

A Paper on Multisite Framework for Web page Recommendation Using Incremental Mining

Mr. Rahul B. Patil¹, Prof. K. B. Manwade²

PG Student, Ashokrao Mane Group of Institution, Vathar, Kolhapur, Maharashtra, India
Associate Professor, Ashokrao Mane Group of Institution, Vathar, Kolhapur, Maharashtra, India

ABSTRACT

Large amounts of information present on web due to that web user finds difficulty when surfing the web. Useful and Effective suggestions provided by Recommendation systems which help web users. User's future interest evaluate by using the learning and prediction models which learn user's behavior. Web page recommendation improves web user satisfaction and website usage. Semantic enhanced "Multisite framework for Web page Recommendation using Incremental Mining" improves web page recommendation. In which web page recommendation will be done by using multiple website in same domain. Will be give effective recommendation results given by Different sites. To identify and collect more appropriate web usage data Dynamic web page recommendations tool is generated. Related terms identified which help to generate more appropriate data. Incremental mining update web usage knowledge and updates the knowledge base.

Keywords: Incremental Mining, IPNAR, Semantic Network Model, Ontology, Semantic Knowledge, FWAP

I. INTRODUCTION

Web users make poor decisions because of enormous amounts of information. Recommendation systems have proved in recent years to be a valuable means of helping web users by providing useful and effective suggestions. The core techniques in recommendation systems are the learning and prediction models which learn user's behaviour and evaluate what users would like to view in the future.

Web page will be recommended to particular user based on his current web navigation behaviour. It will be improved web user satisfaction and website usage. Knowledge bases used in web page recommendation system includes web usage knowledge domain and website knowledge domain. Web page recommendation system extracts the semantic of web pages from web pages title. It generates domain term of web pages from single site only. This can be improved by Semantic enhanced multisite framework for Web page Recommendation using Incremental Mining.

II. METHODS AND MATERIAL

A. Literature Review

1. Thi Thanh Sang Nguyen, Hai Yan Lu, and Jie Lu et al proposed a novel method to efficiently provide better web page recommendation through semantic enhancement by integrating the domain and Web usage knowledge of a website.
2. M.Venu Gopalacharii, Po Sammulal et al presented an effective method to integrate the domain knowledge and web usage knowledge of a website through semantics. A new model is framed to construct a semantic hierarchy of the web log data and the domain contents, which represents the integrated usage knowledge and domain knowledge.
3. Nazneen Tarannum, S.H.Rizvi, Prof. R.R. Keole et al proposed a desktop search utility which uses web usage mining process for finding term patterns in web query data which can be used for predicting the possible next pages in browsing sessions.
4. Mohammad Amir Sharif, Vijay V. Raghavan et al presented a novel hybrid recommendation system based on clustering of items using co-occurrence

- information of pages and content information of pages. These two different types of clusters are used in a parametric form to get aggregated recommendations based on the available preference information of users.
5. Anuradha Veleti, T.Nagalakshmi et al proposed an incremental algorithm (IPNAR) that mines positive and negative association rules in web usage data. The incremental based algorithm incrementally update web log association rules by utilizing the metadata of old database transactions as well as old mined rules, performs single scan over the dataset.

6. Archana Godbole, Mahendra Kumar Rai et al explores the deployment of clicking pattern algorithms in a distributed grid computing environment and demonstrates its effectiveness by empirical cases. The basic task is a weblog analysis for huge, widely distributed, hypertext information repository of World Wide Web.

B. Why Incremental Mining

Websites and the knowledge bases have been evolving with respect to time, Therefore domain and Web usage knowledge bases need to be updated with respect to time. Web usage data source is the Web log file. So for the updating domain and Web usage knowledge bases the system can take a limited segment of the log file. Because to build the Web usage knowledge base the size of the log file can be huge. So we can use incremental mining method for discovered Web usage knowledge and keeping it updated, incremental mining method can dynamically update the knowledge bases.

C. Proposed Work

In proposed system, web page recommendation will be done by using multiple website in same domain. The system will take the log files from the websites as the input. The semantic analysis of domain terms in different sites will be give effective recommendation results. The existing system works with static Web pages. For dynamic web page recommendations advanced tools will be generated to identify and collect more appropriate web usage data than web logs. Dynamic web click stream analysis will be conducted in the data preparation stage, in which the web page will be identified as dynamic contents rather than static pages. Synonyms of the domain

terms will be identified and clustered into common topics. This would help to cluster web page in to relevant topics and there by optimize the semantic network model of web page. Websites have been evolving over time therefore the knowledge bases, including the domain and Web usage knowledge bases need to update accordingly. To keep discovered web usage knowledge up to date, incremental mining will be developed for dynamically update the knowledge base.

The proposed work has following objectives,

1. To build multisite framework for the web page recommendation.
2. To develop tools for identify and collect appropriate web usage data from dynamic web page.
3. To enhance domain knowledge discovery by using synonyms of domain terms.
4. To update web usage knowledge base by using incremental mining method.

D. Proposed Architecture

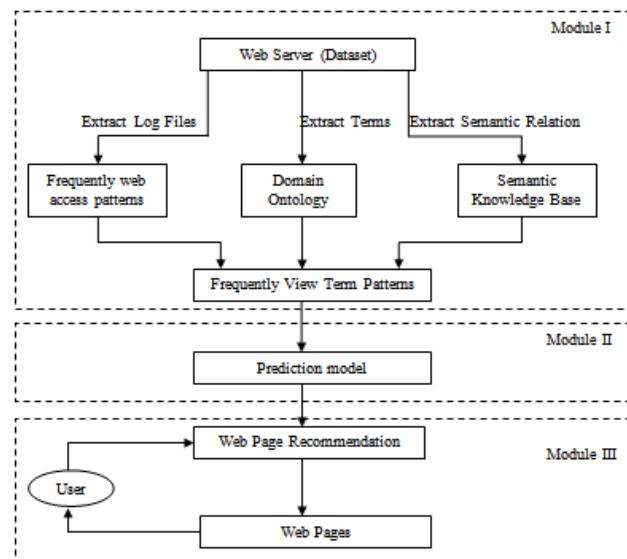


Figure 1: Architecture of Multisite Framework for Web page Recommendation Using Incremental Mining

Following terms will be used for proposed work.

1. Data Collection

Frequent web access pattern will generated from web usage data by using crawler from dataset. It will give the count of term. Domain Ontology will be generated from web usage data. It will give domain knowledge

of extracted term. Symantec knowledge will be generated from web structure data. It will give association between domain term and web pages. In the dataset, users will be identified by id and web pages will be identified by their titles and URLs. The dataset is divided into two subsets, one for training and one for testing. Pre-processing step will be performed on log data to analyze the web log files and produce a list of URLs of web pages that were accessed by users. Crawler will access the current web page browsed by user and to extract the titles from web pages. Similar domain term data will be collected from multisite. Dynamic web click stream will be used to collect data dynamically.

2. Generation of Frequent View Term Pattern

Frequent view term pattern will be generated by extracting frequent web access pattern, domain ontology and semantic knowledge from web usage data and web structure data. Frequent web access pattern will give the count of term which is used by user frequently. It will be extracted from web usage data. Domain ontology represents domain knowledge of extracted term. Taxonomic and non-taxonomic relation of term will find in domain ontology. Semantic Knowledge is kind of knowledge map which represent domain terms, web pages and relation of domain term. Semantic Knowledge also represents association between domain term and web pages. Symantec knowledge will be updated using clustering method where similar domain term will be used for multisite framework.

3. Generation of Prediction Model

Prediction model will develop for generating weighted semantic knowledge of frequently viewed terms pattern. Synonyms of domain term will be used for enhancing domain knowledge discovery. Also to update knowledge bases dynamically incremental mining method will be used. This model will use frequent view term pattern for giving the weight to frequently viewed terms pattern. Weight is probability of the transition between two adjacent terms based on frequently viewed term patterns. Also synonyms of users domain term will be find out and weight is applied to it. Weighted semantic knowledge of frequently viewed terms will use for semantic enhanced web page recommendation.

4. Web Page Recommendation

This model will develop for web page recommendation to web user by using weighted semantic knowledge. User query will take and terms will be compared with weighted semantic knowledge of frequently viewed terms pattern. Term having high probability will be compared with users query. Web page recommendation will use high probability term for web page prediction. Frequently viewed term related with user query will find out and most probable web page will be recommended to web user.

First part of this module will work in order to collect data. Data will be collected from data set of web server. Collection of data consists of web log files and web structure data from multisite. Pre-processing unit will analyze web log file and produce list of URLs of web pages that were accessed by user. Crawler will use for pre-processing unit to extract the titles. Advanced tool will develop to collect more appropriate web usage data like Click stream data.

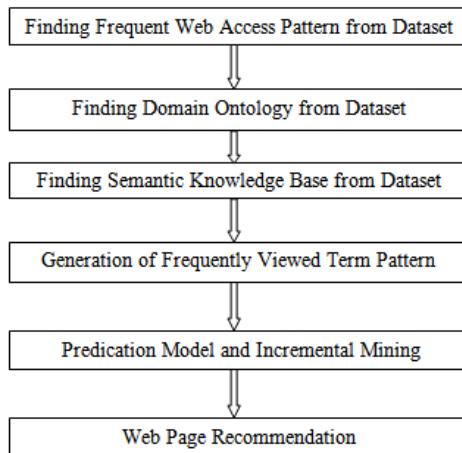
In second part, frequent web access pattern, domain ontology and semantic knowledge will be generated from collected data. Frequent web access pattern will give the count of term which is used by user frequently. Domain ontology represents domain knowledge of extracted term. Semantic Knowledge is kind of knowledge map which represent domain terms, web pages and relation of domain term. Semantic Knowledge also represents association between domain term and web pages. This module will generate frequent view term pattern by using frequent web access pattern, domain ontology and semantic knowledge.

Prediction model will develop for generating weighted semantic knowledge of frequently viewed terms pattern. This model will use frequent view term pattern for giving the weight to frequently viewed terms pattern. Weight is probability of the transition between two adjacent terms based on frequently viewed term patterns. Also synonyms of users domain term will be find out and weight is applied to it. Weighted semantic knowledge of frequently viewed terms will use for semantic enhanced web page recommendation.

In web page recommendation, users query will take as input and compare this query to frequently viewed

term. The term having most probability will be used for recommend of web page. Finally predicted web page will recommend given to user.

E. Implementation Steps



The implementation steps are as follows:

Step I: In step I Frequent web access pattern will generated from web usage data by using crawler from dataset. It will give the count of term which is used by user frequently. Further frequent web access pattern will used in Generation of Frequently Viewed Term Pattern.

Step II: In step II Domain Ontology will be generated from web usage data. It will give domain knowledge of extracted term. Further this domain knowledge will used in Generation of Frequently Viewed Term Pattern.

Step III: In step IV Symantec knowledge will generated from web structure data. It will give association between domain term and web pages. Further this semantic will used in Generation of Frequently Viewed Term Pattern.

Step IV: In step V term generated in above step i.e. Symantec knowledge, Domain Ontology and frequent web access pattern will used to generate frequently viewed terms pattern.

Step V: In this step Prediction model will develop for generating weighted semantic knowledge of frequently viewed terms pattern. Also to update knowledge bases dynamically incremental mining method will be used.

Step VI: In this step using weighted semantic knowledge, web page will recommend to the user by web page recommendation module.

F. SCOPE

Web page recommendation system generates domain term of web pages from single site only. The proposed framework makes web page recommendations for multiple websites in the same domain. The system should take the log files from these websites as the input.

With the advancement in Web technology, pages have been evolving in to dynamic structures. The system will works with dynamic web pages and develop advanced tools to identify and collect more appropriate web usage data than web logs such as click stream data. In proposed system dynamic web click stream analysis will be conducted in the data preparation stage.

System will make usage of synonyms of the domain terms; need to be clustered into common topics. This would help to cluster web pages into relevant topics and thereby optimize the semantic network model of web page.

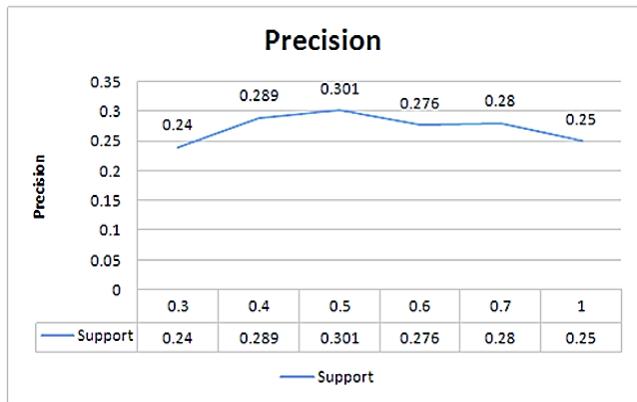
In the traditional web usage data source like web log file, the system can only take a limited segment of the log file to build the web usage knowledge base due to the fact that the size of the log file can be huge. Incremental mining method will keep the discovered web usage knowledge up to date and dynamically update the web usage knowledge bases.

III. RESULT AND DISCUSSION

For Result and analysis we have divided our data set in to training and testing set .we use minsup as threshold value for analysis. It is minimum support value for web access sequence on Frequent Web Access Patten (FWAP).Minsup ranges from 0.3 to 1.Measure value foe evolution decided. They are Precision and Satisfaction.

PRECISION: -Let R_c be the sub-set of R , which consists of all correct recommendation rules. The Web-page recommendation precision is defined as:
 $Precision = |R_c|/|R|$.

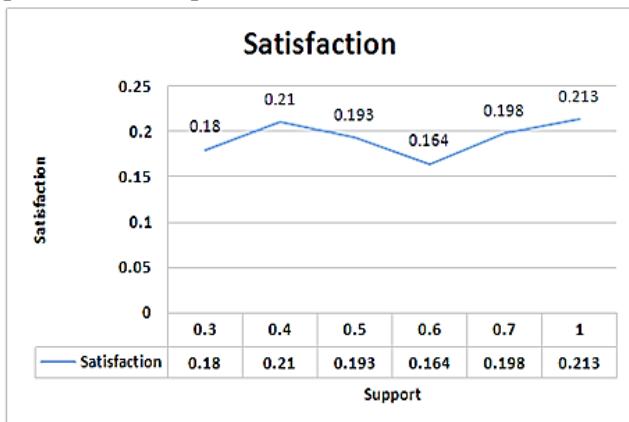
Precision specify ratio of correct recommendation against all recommendation.



SATISFACTION: - Let R_s be the sub-set of R , which consists of all satisfied recommendation rules. The Satisfaction for Web-page recommendation is defined as:

$$\text{Satisfaction} = |R_s|/|R| .$$

Satisfaction specifies ratio of satisfactory recommendation against all recommendation. Result proves that with our approach results are better than previous technique.



IV. CONCLUSION

In proposed work three modules are proposed. Three modules has been completed, namely

1. Generation of frequent view term pattern
2. Generation of prediction model
3. Web page Recommendation

V. REFERENCES

- [1] 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, October 2014.

- [2] M.Venu Gopalacharii, PoSammulal,"Personalized web page Recommender System using integrated Usage and Content Knowledge", IEEE International Conference on Advanced Communication Control and Computing Technologies (ICACCCT), 2014.
- [3] Nazneen Tarannum, S.H.Rizvi, Prof. R.R. Keole "Web-Page Recommendation In Information Retrieval Using Domain Knowledge And Web Usage Mining", International Journal of Science, Engineering and Technology Research (IJSETR).
- [4] Mohammad Amir Sharif, Vijay V. Raghavan, "A Clustering Based Scalable Hybrid Approach for Web Page Recommendation", IEEE International Conference on Big Data, 2014.
- [5] Anuradha Veleti, T.Nagalakshmi, "Web Usage Mining: An Incremental Positive and Negative Association Rule Mining Approach", (IJCSIT) International Journal of Computer Science and Information Technologies, 2011.
- [6] Archana Godbole, Mahendra Kumar Rai,"A Novel Approach for Web Usage Mining with Clicking Pattern in Grid Computing Environment", IJARCSS,February,2014.