



## A Proficient Web Recommender System using Hybrid Possibilistic Fuzzy Clustering and Bayesian Model Approach

Sumanth Venugopal<sup>1\*</sup>      Guruprasad Nagraj<sup>2</sup>

<sup>1</sup> *Department of Computer science & Engineering, Rajareddy Institute of Technology, India*

<sup>2</sup> *Department of Computer science & Engineering, New Horizon College of Engineering, India*

\* Corresponding author's Email: [ssumanth6@gmail.com](mailto:ssumanth6@gmail.com)

**Abstract:** In day to day life, the process of projecting the correct information to the authorized person is more difficult, which makes complexity to the decision making process. Web Page Recommendation Systems (WPRS) used in various fields to identify the customer needs and to help the users to take appropriate decisions over the service or product according to his/her preference. The group of users with similar preference will be identified by using Possibilistic Fuzzy C-Means (PFCM) algorithm with an S3I Similarity Measure (SM). The proposed method will determine the gain and loss of the web users based on the web directories which can be modified by using Relevance Feedback Bayesian Network (RFBN) technique. The experimental results are conducted on the MNSBC dataset and the outcomes are compared with the existing methods like Singular Value Decomposition (SVD) methods. The method predicts the accuracy up to 85% when compared with the existing methods and the outcome results proved the effectiveness of the PFCM - RFBN method.

**Keywords:** Web page recommendation systems, Possibilistic fuzzy c-means, Relevance feedback Bayesian network, MNSBC dataset, Singular value decomposition.

### 1. Introduction

As there are very fast development and a wide range of application of the internet, World Wide Web (WWW) has become an interesting medium for pool, exchange, sharing of information and efficient channel for collaborative work [1]. The growth of web technologies and internet had extended to various fields such as education, entertainment and e-shopping. It is a tough job to satisfy the customer needs by finding the suitable needs of the users [2]. Existing recommendation systems have many complexities like recommending based on relations between the products in previous transactions and another possibility is recommending based on results given by user about a product and similar feedbacks from different users. The second kind of recommendation system is called as Collaborative Filtering (CF) methods [3]. The CF methods have several disadvantages in their nature, especially usual neural network algorithms will be based on performance and their scalability. And also the

number of iterations required by this method is creating a sparsity problem if an insufficient evaluation is done by the user. Mining algorithms have been employed for resolving the sparsity problem in CF based recommendation systems [4]. These methods are introduced before user logs into the system. Hence, the user response is not delayed by this mining algorithm.

The main aim of the web mining is used to extract knowledge, such as usage logs of websites, hyperlinks between web documents and mine useful knowledge from the web data. [5]. The web is a universal information platform space which can be accessed by companies, universities, businessman etc. Generally, internet hold there are numerous sources of information like internal sources and external sources [6]. The web experience of a user can be personalized to user's taste and this action called as web personalization. The firm initiated is defined as preferences of individual rather than a group's interest in web personalization when compared to mass

customization (i.e. Customer initiated) [7]. Based on user's browsing behavior and demographic data, the navigation and content are personalized to each user in web personalization. It might be simple advertisements, product suggestions or complete personalization where entire content over the website varies from user to user. In the case of mining through the internet, the problem of analyzing user behavior is a major issue in recommending web pages [8]. This proves that the increase website visitors will not increase by including lots of data on the website; it will improve by simplifying the access of content in the website.

Web Usage Mining (WUM) has been then employed to serve more people by understanding their usage behaviors and recommends web pages to them based on their activity and also improves the computational process of a website [9]. Data mining refers to the exploring the huge raw data and identifies the pattern in large datasets using Artificial Intelligence (AI) or machine learning techniques. WUM is further classified into web server log which is the data of access history stored by the server. To identify usage patterns of the user many pattern mining techniques can be employed which can discover relations between variables in huge databases [10]. In the proposed work, the technique PFCM Algorithm is used for mining frequent pattern of user logs based on the user behavior and according to the similarity of the content, the related web page is recommended to the user. The proposed methodology is focused to overcome the drawbacks of both semantic SMment and availability of the recommended suggestions for a new user depends upon their profile. The proposed methodology will suggest a web page based on three factors, namely user log, the web page content relevant to user search and the user profile. Based on these three factors proposed methodology suggest a web page to the user and proves the accuracy of prediction and recommended web pages to be better than other existing systems.

This paper is composed as below. Section II surveys several recent papers in the webpage recommendation system. In section III, PFCM with S3I SM and RFBN for recommending the related web pages. In Section IV, comparative analysis for proposed WPRS using MSNBC dataset is presented. The conclusion is made in Section V.

## 2. Literature review

Several techniques are suggested by researchers in the web-page recommendation system. In this section, a brief evaluation of some important

contributions to the existing WUM techniques are presented.

R. Katarya, and O.P. Verma, [11] presented a web-based recommender system based on pattern information of user's web behavior. This method tried to solve the problem of user's web page forecasting by using Fuzzy C Means clustering with similar users as a cluster. This methodology tried to recommend web pages to the user based on the usage pattern of the user and achieved an accuracy of 33% with 5000 weblog entries. The main issue with this kind of approach was that the method doesn't concern about the content of the web page that was recommending and also not suitable for the new user, since the method required at least six weblogs per user.

X. Xie, and B. Wang [12] proposed a twofold clustering based WPRS with the aim to achieve better efficiency and effectiveness with the input as the popularity and date of entry parameters. This method combined both density-based clustering and K-means clustering to cluster data to get the number of clusters and the initial centroid of each cluster. This method has a drawback of lagging from the concern towards user's individual preferences and has a randomized and complex clustering mechanism which doesn't say the accuracy of relevance to the web page to the user requirements.

J. Liu, C. Fang, and N. Ansari [13] had introduced the concept of the Request Dependency Graph (RDG), which modelled the dependency relationships among HyperText Transfer Protocol (HTTP) requests to analyze the behavioral characteristics of web traffic. The graph was made by using this methodology by mining the causal information and temporal among aggregated HTTP requests. The method enhanced the quality of web usage mining, the analysis of user behaviors and the performance of network and web server. The method has a drawback of large and complex RDG to decompose and visualize from massive traffic logs.

C. Suruchi, [14] proposed an approach which used Genetic Algorithm (GA) to generate the cluster based optimal ranked clicked URLs for effective Personalized Web Search (PWS). The study was conducted on the clustered query session data set to confirm the effectiveness of personalized web search. The performance of the method was evaluated during online processing using the average precision of the search results with the optimal ranked clicked URLs in the different domains. The effectiveness of PWS based on clustered query session was increased and satisfied the needs of the user. The method doesn't classify all the web pages based on some key features or any specific link towards each web page.

R. Geetharamani, P. Revathy, and S.G. Jacob [15], focus on predicting the forecasting the web pages based on the user history of web pages enlisted in their database based on their usage behavior using Apriori Prefix Tree (APT) algorithm. The rules generated based on these algorithms was ranked based on their support, confidence and lift evaluation measures. The main pitfall of this approach was that the method doesn't focus on semantic similarity of the web pages and also the system doesn't recommend the user based on the suggestion of similar user. This will create a major drawback when comes in the case of a new user or a user with fewer user logs will receive bad forecasted web pages.

S. Sahoo, and B.K. Ratha [16] implemented a personalized web recommendation system to overcome the issues such as content overspecialization and sparsity. The number of sequences were converted into web pages in the approach. The identification of neighbourhood of user profiles were provided by an effective Possibilistic Fuzzy C-Means (PFCM) clustering algorithm. The frequent web pages were mined by using the association rule mining technique called Rapid Frequent Pattern Growth (RFPG) algorithm. The experiments were conducted on training datasets such as MSNBC dataset, and castings technology international dataset. The results showed that the high priority of web pages were delivered by the PFCM-RFPG approach. The accuracy of the method provides poor performance because the method highly concentrates on sparsity problem.

### 3. Proposed methodology

The specific type of information filtering system techniques (i.e. webpage recommender system) attempts to recommend the social items or elements which is related to the users. The web recommender system may also suffer from the following issues that are the system provide the pattern information about user's behavior doesn't concern about the content of the web pages and also not suitable for the new user. The WPRS method is lagging towards user's individual preferences and the accuracy of the relevance of the web page is not efficient due to complex clustering mechanism to the user requirements. The major drawback of the Context-Aware Personalization System (CAPS) is less effective because at each node which contained limited knowledge and the approach is overloaded due to a large number of users. The method suffered from a cold start problem and also suffered from the problem of over-specialization. The PWS method is

lacking from the classification of all the web pages based on any specific link towards each web page.

To overcome the above-mentioned issues, the methodology includes PFCM-RFBN for web personalization. Initially, the web usage data will be collected from the access log files of ISP cache proxy servers and the preprocessing approach will be carried out to clean the unwanted data from the collected data in order to reduce the time. Subsequently, user sessions will be identified from the preprocessed data by using IP address and a predefined threshold. Next, the web directory will be initialized by analyzing the association of the web pages. After determining the mapping and the associations between Web pages, user sessions, and Web directory categories, interesting behavioral patterns will be identified by using PFCM. Subsequently, the gain and the loss of the web users will be determined and based on the gain (the users find what they are looking for) and the loss (the users do not find what they are looking for) the web directory will be modified with the help of the RFBN technique which in turn reduce the loss of the end users. The structure of the proposed WUM technique is described as in the following Fig. 1.

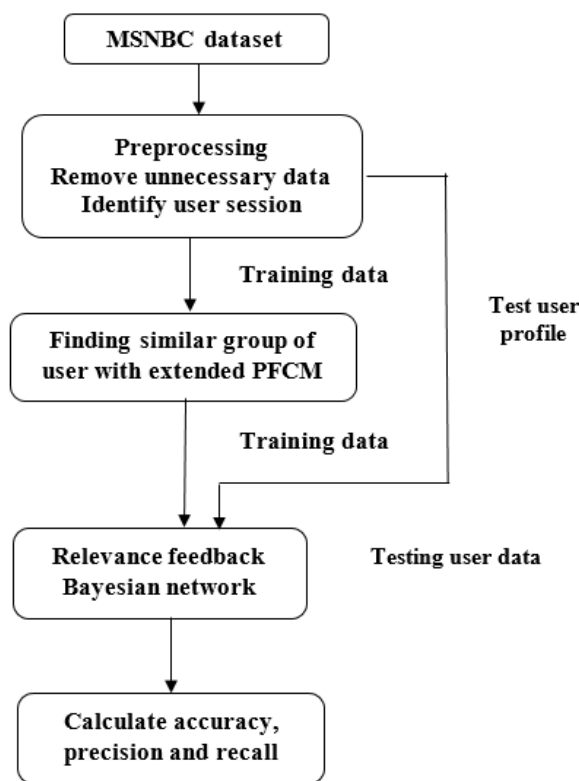


Figure.1 Structure of the web usage mining technique

### 3.1 Preprocessing

#### 3.1.1. Remove unnecessary data

The web usage data are collected from the access log files of ISP cache proxy servers. After the collection process of usage data, the preprocessing approach should be needed. The 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 which have low discriminating values as well as reducing their ambiguity. After that, access pattern analysis is unsuitable in raw proxy server log files and these log files requires preprocessing effectively to remove irrelevant data for analysis.

#### 3.1.2. Identify user session

The user session will be identified from the preprocessed data using 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 set of URLs forms the session should satisfy the requirements and the time elapsed between two consecutive requests is reduced. After preprocessing, the data are 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.

### 3.2 Possibilistic fuzzy c-means clustering algorithm (PFCM)

The large input dataset makes handling the gigantic measure of data to a great degree mind-boggling, furious and time-consuming. Furthermore, the various data leads to poor results by the increments of mistakes and to resolve this issue, the data is preprocessed by clustering technique which is used before characterization. The pre-processed input data is connected to a clustering method called PFCM which is used to improve the clustering process and accuracy of the method. The PFCM constitutes a data clustering technique where each data point is a part of the cluster to a level indicated by the membership grade. The PFCM is dependent on the reduction of the objective function which is described in the following Eq. (1).

$$PFCM_{fun}(M, T, C, S) = \sum_{k=1}^n \sum_{i=1}^c (aM_{ik}^m + bT_{ik}^\varphi) \times \|S_k - C_i\|_A^2 + \sum_{i=1}^c \varphi_i \sum_{k=1}^n (1 - T_{ik})^\varphi \quad (1)$$

Where,  $M$  described the membership matrix,

$T$  represents the typicality matrix,

$C$  labelled the resultant cluster center,

$S$  defined the set of input SM and the constant  $a \wedge b$  represent the comparative significance of membership of fuzzy and typicality values in the objective function.

The gradual procedure of PFCM is described effectively as follows,

#### Step 1: Initialization

At first, the cluster is furnished by the user, which is identical in respect of every segment.

#### Step 2: Calculation of Distance matrix

At the outset, the number of clusters ( $C$ ) is furnished by the user, which is identical in respect of every segment. When the  $C$  is determined, the evaluation of the distance between the centroids and data point for each segment is carried out. In the document, the Euclidian distance function is elegantly employed to evaluate the distance between centroids and data points with the ultimate motive of evaluating the distance matrix. Further, the distance matrix is determined for each and every cluster by using S3I Similarity Matrix (SMa). The SMa can be explained as follows,

#### Step 2.1: S3I similarity matrix

In our system, the researchers formed the knowledge about web users can be acquired by clusters and the learning capability can be used in classification techniques for generating recommendation. The user may contain multiple clusters because of his various interests for processing the SMa. Clustering algorithm is used to group the users according to the SMa. These matrices are used to evaluate the similarity among users and objects. The Hamming Distance, Levensthein Distance, Longest Common Sub Sequence (LCS) are examples of sequence-based similarity/distance measures. The proposed method uses an S3I SM in Eq. (4) to capture the sequential behaviour of the user while forming clusters. SM is a linear combination of Jaccard SM in Eq. (3) and the Length of Longest Common Subsequence (LLCS) can be measured by sequence SM. The Eq. (2) describes the sequence similarity between two clusters.

$$L1 = |S1|$$

$$L2 = |S2|$$

$$D_{ki} = S3I_{ki}$$

$$SeqSim_{betweenS1 \wedge S2} = \frac{LLCS}{Max(L1,L2)} \quad (2)$$

$$S2 = \frac{SetSim(JaccardSimilarity)_{betweenS1 \wedge S2}}{\frac{|S1 \cap S2|}{|S1 \cup S2|}} \quad (3)$$

$$S3I = p \times SeqSim + q \times SetSim \quad (4)$$

**Step 3: Calculation of typicality matrix**

The typicality matrix is calculated after the estimation of the distance matrix which in turn the PFCM provides the typicality matrix. The Eq. (5) can be represented as,

$$T_{ki} = \frac{1}{1 + \left[ \frac{D^2(x_k, v_i)}{\eta_i} \right]^{\frac{1}{m-1}}} \quad (5)$$

**Step 4: Calculation of membership matrix**

The PFCM gathered the membership value of data point which can be accessed. These membership values are used for the evaluation of the membership matrix which is denoted by M. The membership equation is as following Eq. (6),

$$U_{ik} = \frac{1}{\sum_{j=1}^c \left( \frac{\|x_k - v_i\|}{\|x_k - v_j\|} \right)^{\frac{2}{m-1}}} \quad (6)$$

**Step 5: Updation of centroid**

The modernization of the centroids is updated after the generation of clusters. The relative procedure is performed again and again till the modernized centroids of each and every cluster becomes identical similar in successive iterations. The performance of updation of the centroid is as follows in the form of Eq. (7),

$$v_i = \frac{\sum_{k=1}^n (U_{ik}^m + T_{ik}^n) x_k}{\sum_{k=1}^n (U_{ik}^m + T_{ik}^n)}, 1 \leq i \leq c \quad (7)$$

**3.3 Relevance feedback Bayesian network (RFBN)**

**3.3.1. Bayesian belief network**

A collection of random variables and their dependencies represents the graphical view of Bayesian belief network. In a BN, vertices represent the random variables  $X = \{x_1, x_2, \dots, x_n\}$  and direct

arcs correspond to conditional dependencies between variables in the graph. The  $Pa(x_i)$  can be defined as parents of a node  $x_i$ . The conditional probability  $P(x_i \vee Pa(x_i))$  represents the interactions between nodes. The other variables that are contained in a BN include prior beliefs, evidence and posterior probabilities. By applying the conditional probabilities, the joint probability distribution over the complete set of variables can be expressed as Eq. (8).

$$P(X) = P(x_1, x_2, \dots, x_n) = \prod_{i=1}^n P(x_i \vee Pa(x_i)) \quad (8)$$

The BN model provides the probabilistic relationships among objects for making the proposed method to handle the probability distributions over variable's situations. Furthermore, the structural representation of the BN is easy to build and highly adaptive. The feature representations are integrated by BN which provides a good framework and the also BN offers easy maintenance while adding new features.

**3.3.2. Relevance feedback Bayesian network**

The proposed method needs to formulate the problems to represent the distinct elements in terms of creating a set of variables for relevant data adoption by applying BN to the proposed approach. Then, the dependency relationships between variables are generated, i.e. the links are created from the parent to child nodes. The numeric probabilities need to be assessed for each link and variables after the generation of relationship. In the existing method, the retrieval results are found by searching the whole database with the help of BN data retrieval engine. The proposed work uses the similar network of BN, but only for processing the relevant data. The network aggregates all the features compared with their separate networks for each feature in the feature set. Moreover, the BN model can be adjusted with various relevant data specified by the user. The feature vector provides the weight of a link to the relevant image node that is represented by  $p(o_k | r_j)$  is hard to obtain directly. However, the equation can be calculated by using Bayes' rule presented in Eq. (9).

$$p(o_k | r_j) = \frac{p(r_j \vee o_k) p(o_k)}{p(r_j)} \quad (9)$$

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  $P(r_j \vee o_k)$  can be calculated as Eq. (10).

$$p(r_j|o_k) = \frac{1}{d_{jk}} \quad (10)$$

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 which are represented in Eq. (11)

$$p(r_j|o_k) = \frac{p(r_j \vee o_k)}{\sum_j p(r_j \vee o_k)} \quad (11)$$

The method can initialize the query node and perform the inference propagation throughout the network for updating the belief values of all the relevant data after assigning the prior beliefs of all the nodes and weights of links in the network. Then, these values can be used as the probabilistic ranking of the relevant data.

### 3.3.3. Network evaluation for relevant data adoption

The system obtains new information about the current relevance level based on the user's feedback information, which is used for relevant data adoption in the BN model for all the data to the query. According to the result of query and user's relevance feedback, the belief values of all the relevant data nodes in the network are re-evaluated at each stage of iteration. The algorithm for the relevant data adoption is described as follows:

- The BN model can be built by using all the relevant data and the links are calculated after the initial feedback.
- The inference propagation is performed for updating the belief values of all the nodes in the network.
- The relevant data can be selected by choosing the belief value which is above the threshold as the positive feedback information.
- The chosen relevant data is used to update the link weights. The updated belief values are used to set as new prior beliefs for the next iteration.
- The updated positive feedback information is used to start a new iteration of retrieval.
- If the user continues to visit the same web pages, then the priority will be double.
- Else stop the retrieval process and wait for the new query.

## 4. Experimental outcome

The outcome of the PFCM-RFBN are compared with the existing method SVD in [17] are validated by using performance measures. R. Mishra, P. Kumar, and B. Bhasker [17] implemented a recommendation system using similarity upper approximation and SVD to generate the recommendations for users. The overlapping clusters were generated by rough set based clustering technique. The method tried to predict the user's next pages, but in this approach, navigation behavior of users was used. Demographic variables were needed to solve the cold-start issue. The performance of PFCM-RFBN is performed on the MNSBC data and predict the evaluation matrices are described in the following section:

### 4.1 MNSBC database description

The real dataset is used for the experiments that are taken from the repository of UCI dataset. The news-related portions of msn.com and Internet Information Server (IIS) logs for msnbc.com are presented in the UCI dataset. The visits of the weblogs of the users are recorded by the Uniform Resource Locator in time order. In the dataset, each sequence resembles the pages viewed by the users and the event in the sequence represent the requests of a customer for pages.

### 4.2 Performance measure

#### 4.2.1. Accuracy

The accuracy is measured as the proportion of the number of positively mapped web pages to correct concepts from an RO to the all web pages visited by the user. The calculation of the accuracy is evaluated by the following Eq. (12).

$$Accuracy = \frac{|Positivemappedwebpages|}{|Allwebpages|} \quad (12)$$

#### 4.2.2. Precision

The precision of learning and adapting the user profiles for each day of the experiment duration. The precision can be calculated as Eq. (13).

$$Precision = \frac{|Numberofcorrectlearned\wedge adaptedinterests|}{|Totalnumberofactualinterests|} \quad (13)$$

The precision in this equation is computed as the ratio of correctly modelled interests in user profiles to the actual number of interests in the scenarios.

Table 1. Accuracy comparison of existing and proposed method using MSNBC dataset

Methods	No. of predictions	Random (%)	Sample 1(%)	Sample 2(%)	Sample 3(%)	Sample 4(%)	Sample 5(%)
SVD [17]	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
PFCM-RFBN	1	24.83	24.57	41.62	32.68	56.83	62.12
	2	46.3	53.21	46.25	65.98	62.51	64.32
	3	76.45	78.21	73.29	82.15	84	82.56

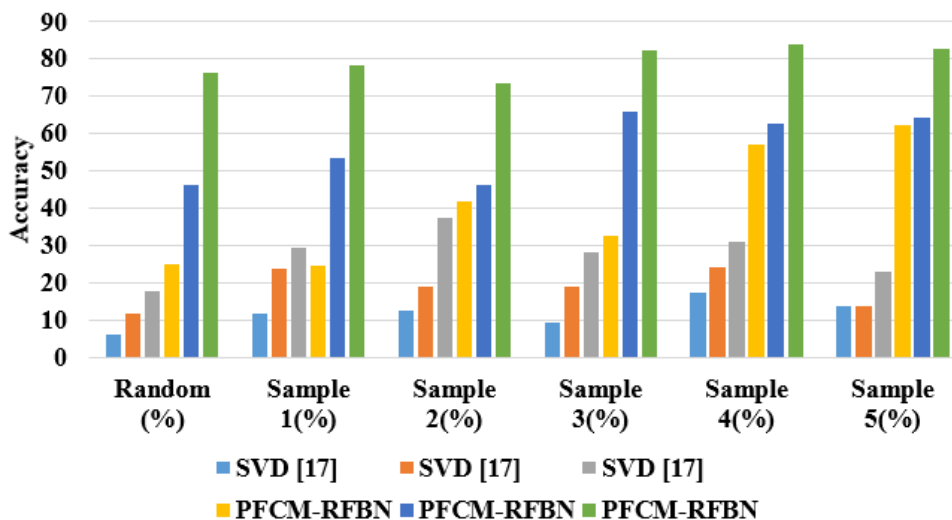


Figure.2 Performance comparison using MSNBC dataset

### 4.3 Results of accuracy for MNSBC dataset

MSNBC dataset is assessed for comparing the performance evaluation of existing methods and the proposed scheme in the experimental analysis. In Table 1, the accuracy of the proposed and existing methodology is compared to the five samples and three predictions. The average accuracy of existing methods (SVD) for each prediction delivers 37.50%, 31.03%, 29.41% and 28.13%. Likewise, the average accuracy of the proposed method (PFCM with S31 SM and RFBN) for each prediction delivers 84%, 82.563%, 78.21% and 76.45%. The performance evaluation of MSNBC dataset is denoted in the following Fig. 2.

### 4.4 Results of Precision for MNSBC dataset

Precision is a metric of information retrieval domain and is considered as an important parameter of recommendation systems along with accuracy. The confusion matrix (Table 2) represents the relevant items among all retrieved items. The proposed method gives a recommendation that is generated based on a mathematical model and computation for all the inputs. The performance of

Table 2. Precision comparison of existing and proposed method

Methods	MSNBC dataset			
	First (%)	Second (%)	Third (%)	
SVD [17]	M=20	12.61	19.8	28.12
	M=22	10.6	16.7	24.38
	M=32	11.02	17.9	28.8
	Random	5.88	11.8	17.65
	PFCM-RFBN	M=20	42.36	53.26
	M=22	40.36	52.56	87.56
	M=32	42.78	58.69	89.32
	Random	43.92	54.62	87.26

the various clustering techniques over several methodologies has been illustrated and the comparison of proposed method performance over the other existing methodologies in terms of precision for MSNBC datasets. The proposed method gives better precision for various clustering samples when compared to other methods is also graphically represented in Fig. 3.

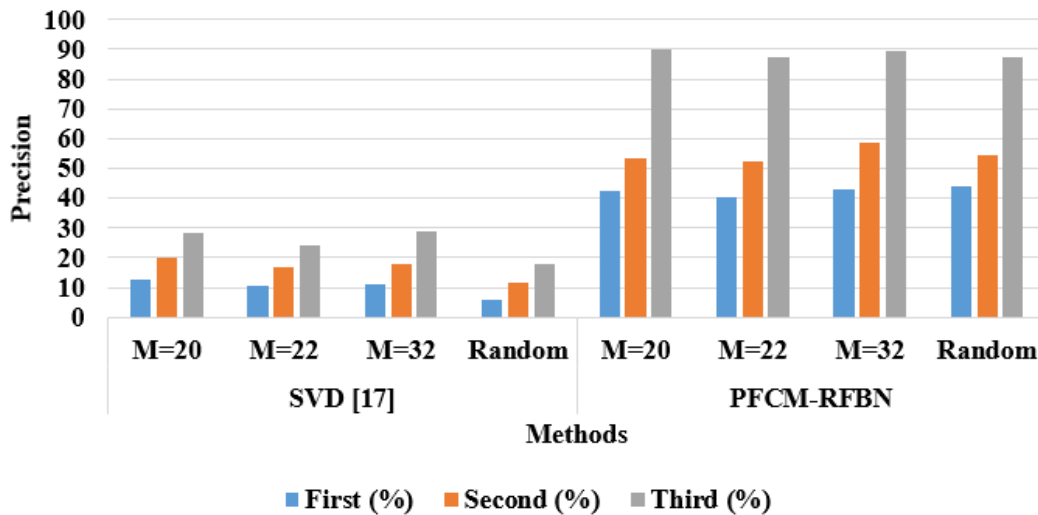


Figure.3 Precision comparison for MSNBC dataset

## 5. Conclusion

Web recommender systems are utilized to judge the web users' behavior depending on their usage patterns because of their popularity in e-commerce applications. A user's behavior plays a major role for e-marketers to customize its offerings to a web customer. In this paper, a method for predicting a user's next page visit is proposed in addition the method can able to give useful information about their tastes and likes. This information can be explored and utilized for projecting the desired category or product to the user. In this way, the probability of purchase for e-commerce organizations can be enhanced, which in turn will increase the expected revenue for e-commerce organizations. The proposed method used the PFCM with S3I SM and RFBN for developing a novel recommendation system, moreover the model considers the sequential information for predicting a user's next visit. The results of the proposed model are compared with the SVD for the prediction of accuracy and precision of the value. The experiments are conducted on the MNSBC dataset and the outcome of the experimental results showed the feasibility of the proposed work. The proposed method obtained nearly 83% accuracy and the precision value for MSNBC dataset acquired nearly 89.5%. The experimental results stated that the proposed method can predict the user's next page visit accurately by using his sequential visits. In future work, the process of updating the recommendation using relevance feedback will be improved with different techniques and provide better efficiency of the proposed method.

## References

- [1] S. Ahila, Sobitha, and K.L. Shunmuganathan, "Role of agent Technology in Web Usage Mining: homomorphic encryption based recommendation for E-commerce applications", *Wireless Personal Communications*, Vol.87, No.2, pp.499-512, 2016.
- [2] R. Ali, M. Afzal, M. Hussain, M. Ali, M.H. Siddiqi, S. Lee, and B.H. Kang, "Multimodal hybrid reasoning methodology for personalized wellbeing services", *Computers in biology and medicine*, Vol.69, pp.10-28, 2016.
- [3] Abdul-Aziz and R. Al-Azmi, "Data, Text and Web Mining for Business Intelligence: A Survey", *International Journal of Data Mining & Knowledge Management Process*, Vol.3, No.2, pp.1-21, 2013.
- [4] V. Krishnaraju, S.K. Mathew, and V. Sugumaran, "Web personalization for user acceptance of technology: An empirical investigation of E-government services", *Information Systems Frontiers*, Vol.18, No.3, pp.579-595, 2016.
- [5] V. Subramaniaswamy and R. Logesh, "Adaptive KNN based Recommender System through Mining of User Preferences", *Wireless Personal Communications*, Vol.97, No.2, pp.2229-2247, 2017.
- [6] M. Li, G. Sagl, L. Mburu, and H. Fan, "A contextualized and personalized model to predict user interest using location-based social networks", *Computers, Environment and Urban Systems*, Vol.58, pp.97-106, 2016.
- [7] R. Mahajan, J.S. Sodhi, and V. Mahajan, "Usage patterns discovery from a web log in an Indian e-learning site: A case study", *Education and*



- Information Technologies*, Vol.21, No.1, pp.123-148, 2016.
- [8] H. Li, Z. Xu, T. Li, G. Sun, and K.K.R. Choo, “An optimized approach for massive web page classification using entity similarity based on semantic network”, *Future Generation Computer Systems*, Vol. 76, pp.510-518, 2017.
- [9] A.I. Saleh, M.F. Al Rahmawy, and A.E. Abulwafa, “A semantic based Web page classification strategy using multi-layered domain ontology”, *World Wide Web*, Vol.20, No.5, pp.939-993, 2017.
- [10] K. Makwana, J. Patel, and P. Shah, “An Ontology Based Recommender System to Mitigate the Cold Start Problem in Personalized Web Search”, In: *Proc. of International Conf. On Information and Communication Technology for Intelligent Systems*, Cham, pp.120-127, 2017.
- [11] R. Katarya, and O.P. Verma, “An effective web page recommender system with fuzzy c-mean clustering”, *Multimedia Tools and Applications*, pp.1-16, 2016.
- [12] X. Xie, and B. Wang, “Web page recommendation via twofold clustering: considering user behavior and topic relation”, *Neural Computing and Applications*, pp.1-9, 2016.
- [13] J. Liu, C. Fang, and N. Ansari, “Request dependency graph: A model for web usage mining in large-scale web of things”, *IEEE Internet of Things Journal*, Vol.3, No.4, pp.598-608, 2016.
- [14] C. Suruchi, “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, 2016.
- [15] R. Geetharamani, P. Revathy, and S.G. Jacob, “Prediction of user’s webpage access behavior using association rule mining”, *Sadhana*, Vol.40, No.8, pp.2353-2365, 2015.
- [16] S. Sahoo, and B.K. Ratha, “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, pp.237-245, 2017.
- [17] P.K. Mishra, and B. Bhasker, “A web recommendation system considering sequential information”, *Decision Support Systems*, Vol.75, pp.1-10, 2015.