Journal of Theoretical and Applied Information Technology
20th October 2013. Vol. 56 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A TECHNIQUE FOR WEB PAGE RECOMMENDATION USING MARKOV MODEL ASSOCIATED WITH QUALITY AND TIME BASED FP MINING

[1]R. SUGUNA, [2] Dr. D. SHARMILA
[1]Assistant Professor, Department of Computer Science and Engineering
Arunai College of Engineering, Thiruvannamalai – 638 401.Tamil Nadu, India.
[2]Prof and Head, Department of Electronics and Instrumentation Engineering
Bannari Amman Institute of Technology, Sathyamangalam-638 401 Tamil Nadu, India.
E-mail: [1]sugunarajasekar@gmail.com

ABSTRACT

The aim of the web page recommendation is to predict the navigation of the users using web usage mining technique. In recent days, the researchers are exploring to develop an algorithm for web page recommendation using pattern mining model. Here, the data are prepared from the web log file and identify the users based on the ip address. Boid algorithm is used to cluster the logs and the quality and time based frequent pattern growth algorithm is used to mine the frequent patterns. Markov model is applied to generate the recommendation. The frequent pattern tree is formed by calculating the total support values of each web page based on the quality and time duration of the web pages and this frequent pattern tree is used to recommend the web page using Markov model. The artificial dataset is generated for experimentation to compare the performance of proposed technique with the existing technique.

Keywords: *Web Page Recommendation, FP-Growth Algorithm, Markov model, clustering*

1. INTRODUCTION

In recent days data is developed enormously. To examine the data for meaningful information is very vital and perceptive task. A key feature for many computation applications [8] is data examination. Data analysis can be separated into two groups exploratory or confirmatory based on the availability of data models. But the important factor which governs the procedure of data analysis is grouping the data based on specific features. These grouping of data in computer lingua franca are called clustering of data on certain parameters. Clustering can be defined as identification of groups and the identification of fascinating distributions in the data [9]. The demand of clustering in large mass of data is the proper arrangement of data in a generic way and it can be analyzed more easily [2]. To make the web content apprehensible by humans and computers is the goal of the semantic web. Thus to allow the software agents for search and find the desired contents, the information and knowledge are shared [10]. Therefore, using organizations [11, 12] there is an increasing effort in annotating the web pages and objects in terms of semantic information.

Many papers were proposed in medical, engineering and biological field [3] based on clustering and its application. Diverse algorithms were used depending on the application. The most common algorithms are K-means algorithm, spectral clustering algorithm, hieratic clustering and fuzzy c-means clustering. The goal of the K-means is to split the P point in N dimension into M clusters and to minimize the sum of square [4]. The hieratical clustering works on the hierarchy of clusters, a binary tree of data that combines same group of points is prepared and good insight in data is provided by the analysis of this binary tree. The fuzzy means allocate a single data into two or more groups based on certain features.

Web mining is one of the successful domains of data mining that extracts meaningful and relevant data from the World Wide Web [13]. The extraction of raw data that is available in the web pages is focused by the web content mining which is a branch of web mining. Typically, the source data contains the textual data like words, tag, etc. in the web pages. The web content mining [14] has two successful applications which are content based categorization and content based ranking of web pages. The structure of the web sites is concentrated by the web structure mining which is a part of the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

web mining. The source data primarily includes the structural information of the web pages like links to the other pages. The web usage mining is also one of the branches of web mining that extracts the relevant data from the server log files.

In this paper, a new technique is proposed for web page recommendation using Markov model which is associated with the quality and time based frequency pattern mining. The steps involved in new technique are data preparation, clustering, mining of constraint association patterns and recommendation. In data preparation, preprocessing the data is carried out and identifying the user based on the ip address. Session is constructed based on user interest level for particular website. The biased frequency database is generated based on the web pages, duration of time and the quality of the web page. The clustering is done by the boid algorithm which is modified from the bird flocking algorithm. The quality and time based frequency pattern growth algorithm is used to mine the frequent patterns. The frequent pattern tree is formed using the total support values of the web pages which are calculated using the quality and duration of the web page. The recommendation is generated using the Markov model based on the frequent pattern tree.

This paper is organized as follows: the second section gives a brief explanation about the related works and the third section shows the technique which proposed to recommend the web page and the fourth section shows the result obtained for this technique and it is compared with the existing techniques and the fifth section concludes the performance of the proposed technique.

2. RELATED WORKS

Different researchers have presented different algorithms for the recommendation of web pages. In this section, we have presented some of the recent works related to the web page recommendation.

Qingyan Yang *et al.* [1] have proposed personalizing web page recommendation through collaborative filtering and topic aware Markov model. Their technique could guide the web users to identify more useful pages without enquire them. It has fascinated more attention in the web mining community. They have proposed a personalized web page recommendation technique which is known as PIGEON through a collaborative filtering and a topic aware Markov technique. They recommended a graph based iteration algorithm to identify the user's interested topics. They also recommended a topic aware Markov model to study the user's navigation patterns that capture both temporal and topical relevance of pages to recommend the topically coherent pagers.

R. Forsati *et al.* [5] have an effective page recommendation algorithm based on distributed learning automata and weighted association rules. They have proposed three algorithms: in their first algorithm, they have used distributed learning automata to study the behavior of the previous users and suggest the pages to the current user derived from the learned patterns. They used the second algorithm for recommendation by introducing a weighted association rule mining algorithm. To deal with the unvisited or newly added pages are the challenging problem in recommendation system. Considering this issue and improving the efficiency of the first two algorithms, they have presented a hybrid algorithm derived from distributed learning automata and suggested weighted association rule mining algorithm. They have also used the HITS algorithm in the hybrid algorithm to extend the recommendation set.

Jia Li *et al.* [6] suggested a technique by combining usage, content and structure data to enhance the web site recommendation. They have investigated a web recommender system that combines usage data, content data and structure data in a web site to create user navigational models. These models are fed back into the system to recommend users page resources. They also proposed an evaluation mechanism to evaluate the quality of the recommender systems.

Xiong Haijun *et al.* [7] suggested a rough set web page recommendation using fuzzy semantic log. Their technique first alters the access logs in the web into fuzzy semantic logs and then matches the current session with the rules founded and eventually provides the recommendation set of web pages to the users. They have used the backward path ratio technique to evaluate the effectiveness of the algorithm.

Mingyu Lu *et al.* [18] have proposed an association based technique for web page recommendation. Their motto is to create recommendatory web page links for the user derived from the data associated with users query. The meaning and variety of association are introduced and the technique to realize web page recommendation derived from the generated associational data related with users query is explained.

Journal of Theoretical and Applied Information Technology
20th October 2013. Vol. 56 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3. QUALITY AND TIME BASED FP MINING FOR WEB RECOMMENDATION USING MARKOV MODEL

This section explains the detailed description of proposed method to recommend the web page. The Fig.1 shows the sample block diagram which explains the process of proposed technique.

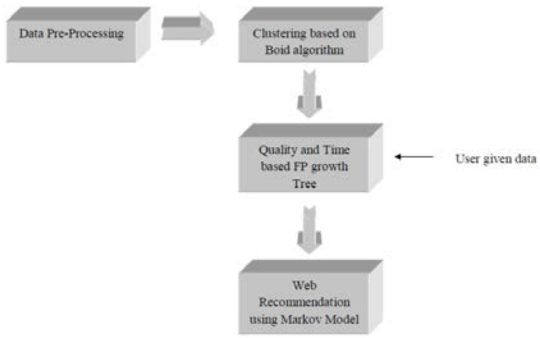

Fig.1 Block Diagram of our Proposed Technique

3.1. Preprocessing

This section details the preprocessing of web log file which is used as input for the new technique. The format of the web log file is as follows: IP address, access time, HTTP request type used, URL of the referring page and the browser name. A sample web log file is as follows: *192.167.28.11 [28/Sep/2012:10:24:30] "GET / HTTP/1.1" "http://www.loganalyzer.net"Safari/3.2.2 Windows 07*. Initially, web logs are preprocessed for applying the mining process.

3.1.1. Identifying User

This is an important step to classify the user. The user is classified based on the IP address and the session which the IP address is used. Therefore, the IP address and the session are used to classify the unique user. The time period is set to classify the unique users from the same IP address. When the time period gets end, the transaction of the current user will get end and from the next second the new user's session will get start. The transaction is formed based on this procedure from the web log file.

3.1.2. Biased Frequency Database Generation

This database is generated based on the web pages which the users visited and the time duration which the users spend to view those web pages and the quality of those web pages which is based on the rate that the users gave. This database would be in two sections; the first section contains the web pages which are visited by the users and the second section contains the time duration and quality of the web pages based on the first section.

$$BFD_i = \{p_1, p_2, \ldots\ldots\ldots, p_n\}$$
$$BFD_j = \{t_1 q_1, t_2 q_2, \ldots\ldots\ldots, t_n q_n\}$$

Where,

p → Web Pages

t → Duration of time spent for the web page p

q → Quality based on user rating of web page p

Table.1 Sample Database Contains The Users With The Web Pages visited by them

Users	Web Pages Visited
U_1	w,n,b,j,o,h
U_2	n,s,b,w,k
U_3	s,w,m,c
U_4	s,b,i,g
U_5	n,w,b,d,k,v

Table.2 Sample Database With The Values Of Time Duration And quality

Users	(time duration, quality)
U_1	(11, 4.2), (15, 3.8), (13, 4.5), (22, 2.5), (18, 2.8), (14, 3.5)
U_2	(18, 4.2), (16, 3.4), (23, 4.5), (20, 5), (24, 4.4)
U_3	(22, 4.8), (16, 3.9), (11, 3.1), (13, 1.5)
U_4	(23, 4.7), (17, 4.0), (11, 2.7), (18, 3.6)
U_5	(13, 2.4), (18, 3.3), (19, 4.0), (17, 3.7), (15, 4), (23, 4.6)

The Table.1 shows the sample database which contains the users and the web pages which were visited by them. The Table.2 shows the sample database that contains the respective values of time duration and quality of the web pages visited by the users on Table.1.

3.1.3. Parameters taken for Biased Frequency Database Generation

This section details about the parameters which are considered to generate the biased frequency database. The parameters taken to generate the biased frequency database is duration of time spend on a particular web page by the user and the quality of the web page which is based on the user given rate.

The duration of time spend on a particular web page by a user is one of the important field to identify the preference of the web page. Because if a web page has more importance, the user would access it for a long time and it shows the interest of that user on that particular web page. So, these

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

parameters are applied to generate the biased frequency database effectively.

Another parameter which is considered to generate the biased frequency database is quality of the web page which is based on the user given rate. The rate which user gives is based on the user's desire. The rating points would be amid zero and five. If the web page got more five points by diverse users, then that web is considered as most desired web page by different users. So this is also an important factor which is considered to generate the biased frequency database.

3.2. Boid Algorithm for Clustering the Logs

The Boid algorithm is used to group the websites. The boids are considered as moving data points and hence the clustering should be considered in diverse manner. The parameters used to cluster the boids are affinity and centroid based calculation. If there are n number of objects, then there would be n number of boids considered in a dataset.

$$D = [x_1, x_2, \ldots, x_n]$$

Where, D denotes the dataset and x denotes the boids in the dataset and n is the number of boids.

3.2.1. Affinity Calculation

The affinity value defines relationship of two boids which has the similarities amid same objects. The similarity amid the related objects is indicated quantitatively by the affinity amid the two boids. The affinity amid the related boids is high, if it has more similar objects. To group the data the boids algorithm uses the affinities of boids as an important parameter. The affinities amid two boids are calculated using the following formula:

$$A_{ij} = \left(\sqrt{\sum_{k=1}^{L} \left(x_{ik} - y_{jk} \right)^2} \right)$$

Where,

x_i, y_j → boids considered for similarity calculation

L → Total length between x_i and y_j

3.2.2. Centroid rule and Merging rule

Every boid travelling in the environment is considered as centroid and each centroid has many objects as per their similarities. For instance, if a database has hundred objects, the algorithm will have hundred boids at first and each cluster contains single object. A probabilistic rule is then applied to merge centroid. Another probabilistic rule is also applied to create new centroid, to split the previously merged groups and to form two discrete groups.

To merge two boids into single centroid, the merging centroid rule is used. Initially, each group has single object in the database. When two boids are within the sight area of each other, the probability Pm_{xy} of two groups x and y are merge as a single group. The probability is proportional to the affinity amid the two boids.

$$Pm_{xy} = \sqrt{\sum_{i=1}^{n} \left(x_i - y_i \right)^2}$$

When two boids merge, one boid get disappear from the environment. But sometimes it is possible that the merged boid may not belong to the same group and it may get divide from that group. To exclude that problem, a new rule is introduced to create a new centroid. When two or more objectes are grouped, there is a probability that one of them leaves the group and return to the environment to look for a new group. The probability of an individual leaving its cluster is proportional to the difference between its affinity with the current group and its greater affinity with the other groups. Initially, the current affinity with its group (cag) is estimated by finding the affinity between the boid and the centroid of the group which it belongs. Then it calculates the affinity of this boid with the centers of all the other groups and the highest one is named as gag (greatest affinity group). The probability P_i of the boid leaving the group is directly proportional to the difference between cag and gag.

$$cag = \sqrt{\sum_{i=1}^{n} \left(b_i - c_i \right)^2}$$

$$gag = \sqrt{\sum_{i=1}^{n} \left(b_i - g_i \right)^2}$$

$$P_i = gag - cag$$

From the above formulas, b is the object that is represented by the boid, c is the centroid (average values) of the current group and g is the centroid of the group with greatest affinity with b. If the value of the object changes, this rule reflects this change and it may remove the object from its current group.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3.2.3. New Separation, Alignment and Cohesion Rules

According to the new rules, the three behavioral rules are redefined. The rules are redefined based on their affinity value. The rules are as follows:

- Separation rule is defined as the strength of the separation between two boids is inversely proportional to the affinity between the boids. If the value of affinity is less, the separation force between them is high.
- Alignment rule is defined as the degree of alignment varies with respect to the affinity between the boids. If the affinity value between the boids is high, the alignment between them is harder.
- Cohesion rule is defined as the strength of cohesion varies with respect to the boids affinity value. If the affinity value is high, the cohesion is stronger.

3.3. Q and T based FP-Growth Algorithm

The Frequent Pattern Growth algorithm is abbreviated as FP-Growth algorithm which is used to mine the frequent patterns. Here, the total support values for each web page are calculated to mine the frequent patterns. The total support value is calculated based on the quality and time of the web pages and the formula to find the total support value is as follows:

$$TSV = \frac{(S_f + S_t + S_q)}{3}$$

Where,

TSV → Total Support Value

S_f →Support value of frequency for the web page p

S_t →Support value of duration of time spent for the web page p

S_q →Support value of quality for the web page p

The support value of the frequency would be either in zero or one based on the number of times that a particular user visited a particular web page. The support value of the duration of time visited for a particular web page on a particular user is calculated by dividing the duration of time spend for a web page by the maximum duration of time spend for that web page among the whole user. The support value of quality is calculated by dividing the quality value of a web page with its maximum quality value among the whole user.

The total support values which are calculated using the quality and time are applied in the FP-Growth algorithm to mine the frequent patterns. Minimum support value is set and compares it with entire total support value of the web pages and to take the web pages which have the total support value more than the minimum support value. After choosing the web pages that has total support value more compared to the minimum support value which is set, the web pages are ordered in descending order for each user based on the total support values. The Table.3 shows the sample web pages which have the total support value as high compared to the minimum support value and arranged in descending order for each user based on the total support value.

Table.3 Web Pages Which Have High Total Support Value Than The Minimum Support Value

Users	Web Pages Visited
U_1	b,o
U_2	s,b,w
U_3	s,w,m
U_4	s,b,i
U_5	b,w,d

After arranged all the web pages in descending order for every user based on the total support values, we have to generate the FP tree. The FP tree is generated based on every user sequentially. First FB tree is formed for the first user. The FB tree based on the first user is shown in Fig.2.

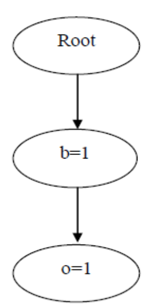

Fig.2 FP Tree For First User

After plotting the FP Tree for the first user,same procedure is applied to the the second user's web pages and plot the FP Tree. The Fig.3 shows the FP Tree after applying the web pages of the second user.

Journal of Theoretical and Applied Information Technology
20th October 2013. Vol. 56 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

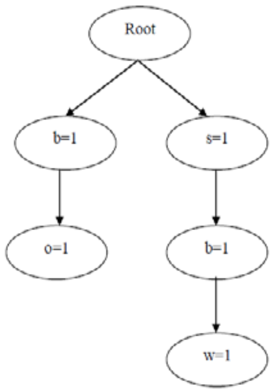

Fig.3 FP Tree After Applying Second User

The Fig.4 shows the FP Tree after applying the web pages of the third user and the Fig.5 shows the FP Tree after applying the fourth user's web pages which has more total support value compared to the minimum support value.

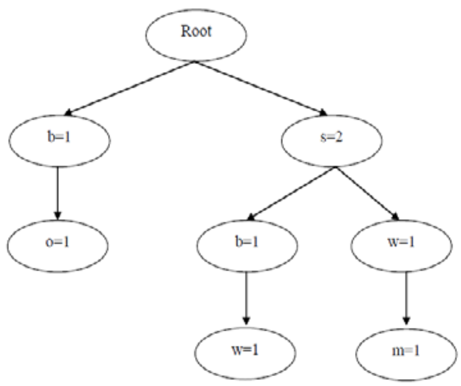

Fig.4 FP Tree After Applying Third User

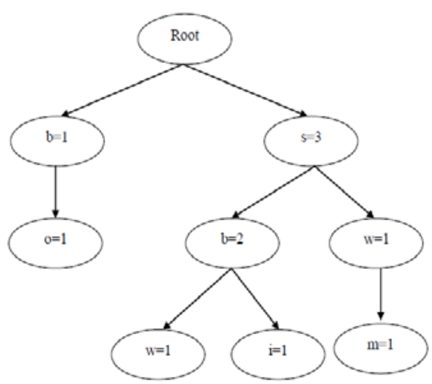

Fig.5 FP Tree After Applying Fourth User

Similarly, the same procedure is applied to all the users and generates the FP Tree. The final FP Tree after applying all the users is shown in the Fig.6.

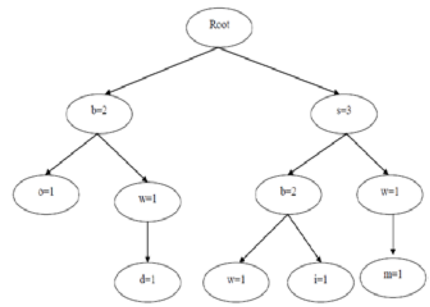

Fig.6 Final FP Tree After Applying All The Users

After plotting the final FP Tree, FP Tree is mined. The mining of FP Tree is as follows: first take the lowest node m. Here, the frequent pattern of m is 1 from the final FP tree. Thereafter, have to construct the conditional pattern base of m from the final FP tree. The conditional pattern base of m is swm which have the pattern values as 1 for each. Take the conditional patterns as sm and wm. The pattern values for both sm and wm is 1. Now, consider the node i and find the frequency pattern for it. The frequency pattern of i is 1 and the conditional pattern base of i is sbi. The pattern values of si and bi is also 1. Similarly, need to find the conditional frequent patterns for every node. The Table.4 shows the conditional frequent patterns for every node in the Fig.6 and the frequent values.

Table.4 Nodes With Respective Conditional Frequent Patterns And Values

Node	**Conditional Frequent Patterns with values**
m	$(m=1),(sm=1),(wm=1)$
i	$(i=1),(si=1),(bi=1)$
w	$(w=3),(sw=2),(bw=2)$
d	$(d=1),(bd=1),(wd=1)$
o	$(o=1),(bo=1)$
b	$(b=4),(sb=2)$
s	$(s=3)$

3.4. Generation of Recommendation using Markov model

This section explains the Markov model which is used to recommend the web pages. Here, Markov model is used for the identification of next pages based on the sequence of previously visited pages by the users. When a new user comes to get the recommendation of web page, the sequence path of that user is compared with the FP tree and it would recommend the web page using the probability definition.

Journal of Theoretical and Applied Information Technology
20th October 2013. Vol. 56 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Consider the input sequence which the user given as $p_1, p_2, \ldots\ldots, p_n$, where $p_1, p_2, \ldots\ldots, p_n$ are the sequence of web pages visited by a new user. Initially, the sequence which the user giving is compared with the FP tree and provides the matching results like

$p_1, p_2, \ldots\ldots, p_n, p_{n+1}, \ldots\ldots, p_k$ and

$p_1, p_2, \ldots\ldots, p_n, p_{n+1}, \ldots\ldots, p_l$ and

$p_1, p_2, \ldots\ldots, p_n, p_{n+1}, \ldots\ldots, p_m$. Here, every sequence has separate total support value and the most important sequence is recommended to the user by calculating the probability. The following equation gives the calculated value for the matched result.

$$p_{n+1}^{(1)} = \Pr o\left(p_1, p_2, \ldots\ldots, p_n, p_{n+1}, \ldots\ldots, p_k / p_1, p_2, \ldots\ldots, p_n\right)$$

$$p_{n+1}^{(2)} = \Pr o\left(p_1, p_2, \ldots\ldots, p_n, p_{n+1}, \ldots\ldots, p_l / p_1, p_2, \ldots\ldots, p_n\right)$$

$$p_{n+1}^{(3)} = \Pr o\left(p_1, p_2, \ldots\ldots, p_n, p_{n+1}, \ldots\ldots, p_m / p_1, p_2, \ldots\ldots, p_n\right)$$

This probability $\Pr o\left(p_{n+1}/p\right)$ is calculated using the sequence of all users in the FP tree constructed from biased frequency database.

$$\Pr o\left(p_1, p_2, \ldots\ldots, p_n, p_{n+1}, \ldots\ldots, p_k / p_1, p_2, \ldots\ldots, p_n\right) = \frac{T_{Sup}\left(p_1, p_2, \ldots\ldots, p_n\right)}{T_{Sup}\left(p_1, p_2, \ldots\ldots, p_n, p_{n+1}, \ldots\ldots, p_k\right)}$$

The final recommendation is based on the following equation:

$$p_{n+1} = \arg\max\left(p_{n+1}^{(1)}, p_{n+1}^{(2)}, p_{n+1}^{(3)}\right)$$

4. RESULT AND DISCUSSION

This section details the result obtained for proposed technique. The synthetic dataset is created to evaluate the performance of the result obtained based on the evaluation of precision, applicability and hit ratio is compared with the Niranjan *et al.*'s [16] and S.A.Sahaaya Arul Mary *et al.*'s technique [19] for the same evaluations.

4.1. Experimental setup and Dataset description

The system is applied in Java (jdk 1.6) which the system has i5 processor with 4GB RAM.

Evaluation metrics

The proposed system is evaluated in terms of precision, applicability and hit ratio. The description and the formula of the terms used to evaluate the proposed technique are as follows:

$$Precision = \frac{C^+}{C^+ + I^-}$$

$$Applicability = \frac{C^+ + I^-}{|N|}$$

$$Hit\ ratio = Precision \times Applicability = \frac{C^+}{|N|}$$

In the above formulas, C^+ denotes the number of correct recommendations and I^- denotes the number of incorrect recommendations and $|N|$ denotes the total number of given requests.

The term precision is defined as the number of correct recommendations which is divided by the sum of number of correct recommendations and number of incorrect recommendations. The term applicability is defined as it is the sum of number of correct recommendations and number of incorrect recommendations divided by the total number of requests given. The hit ratio is the product of precision and applicability.

4.2. Performance Comparison

This section compares the performance of proposed technique with the Niranjan *et al.*'s technique [16] and S. A. Sahaaya Arul Mary *et al.*'s technique [19]. The performance of both the techniques is compared in terms of precision, applicability and hit ratio. These terms are evaluated with different minimum support values.

The Fig.7 shows the performance of proposed technique and Niranjan *et al.*'s technique in terms of precision with different minimum support values. From this graph, the precision of proposed technique varying up and down while increasing the minimum support value and the precision of

Journal of Theoretical and Applied Information Technology
20th October 2013. Vol. 56 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Niranjan *et al.*'s technique get decrease after certain point of minimum support value while increasing the minimum support value. Here, when set the minimum support value as ten, the precision of proposed technique is 85.71 and the precision is 85.17 for Niranjan *et al.*'s technique and it is 45.45 for S.A.Sahaaya Arul Mary *et al.*'s technique. When the minimum support value is twenty, the precision of proposed technique is 82.85 and the precision is 85.17 for Niranjan *et al.*'s technique 45.45 for S.A.Sahaaya Arul Mary *et al.*'s technique. When the minimum support value is thirty, the precision of our technique is 82.71 and the precision is 83.79 for Niranjan *et al.*'s technique and it is 45.45 for S.A.Sahaaya Arul Mary *et al.*'s technique. For the minimum support value forty, the precision is 80 for our technique and it is 58.42 for Niranjan *et al.*'s technique and it is 45.45 for S.A.Sahaaya Arul Mary *et al.*'s technique. When set the minimum support value as fifty, the precision of proposed technique is 80 and the precision is 42.69 for Niranjan *et al.*'s technique.

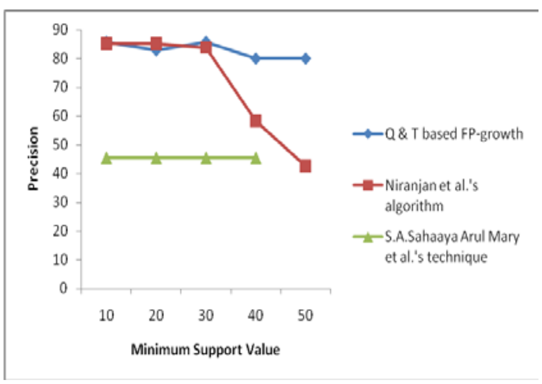

Fig.7Performance Comparison In Terms Of Precision

The performance in terms of applicability is shown in Fig.8 for proposed technique as well as Niranjan *et al.*'s technique and S.A.Sahaaya Arul Mary *et al.*'s technique with different minimum support values. Here, the applicability as hundred for proposed technique and for S.A.Sahaaya Arul Mary *et al.*'s technique for the different support values which is given. But the applicability of Niranjan *et al.*'s technique is hundred for the minimum support values ten, twenty and thirty. When the minimum support value is forty, the applicability is 92.06 for Niranjan *et al.*'s technique. When set the minimum support value as fifty, the applicability of Niranjan *et al.*'s technique is 58.96.

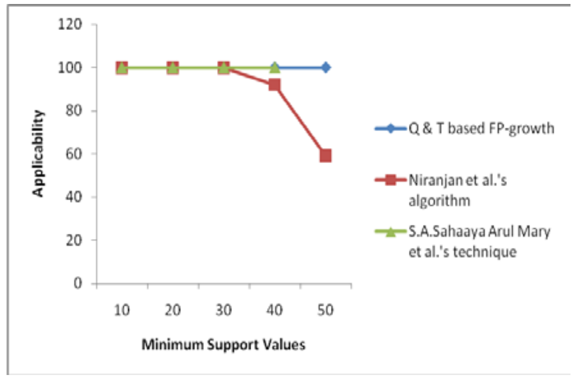

Fig.8 Performance Comparison In Terms Of Applicability

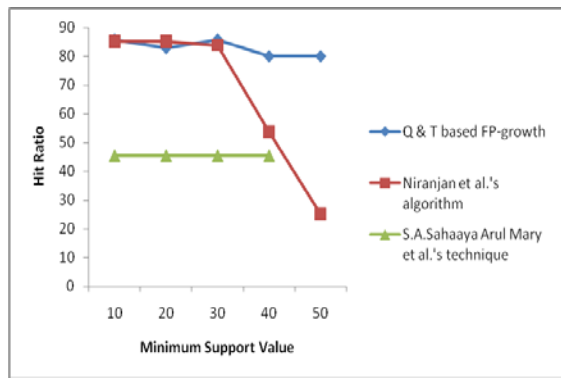

Fig.9 Performance Comparison In Terms Of Hit Ratio

The Fig.9 shows the performance comparison between proposed technique and the Niranjan *et al.*'s technique and S.A.Sahaaya Arul Mary *et al.*'s technique based on the hit ratio. Different minimum support values are used to evaluate the performance in terms of hit ratio. When set the minimum support value as ten, the hit ratio of our technique is 85.71 and the hit ratio of Niranjan *et al.*'s technique is 85.17. For the minimum support value twenty, the hit ratio is 82.85 for proposed technique and it is 85.17 for Niranjan *et al.*'s technique. When the minimum support value is thirty, the hit ratio is 85.71 for proposed technique and the hit ratio is 83.79 for Niranjan *et al.*'s technique. The hit ratio of proposed technique is 80 for the minimum support values forty and fifty and the hit ratio of Niranjan *et al.*'s technique is 53.79 for the minimum support value forty and 25.17 for the minimum support value fifty. The hit ratio obtained for S.A.Sahaaya Arul Mary *et al.*'s technique is 45.454 for all the minimum support values set as threshold.

5. CONCLUSION

In this paper a new technique is proposed to recommend the web page. To mine the frequent

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

patterns, quality and time based frequent pattern growth algorithm is used and to generate the recommendation, the Markov model is used. The experimentation is done using the synthetic dataset and the performance is evaluated in terms of precision, applicability and hit ratio. The performance of new technique is compared with the Niranjan *et al.*'s technique and S.A.Sahaaya Arul Mary *et al.*'s technique for the same terms. From the performance comparisons which is detailed in the result section, proposed technique shows better performance than the existing techniques.

REFERENCES

[1] Qingyan Yang, Ju Fan, Jianyong Wang and Lizhu Zhou, "Personalizing Web Page Recommendation via Collaborative Filtering and Topic-Aware Markov Model", *IEEE 10th International Conference on Data Mining*, PP. 1145-1150, 2010.

[2] Bonner, R E, "Some clustering techniques", Volume: 8, Issue:1, *IBM journal of Resarch and Development*, pp.22-32,1964.

[3] Richard Dubes, Anil K.Jain, "Clustering techniques: The user's dilemma", Volume: 8, Issue: 4, pp.247–260,1976.

[4] J A Hartigan, M A Wong , "A k-means clustering algorithm", *Published by:Blackwell Publishing*,vol 28, pp 100-108, 1979.

[5] R Forsati and M R Meybodi, "Effective page recommendation algorithms based on distributed learning automata and weighted association rules", *Expert Systems with Applications*, Vol. 37, No. 2, PP. 1316-1330, 2010.

[6] Jia Li and Osmar R Zaïane, "Combining Usage, Content, and Structure Data to Improve Web Site Recommendation", *E-Commerce and Web Technologies Lecture Notes in Computer Science*, Vol. 3182, PP. 305-315, 2004.

[7] Xiong Haijun and Zhang Qi, "Using Fuzzy Semantic Log for Rough Set Web Page Recommendation", *International Symposium on Information Engineering and Electronic Commerce*, PP. 249-253, 2009.

[8] A K Jain M, N Murthy, P J Flynn, "Dataclustering:A Review" Volume: 31, Issue: 3, Publisher: ACM, Pages: 264-323,1999.

[9] Sudipto Guha, Rajeev Rastogi, Kyuseok Shim,"CURE:an efficient clustering algorithm for large database" Volume 27 ,Issue 2,Pages 73-84, , June 1998.

[10] Berendt B, Hotho A, Stumme,"Towards semantic web mining. In: Proceedings of internationalsemantic web conference", *(ISWC) Springer, Berlin*, pp 264–278, 2002.

[11] Richards D, "Addressing the ontology acquisition bottleneck through reverseontological engineering", *KnowlInfSyst 6(4)*, pp.402–427, 2004.

[12] Rohn E,"Generational analysis of variety in data structures: impact on automatic data integrationand on the semantic web", *KnowlInfSyst 24(2)*, pp.283–304, 2010.

[13] Oren Etzioni, "The world-wide web: Quagmire or gold mine?", *Communications of the ACM*, vol. 39, no.11, pp.65–68, 1996.

[14] Kosala and Blockeel, "Web mining research: A survey", ACM SIGKDD Explorations Newsletter, vol. 2, no. 1, pp. 1–15, 2000.

[15] Faten Khalil, Jiuyong Li and Hua Wang, "Integrating Recommendation Models for Improved Web Page Prediction Accuracy", *in Proceedings of the thirty-first Australasian conference on Computer science*, Vol. 74, 2008.

[16] Utpala Niranjan, R B V Subramanyam and V Khanaa, "Developing a Web Recommendation System Based on Closed Sequential Patterns", *Communications in Computer and Information Science*, Vol. 101, no. 1, pp. 171-179, 2010.

[17] Utpala Niranjan, R B V Subramanyam, V Khana, "An Efficient System Based On Closed Sequential Patterns for Web Recommendations", *IJCSI International Journal of Computer Science Issues*, Vol. 7, Issue 3, No 4, May 2010.

[18] Mingyu Lu, Jiantao Sun and Yuchang Lu, "Association-based approach to Web page recommendation", *Fifth World Congress on Intelligent Control and Automation*, Vol. 4, PP. 3150-3154, 2004.

[19] S A Sahaaya Arul Mary and M Malarvizhi, "Integrated Web Recommendation Model with Improved Weighted Association Rule Mining", *International Journal of Data Mining & Knowledge Management Process (IJDKP)* Vol.3, No.2, pp.87-105, 2013.