International Journal of Engineering & Technology, 7 (4) (2018) 4024-4029

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET
doi: 10.14419/ijet. v7i4.16224
Research paper

A possiblistic clustering based biased Bayesian relevance feedback model for web usage recommendation

Sumanth Venugopal [1] *, Guruprasad Nagraj [2]

[1] Rajareddy Institute of Technology, Department of Computer science &Engineering, India
[2] New Horizon College of Engineering, Department of Computer science &Engineering, India
**Corresponding author E-mail: ssumanth6@gmail.com*

Abstract

World Wide Web (WWW) consists of a huge amount of web pages and links, provides massive information to the internet users. The de-velopments of websites have become challenging thing and size of web contents is more abundant. The web usage mining technique is em-ployed in web server log for extracting the user information. Presently, the Web Recommendation System (RS) is rapidly developing and major objective is generating the customized data for the end users. The RS is the platform that make personalized recommendations for a particular user by predicting the ratings for different items. In this paper, an efficient web RS that consists of two methods such as Possibil-istic Fuzzy C-Means (PFCM) and Relevance Feedback Biased Bayesian Network (RFBBN) methods are proposed. The PFCM algorithm clusters the similar web page users. In these clusters RFBBN model extract the relevant information and predict the relevant web pages. The proposed method reduces the loss of the end users. The experimental analysis demonstrated that the PFCM-RFBBN approach delivered the high priority of web pages and also recommended the related web pages. Finally, the experimental outcome showed that the proposed ap-proach improved accuracy in web page recommendation up to 31% compared to the existing methods.

Keywords: *Possibility Fuzzy C-Means; Recommendation System; Relevance Feedback Biased Bayesian; Web Usage Mining.*

1. Introduction

In present days, rapid growth of the Internet has shared huge volumes of information to the web users. A lot of websites like news portals, different life style website, e-learning websites and etc., frequently update the information. Each e-commerce website consists of its own RS, that directly or indirectly identifies the user's behaviours [1]. The RSs are broadly used in different areas such as academic, social commerce, e-commerce and etc. RSs majorly provide appropriate information on items or behaviour of personal users is known as personalized recommendation [2, 3]. The web personalization is the process of modifying the website based on the user requirement or group of users [4]. A RS is a web personalization technology that identifies products or information that are of importance to a specific user. One of the application of recommender models is to guide users during their visit on a Web site. Making such a recommendation requires predicting the page that is of interest to a user at a specific time [5]. The significant task in RS is to extract the interest of a user's navigational path through the site in a session [6].

In the context of recommender models, clusters of web sessions represent the groups of similar sessions that determine various navigational behaviours in a given web site [7]. These kinds of recommendation model's performance majorly depend on the cluster's capacity. Therefore, the researchers more concentrate on the improving the clustering methods accuracy and RS system performance. In existing research work, the user sessions are indicated as web page vectors [8]. Then, an efficient clustering technique is applied to estimate the numerical feature vectors in web session data [9]. For predicting the next request of a user, keep track of the order in which the pages are visited in a user session is an important process. The clustering methods such as K-Means, FCM, and etc. have been used for clusters the similar user sessions based on distance function. The distance function calculates the sequence alignment, longest common sequence and etc. [10]. Recently, various techniques have been proposed that exploit either the web access sequences in combination with knowledge extracted from the content of the web pages [11]. More specifically, the problem by extracting association rules from the web server log files using content based knowledge, by means of clustering the web pages by content is tackled in order to enhance the quality of the association rules. Then, the process tries to combine web access log files with web content, by extracting and using N-grams from the web pages [12]. In this paper, the PFCM clustering method is used for mapping the similar web pages, user sessions, web directory categories and behavioural patterns are identified. The proposed methodology is focused to overcome the drawbacks of both semantic similarity measurement and availability of the recommended suggestions for a new user based on their profile. The RFBBN model is used for retrieving the relevant web pages and it reduce the loss of the end users.

This paper is composed as follows. Section II surveys several recent papers in the web-page recommendation system. In section III, PFCM with RFBN for recommending the similar web page users is discussed. In Section IV, comparative analysis for proposed WPRS using MSNBC dataset is presented. The conclusion is made in Section V.

2. Literature review

The researcher suggested several techniques are suggested on the web page recommendation system analysis. In this scenario, a brief evaluation of significant contributions to the existing literatures is presented.

R. B. Wagh, and J. B. Patil, [13] developed a new web personalization technique to provide effective suggestions. This research accurately predicts the user's future requests, but the work was classified into two groups such as online phase as well as offline phase. In offline phase, the similar web pages were clustered using navigation pattern mining. Online phase generates the list of recommended web pages. The enhanced Graph based Partitioning Algorithm (GPA) was used for clustering of web pages and classified the current user activities more accurately. The method helped to improve the browsing experience of the user. The threshold used in the last phase of recommendations affected the GPA in lower coverage.

P. Damodharan, and C. S. Ravichandran, [14] presented a two-tier architecture for capturing recommendations in the form of the intuition list for the user. The intuition list consists of a list of pages visited by the user as well as a list of pages visited by other user of having similar usage profile. The method was experimented on various real time databases like MNSBC and CTI dataset. The experimental results represented that user navigation and to optimize the web server and improved the accuracy of the recommendations. The method improves the coverage parameter but, decrease the Live Session Window (LSW).

D. Anandhi, and MS Irfan Ahmed, [15] dealt with some techniques used for predicting the future visit of potential user in a website. In the web server log, identify the user session after that apply the Rough Set Clustering ((RST) method for clustering the significant sessions. The clustering done based on the maximum number of web pages visited in the web sites. The RST clustering approach achieved better clustering accuracy compared to the traditional approaches such as fuzzy clustering, graph partitioning, path prediction and graph partitioning. An experimental analysis, demonstrated that the proposed fuzzy algorithm achieved better results in terms of MNSBC and CTI database, but to predict the appropriate recommended page a number of iterations required.

R. Forsati et al., [16] proposed Binary Clustering Algorithms (BCA) to partition the binary session data. For clustering the binary data K-Means (KM) algorithm was used to achieve better clustering quality by combining with fine-tuning power. The KM method was compared with existing methods in terms of accuracy, minimum frequency, coverage and F-measure. The BCA method not able to handle directly the cold-start and dynamic pages requires some source of side information.

R. Katarya, and O.P. Verma, [17] presented a top-N clusters FCM algorithm based on sequential information on user's navigation. The method determined the similar users for the target user and also evaluated the weight for each web page. The method solved the problem of RS and also forecast a user's next web page visit. The experimental result showed that the method achieved three times better accuracy performance compared to the existing methods. The FCM method was not much focused on privacy, trust and social networks with hybrid intelligent systems.

3. Proposed methodology

The mining technique identify the frequently visited web pages from web logs have become a significant task for web usage mining to understand the behavior of users. This paper proposed an efficient web page recommendation system based on PFCM clustering algorithm and RFBBN model. The PFCM algorithm cluster the similar kinds of web pages and identifies the interesting behavioral patterns. After that, RFBBN model extracts the relevant web pages and it reduce the loss of the end users. The proposed web usage mining technique implemented in the working platform of JAVA. The figure 1 shows the architecture of proposed web recommendation system.

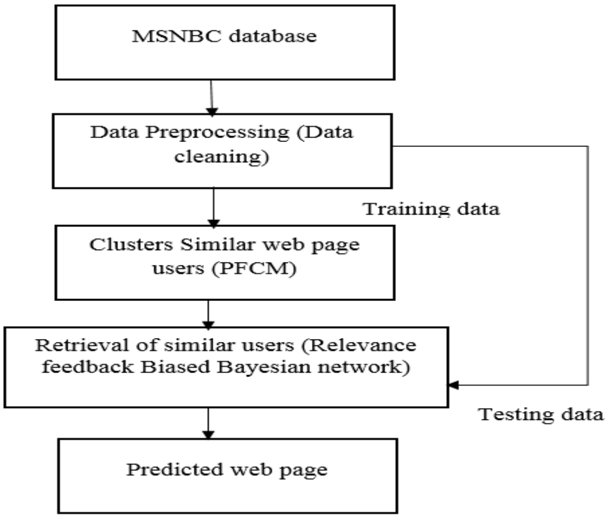

Fig. 1: Proposed Architecture.

3.1. Data acquisition

A real dataset is used for this experiment. The data set is taken from the UCI dataset repository that consists of Internet Information Server (IIS) logs from msnbc.com and news-related portions of msn.com [18]. Visits are recorded at the level of Uniform Resource Locator category and are recorded in time order. Each sequence in the dataset corresponds to page views of a user during that twenty-four-hour period. Each event in the sequence corresponds to a user's request for a page. The method takes data from this database for further processing.

3.2. Pre-processing

The input data is taken from the MSNBC database and it consists of irregular data. These data reduce the whole system performance. So, to remove the unwanted data preprocessing approach should be needed. The data cleaning approach is carried out to clean the unwanted data from the collected data in order to reduce the time. The main aim of preprocessing an input data is the data which is obtained from the logs may be incomplete, noisy and inconsistent. All these contents must be cleaned and tokenized, to reduce the dimensionality of terms by removing all unnecessary ones that have low discriminating values as well as reducing their ambiguity. The user sessions are identified from the preprocessed data with the help of the IP address and a predefined threshold. Each unique IP address is identified as a different user in session identification. The identification process defined by the set of pages visited by the particular user from the particular machine is identified. The relevant set of data constructed for training and testing process. The training process contains maximum 70% of data from the standard datasets and the remaining 30% of data is used for the testing process.

3.3. Possibility fuzzy c-means clustering

After preprocessing, the input data are connected to the clustering process using PFCM. After the mapping determination process and the associations between web pages, user sessions and web directory categories, the behavioral patterns are identified with the help of PFCM.

The PFCM is a data clustering technique, where each data point is a part of the cluster to a level indicated by the membership grade. In the clustering module, PFCM is dependent on the reduction of the objective function, which is illustrated in the equation (1), equation (2) and equation (3),

$$J_{PFCM}(\mathrm{U,T,V}) = \sum_{i=1}^{C}\sum_{j=1}^{n}\left(u_{ij}^{m} + t^{n}\right)d^{2}\left(x_{j}, v_{i}\right) \tag{1}$$

With following constraints:

$$\sum_{i=1}^{C}\mu_{ij}=1,\forall j\in\{1,\dots,n\} \quad (2)$$

$$\sum_{j=1}^{n}t_{ij}=1,\forall i\in\{1,\dots,c\} \quad (3)$$

Where, J_{PFCM} is the objective function, U is represented as the partition matrix, T is represented as the typicality matrix, V is denoted as the vector of cluster centers. The outcome of the objective function is achieved using an iterative approach, where the degree of membership and the cluster centers are mathematically represented in the equation (4), equation (5) and equation (6).

$$p\left(r_j|o_k\right)=\frac{1}{d_{jk}} \quad (4)$$

$$t_{ij}=\left[\sum_{k=1}^{n}\left(\frac{dx_j,v_i}{dx_j,v_k}\right)^{\frac{2}{n-1}}\right]^{-1}, 1\le i\le c, 1\le j\le n \quad (5)$$

$$V_i=\frac{\sum_{k=1}^{n}\left(u_{ik}^{m}+t_{ik}^{n}\right)x_k}{\sum_{k=1}^{n}\left(u_{ik}^{m}+t_{ik}^{n}\right)}, 1\le i\le c \quad (6)$$

Where, n is represented as the number of data points, c is represented as the number of cluster centers, which are described by the coordinates (x_j,v_i) and it is used to calculate the distance between cluster center and data sets. PFCM constructs possibilities and memberships with normal prototypes and cluster centers for every cluster. Choosing the objective function is the important aspect of the performance of cluster methodology for accomplishing better clustering. Whereas, the clustering performance is based on the objective function, which is utilized for clustering. For developing an effective objective function, the following set of requirements are considered.

- Distance between the clusters should be reduced.
- Distance between the data points, which allocated in the clusters should be reduced.

The desirability between the clusters and data are modelled by the objective function. Further, the objective function of PFCM is improved by using driven prototype learning of parameter α. The learning procedure α is a dependent exponential separation strength between the clusters and it is updated at every iteration. The parameter α is represented in the equation (7).

$$\alpha=\exp\left(-\min_{i\neq k}\frac{\|v_i-v_k\|^2}{\beta}\right) \quad (7)$$

Where, β is represented as sample variance, mathematically it is represented in equation (8),

$$\beta=\frac{\sum_{j=1}^{n}\|x_j-\bar{x}\|^2}{n} \quad (8)$$

Where, $\bar{x}=\frac{\sum_{j=n}^{n}x_j}{n}$

Then, a weight parameter is introduced to calculate the common value of α. Each point of the database consists of a weight in relationship with each cluster. So, the usage of weight function delivers a better classification outcome, especially in the case of noisy data. The general equation of weight function is determined in the equation (9)

$$w_{ij}=\exp\left(-\frac{\|x_j-v_i\|^2}{\left(\sum_{j=1}^{n}\|x_j-\bar{v}\|^2\right)\times c/n}\right) \quad (9)$$

Where, w_{ij} is denoted as the weight function of the point j with the class i. The step by step procedure of PFCM is represented in the figure.2, and also effectively explained below.

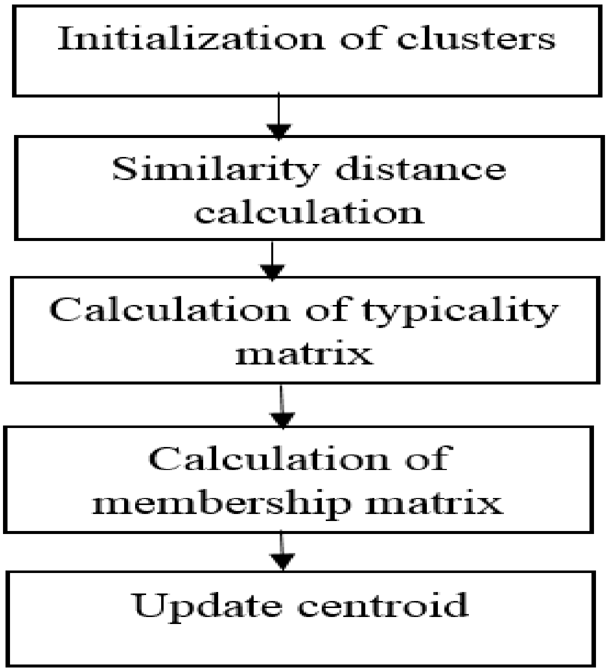

Fig. 2: PFCM Based Web Page Clustering Process.

The figure 2 represents the similar web page clustering process of PFCM algorithm. When the number of clusters are initialized, then the evaluation of the distance between the centroids and data points are calculated. Once the similarity distance is calculated, the typicality matrix is evaluated using PFCM algorithm. After that, with the help of FCM algorithm membership matrix is calculated. The FCM algorithm estimates the membership value of each data point. Finally, update the centroids of each and every cluster. The performance of the proposed technique is assessed in terms of user's gain and loss. Also, the performance of the PFCM utilized in the proposed technique is analyzed in terms of interesting of users behavioral patterns.

3.4. Relevance feedback biased Bayesian network

The Relevance Feedback (RF) technique is used in user data retrieval process to improve the final result set. In particular, the user gives feedback on the relevance of documents in an initial set of results. The gain and loss of the web users are identified and based on the gain and the loss the web directory is modified with the help of RFBBN technique which in turn reduce the loss of the end users. A Bayesian network data retrieval engine search the whole database to find the retrieval results. In our proposed work, the method builds a similar network, but only to process the relevant data. The network aggregates all the features compared with their separate networks for each feature in the feature set. Moreover, the Bayesian network model is a dynamic model, which can be adjusted with different relevant data specified by the user. The weight of a link from the feature vector node to the relevant image node represented by $p(o_k|r_j)$ is hard to obtain directly. However, the equation (10) is calculated by using Bayes' rule.

$$p(o_k|r_j)=\frac{p(r_j\vee o_k)p(o_k)}{p(r_j)} \quad (10)$$

Where, $p(r_j)\wedge p(o_k)$ are prior belief values of the component r_j and the relevant image o_k. The equation $P(r_j\vee o_k)$ is calculated in equation (11).

$$p\left(r_j \middle| o_k\right) = \frac{1}{d_{jk}} \tag{11}$$

Where, d_{jk} is the distance of relevant image o_k and the query example on component r_j . These weights are then normalized to make the sum of the weights equal to 1. The weight calculation is shown in the equation (12).

$$p\left(r_j \middle| o_k\right) = \frac{p\left(r_j \vee o_k\right)}{\sum_j p\left(r_j \vee o_k\right)} \tag{12}$$

The web page links weights and previous beliefs of all the nodes in the network have been allocated. The RFBBN method initializes the query node and perform the interference propagation throughout the network to update the belief value of all the relevant data. Then theses belief values can be used as the probabilistic ranking of the relevant data.

Biased Bayesian Network

The proposed methodology for relevance feedback in the BBN model is evaluating a set of retrieved documents obtained because of running query Q . The natural approach would be to instantiate each evaluated document to its corresponding relevance value. The clusters consist of relevant and irrelevant documents so, based on the relevance value it is easy to identify the relevant documents. The new query Q_1 would be shown in equation (13) and equation (14).

$$Q_1 = Q \wedge \left[D_{k1} = d_{k1}, ..., D_{kh} = d_{kh}, D_{kh+1} = d_{kh+1,......,D_{kh}} = d_{kb}^{-}\right] \tag{13}$$

$$\Delta_Q = Q_1 - Q_0 \tag{14}$$

The main idea is to first derive additional pairwise preferences from the data using the available significant information and in the optimization phase-bias, the optimization procedure draw a certain amount of samples from these additionally available data points. The gain and loss of the web users are identified. The gain is referred to as users found the requested web pages and loss is referred to as the users not possible to found the requested web pages. Then the web directory is modified with the help of RFBBN technique, which in turn reduce the loss of the end users. The web page retrieval process using proposed and existing methods experimental performance is shown in the following sections.

4. Experimental result

The implementation for experimental analysis performed in Netbeans (version 8.2) that was employed on PC with 3.2 GHz with i5 processor. In order to estimate the efficiency of the proposed algorithm, the performance of the proposed method was compared with Singular Value Decomposition (SVD), Markov Model, and Dynamic Nested Markov Model (DNMM) [19, 20] on the simulated recommended web pages. The performance of the proposed methodology was compared in terms of accuracy, precision, and prediction time.

4.1. Performance measure

Performance measure is defined as the relationship between the input and output variables of a system understand by employing the suitable performance metrics like Accuracy, Precision and Prediction Time. The general formula for calculating the evaluation metrics of the web page recommendation analysis is given in the equation (15) and equation (16).

Accuracy

Accuracy is the measure of statistical variability and a description of random errors. The general formula of accuracy for determining the web page recommendation analysis is given in the equation (15).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \tag{15}$$

Where, TP is represented as true positive, FP is denoted as false negative, TN is represented as true negative and FN is stated as a false negative.

Precision

The precision of learning and adapting the user profiles for each day of the experiment duration. The precision can be calculated in equation (16),

$$precision = \frac{\left|Number\ of\ correct\ learned \wedge adapted\ interests\right|}{\left|Total\ number\ of\ actuak\ interests\right|} \tag{16}$$

4.2. Experimental analysis data acquired from MSNBC database

MSNBC dataset is assessed for comparing the performance evaluation of existing methods and the proposed scheme in the experimental analysis. The table 1 represents the accuracy of the existing and proposed web recommendation system's performance analyzed with different five samples and three predictions.

Table 1: Accuracy Performance of Existing and Proposed Methods

Methods	No. of predictions	Random (%)	Sample 1(%)	Sample 2(%)	Sample 3(%)	Sample 4(%)	Sample 5(%)
SVD [19]	1	5.88	11.76	12.50	9.38	17.24	13.64
	2	11.76	23.53	18.75	18.75	24.14	13.64
	3	17.65	29.41	37.50	28.13	31.03	22.73
Markov Model [20]	1	60.89	53.58	61.35	69.85	68.25	67.25
	2	61.24	58.89	62.78	70.12	69.23	70.12
	3	68.56	60.98	63.56	71.41	71.28	72.36
PFCM-RFBBN	1	85.36	88.56	90.12	82.56	88.25	92.0
	2	84.78	87.23	91.25	87.57	90.69	92.87
	3	89.12	91.85	93.75	89.43	91.25	93.25

The SVD algorithm achieves 29.41%, 37.50%, 28.13%, 31.03% and 22.73% of accuracy in different five samples. The Markov model achieves 60.98, 63.56%, 71.41%, 71.28%, and 72.36%. The PFCM-RFBBN model achieves 91.85%, 93.75%, 89.43%, 91.25%, and 93.255 of accuracy in web page prediction. Compared to the existing WRS model, the proposed PFCM-RFBBN based WRS model shows better results.

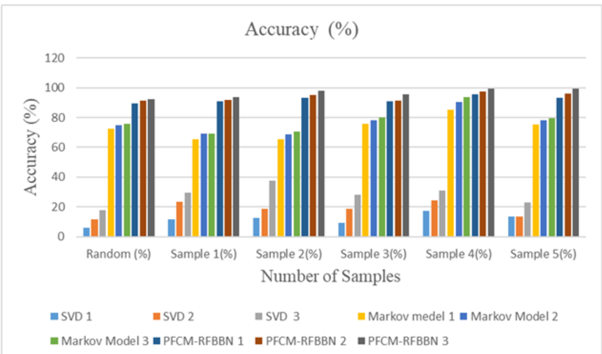

Fig. 3: Accuracy Performance.

The figure 3 represents the graphical representation of accuracy performance. In different samples, the web pages are correctly predicted using different RS. Correct predictions are found out by comparing original and predicted web pages; those predicted web pages that are equal to original web pages are considered as correct prediction.

Table 2: Precision Performance of Existing and Proposed Methods

Methods	No. of predictions	Random (%)	Sample 1(%)	Sample 2(%)	Sample 3(%)	Sample 4(%)	Sample 5(%)
SVD [19]	1	5.88	11.76	12.50	9.38	17.24	13.64
	2	11.76	23.53	18.75	18.75	24.14	13.64
	3	17.65	29.41	37.50	28.13	31.03	22.73
Markov Model [20]	1	72.42	65.35	65.23	75.85	85.25	75.25
	2	74.98	68.98	68.87	78.21	90.52	78.21
	3	75.65	69.25	70.56	80.14	93.89	79.63
PFCM-RFBBN	1	89.63	90.65	93.12	90.65	95.52	93.01
	2	91.14	91.71	95.25	91.49	97.35	95.87
	3	92.28	93.85	97.75	95.65	99.25	99.45

The table 2 depicts the precision performance of the WRS of existing and proposed models. The precision performance is calculated for different five samples. The SVD method is employed in five different samples and three times predicted the web pages. The precision of the SVD algorithm achieved 17.65%, 29.41%, 37.50%, 28.13%, 31.03% and 22.73% in MSNBC database. The Markov model achieved 75.65%, 69.25%, 70.56%, 80.14%, 93.89% and 79.63%. The PFCM-RFBBN model achieved 92.28%, 93.85%, 97.75%, 95.65%, 99.25%, and 99.45% of precision. The graphical representation of precision performance is shown in figure 4.

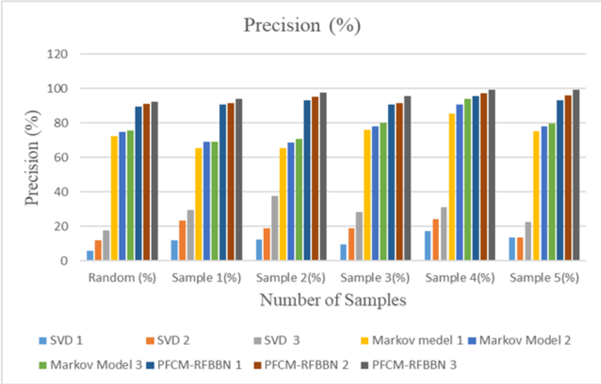

Fig. 4: Precision Performance.

Table 3: Prediction Time of WRS System Models

Time Evaluation	First Order Markov Model	First Order DNMM Model	Second Order Markov Model	Second Order DNMM Model	Third Order Markov Model	Third Order DNMM Model	PFCM-RFBBN
Prediction Time (sec)	12	12	34	23	34	30	9.2

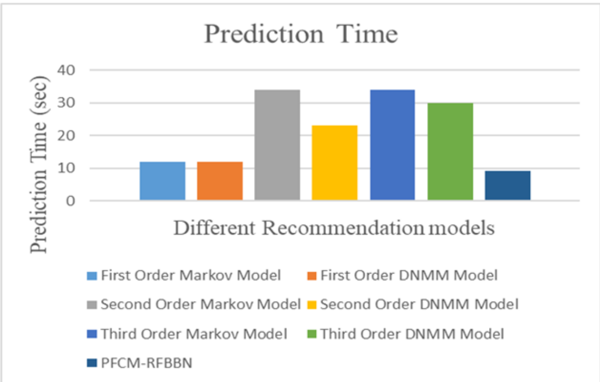

Fig. 5: Prediction Time.

The table 3 represents the prediction time of the web page RS performance of existing and proposed models. The prediction time taken by various orders of the model, which is measured on the testing file. The RS prediction time, majorly depends on the several factors like the network traffic, server response time etc. The number of web pages in the web log file effectively influences the prediction time. According to the figure 5, when the number of web pages increased, the prediction time also increased simultaneously. The first-order DNMM and Markov Model takes almost the same time to forecast the next accessed web page. Compared to the higher-order MM model, the higher-order DNMM takes minimum prediction time. The PFCM-RFBBN model achieved 9.2msec of prediction time. Compared to all the other methods, the proposed PFCM-RFBBN model achieved better results.

5. Conclusion

Web page RSs are used to recommend the future web page views to WWW users. In the recent days, huge amount of web pages, web sites and web documents are available, so the search engine returns the thousands of related links to a search query. The outline of the current websites has overwhelmed the web users by offering many choices. Hence, the web users tend to make poor decisions when surfing the Web, due to an inability of managing the enormous amount of information. In this research paper, PFCM-RFBBN model is proposed for predicting the efficient web pages with the help of PFCM algorithm clusters the similar web page users. The RFBNN model extracts the relevant information from these clusters and predict the relevant web pages. An experimental analysis demonstrated that the PFCM-RFBBN approach obtained 31% of improvement in accuracy and 6% of improvement in precision of WRS for MSNBC database. The PFCM-RFBBN model performance is compared with the existing models namely SVD, Markov model and DNMM. In the future work, for further improving the web page recommendation accuracy, possible to develop a new pattern mining approach with an adaptive clustering methodology.

References

[1] Tewari AS, & Barman AG (2018), "Sequencing of items in personalized recommendations using multiple recommendation techniques," *Expert Systems with Applications*, Vol. 97, pp. 70-82. https://doi.org/10.1016/j.eswa.2017.12.019.

[2] Seo, YD, Kim YG, Lee E, Seol KS, & Baik DK (2018), An enhanced aggregation method considering deviations for a group recommendation," *Expert Systems with Applications*, Vol. 93, pp. 299-312. https://doi.org/10.1016/j.eswa.2017.10.027.

[3] Thiyagarajan R, Thangavel K, & Rathipriya R (2014), "Recommendation of web pages using weighted K-means clustering," *International Journal of Computer Applications*, Vol. 86, No. 14.

[4] Duwairi R, & Ammari H (2016), "An enhanced CBAR algorithm for improving recommendation systems accuracy," *Simulation Modelling Practice and Theory*, Vol. 60, pp. 54-68. https://doi.org/10.1016/j.simpat.2015.10.001.

[5] Zheng L, Qu Y, Qian X, & Cheng G (2018), "A hierarchical co-clustering approach for entity exploration over Linked Data," *Knowledge-Based Systems*, Vol. 141, pp. 200-210. https://doi.org/10.1016/j.knosys.2017.11.017.

[6] Adeniyi DA, Wei Z, & Yongquan Y (2016), "Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method," *Applied Computing and Informatics*, Vol. 12, No. 1, pp. 90-108. https://doi.org/10.1016/j.aci.2014.10.001.

[7] Chawla S (2016), "A novel approach of cluster based optimal ranking of clicked URLs using genetic algorithm for effective personalized web search," *Applied Soft Computing*, Vol. 46, pp. 90-103. https://doi.org/10.1016/j.asoc.2016.04.042.

[8] Li H, Xu Z, Li T, Sun G, & Choo KKR (2017), "An optimized approach for massive web page classification using entity similarity based on semantic network," *Future Generation Computer Systems*, Vol. 76, pp. 510-518. https://doi.org/10.1016/j.future.2017.03.003.

[9] Xie X, & Wang B (2018), "Web page recommendation via twofold clustering: considering user behavior and topic relation," *Neural*

Computing and Applications, Vol. 29, No. 1, pp. 235-243. https://doi.org/10.1007/s00521-016-2444-z.

[10] Xuan J, Luo X, Zhang G, Lu J, & Xu Z, "Uncertainty analysis for the keyword system of web events," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, Vol. 46, No. 6, pp. 829-842. https://doi.org/10.1109/TSMC.2015.2470645.

[11] Husamuddin Md, & Mazarbhuiya FA (2015), "Clustering of Locally Frequent Patterns Over Fuzzy Temporal Datasets," *International Journal of Computer Trends and Technology*, Vol. 28, No. 3, pp. 131-134. https://doi.org/10.14445/22312803/IJCTT-V28P124.

[12] Deng L, & Li D (2017), "Multimedia data stream information mining algorithm based on jointed neural network and soft clustering," *Multimedia Tools and Applications*, pp. 1-24. https://doi.org/10.1007/s11042-017-4964-7.

[13] Wagh RB, & Patil JB (2017), "Enhanced Web Personalization for Improved Browsing Experience," Advances in Computational Sciences and Technology, Vol. 10, No. 6, pp. 1953-1968.

[14] Damodharan P, & Ravichandran CS (2016), "Certain Investigation on Automatic Recommendation for Online Users Using Web Usage Mining," Certain Investigation on Automatic Recommendation for Online Users Using Web Usage Mining, Vol. 1, No. 1, pp. 72-77.

[15] Anandhi D, & Ahmed MSI, "Prediction of user's type and navigation pattern using clustering and classification algorithms," *Cluster Computing*, pp. 1-10, 2017.

[16] Forsati R, Moayedikia A, & Shamsfard M (2015), "An effective Web page recommender using binary data clustering," *Information Retrieval Journal*, Vol. 18, No. 3, pp. 167-214. https://doi.org/10.1007/s10791-015-9252-4.

[17] Katarya R, & Verma OP (2017), "An effective web page recommender system with fuzzy c-mean clustering," *Multimedia Tools and Applications*, Vol. 76, No. 20, pp. 21481-21496. https://doi.org/10.1007/s11042-016-4078-7.

[18] Sahoo S, & Ratha BK (2018), "Rapid Frequent Pattern Growth and Possibilistic Fuzzy C-means Algorithms for Improving the User Profiling Personalized Web Page Recommendation System," *International Journal of Intelligent Engineering and Systems*, Vol. 11, No. 2.

[19] Mishra R, Kumar P, & Bhasker B (2015), "A web recommendation system considering sequential information," *Decision Support Systems*, Vol. 75, pp. 1-10. https://doi.org/10.1016/j.dss.2015.04.004.

[20] Nigam B, Tokekar S, & Jain S (2015), "Evaluation of models for predicting user's next request in web usage mining," *International Journal on Cybernetics & informatics (UCi)*, Vol. 4, pp. 1-13. https://doi.org/10.5121/ijci.2015.4101.