calculating semantic similarity among web-pages is very time consuming and less reusable.
4	B. Zhou [7]	CSB-mine, for mining sequential access patterns from web access sequence databases. There is no need to construct WAP tree	CSB-mine	Suffer from a major drawback in which most users surf websites anonymously via a proxy, and their identities are hidden and difficult to get.
5	G. Stumme [13]	Semantic web mining is used to retrieve the relevant information or web-pages on web search engines.	Ontology	There is need to calculate the semantic similarity between web-pages accurately, which is very time consuming and less reusable.
7	B. Mobasher[2]	Web personalization is achieved using clustering, markov model.	Sequential pattern mining	When data will be made anonymous before clustering, so that there are no personal profiles. Privacy issue is raised in web personalization.
8	N. R. Mabroukeh [4]	The integration of semantic information into probabilistic low-order Markov models to overcome contradicting prediction problem.	Domain Ontology	As the order of the Markov model increases, so does the number of states and the model complexity. Reducing the number of states leads to inaccurate transition probability

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

				matrix thus less predictive power, and less accuracy.
9	S. T. T. Nguyen [9]	A web usage mining process for finding sequential patterns in web usage data which can be used for predicting the possible next move in browsing sessions for web personalization.	Dynamic Markov model	The complexity of a Markov model excessively increases when using it to model a huge number of web pages.
10	C. Ezeife [8]	Incremental web sequential mining	RePL4UP, PL4UP	When there are too many small items in database, RePL4UP, PL4UP algorithms does not perform well.

IV. CONCLUSION

This study aims to address the problems in developing web-page recommender systems, such as "new page" problem and manually construction of domain knowledge. This study used the two new models for representation of domain knowledge of a website. The first model is ontology based which represents domain knowledge of the website. The second model is semantic network to represent domain terms, web-pages and relation among them. This study introduced the conceptual prediction model which integrates the domain knowledge and web usage knowledge, to support effective and better web-page recommendation. This study provides better web-page recommendation by combining domain and web-usage knowledge. This can offer better web-page recommendation through semantic enhancement.

REFERENCES

1. B. Liu, B. Mobasher, and O. Nasraoui, "Web usage mining", in Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data, B. Liu, Ed. Berlin, Germany: Springer-Verlag, 2011, pp. 527–603.
2. B. Mobasher, "Data mining for web personalization", in The Adaptive Web, vol. 4321, P. Brusilovsky, A. Kobsa, and W. Nejdl,Eds. Berlin, Germany: Springer-Verlag, 2007, pp. 90–135.
3. S. Salin and P. Senkul, "Using semantic information for web usage mining based recommendation", in Proc. 24th ISCIS, Guzelyurt, Turkey, 2009, pp. 236–241.
4. N. R. Mabroukeh and C. I. Ezeife, "Semantic-rich Markov models for Web prefetching", in Proc. ICDMW, Miami, FL, USA, 2009, pp. 465–470.
5. J. Borges and M. Levene, "Generating dynamic higher-order Markov models in Web usage mining", in Proc. PKDD, Porto, Portugal, 2005, pp. 34–45.
6. C. I. Ezeife and Y. Lu, "Mining Web log sequential patterns with position coded pre-order linked WAP-tree" , Data Min. Knowl.Disc., vol. 10, no. 1, pp. 5–38, 2005.
7. B. Zhou, S. C. Hui, and A. C. M. Fong, "Efficient sequential access pattern mining for web recommendations", Int. J. Knowl.-BasedIntell. Eng. Syst., vol. 10, no. 2, pp. 155–168, Mar. 2006.
8. C. Ezeife and Y. Liu, "Fast incremental mining of Web sequential patterns with PLWAP tree", Data Min. Knowl. Disc., vol. 19, no. 3, pp. 376–416, 2009.
9. S. T. T. Nguyen, "Efficient Web usage mining process for sequential patterns", in Proc. IIWAS, Kuala Lumpur, Malaysia, 2009, pp. 465–469.
10. S. Boyce and C. Pahl, "Developing domain ontologies for course content", Educ. Technol. Soc., vol. 10, no. 3, pp. 275–288, 2007.
11. M. Eirinaki, D. Mavroeidis, G. Tsatsaronis, and M. Vazirgiannis, "Introducing semantics in Web personalization: The role of ontologies", in Proc. EWMF, Porto, Portugal, 2006, pp. 147–162.
12. A. Bose, K. Beemanapalli, J. Srivastava, and S. Sahar, "Incorporating concept hierarchies into usage mining based recommendations", in Proc. 8th WebKDD, Philadelphia, PA, USA, 2006, pp. 110–126.
13. G. Stumme, A. Hotho, and B. Berendt, "Semantic Web mining: State of the art and future directions", J. Web Semant., vol. 4, no. 2, pp. 124–143, Jun. 2006.
14. Thi Thanh Sang Nguyen, Hai Yan Lu, and Jie Lu, "Web-Page Recommendation Based on Web Usage and Domain Knowledge", IEEE Trans. On knowledge and data engineering, vol.26, no. 10, October 2014.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 11, November 2015

BIOGRAPHY

Ms. Shubhada S. Mali received the Bachelor degree (B.E.) in Computer Science and Engineering in 2014 from SVPM's college of Engineering, Pune University. Currently, she is pursuing Master's degree in Computer Engineering at Vidya Pratishthan's College of Engineering, Baramati, Savitribai Phule Pune University. Her current research interests include Data Mining and Web Mining.

Prof. S. S. Nandgaonkar obtained Engineering degree (B.E.) in Computer Science and Engineering from WIT, Solapur (Shivaji University) and Postgraduate degree (M. Tech.) in Computer Engineering from College of Engineering, Pune (COEP) both with first class and distinction. She is approved graduate and postgraduate teacher of Pune University and having about 14 yrs. of experience. She is life member of the ISTE. Her research interest is in area of Information Retrieval, data mining, web mining and machine learning.