www.ijsrnsc.org

Available online at www.ijsrnsc.org

IJSRNSC

Volume-9, Issue-3, June 2021
Research Paper

Int. J. Sc. Res. in Network Security and Communication

E-ISSN:2321-3256

Providing a Web Recommender System Using Markov Chains and Registration Files Structure

M. Mojarad[1*], M.A. Mohammadshahi[2], M. Saberi Nasab[2], A. Shamsi[2]

[1]Dept. of Computer Engineering, Firoozabad Branch, Islamic Azad University, Firoozabad, Iran
[2]Dept. of Computer Engineering, Liyan Institute of Education, Bushehr, Iran

[*]*Corresponding Author: m.mojarad@iauf.ac.ir, Tel.: +98-91788-63397*

Received: 30/Mar/2021, Accepted: 20/Jun/2021, Published: 30/Jun/2021

***Abstract*—** Today, due to the increasing growth of the internet and the huge amount of information we need systems to be able to recommend the most appropriate services and products to the user. The systems that do this are recommender systems. These systems are intelligently using artificial intelligence techniques to identify the interests of your users on the internet and suggest tailored offers to the user's preferences and interests. Today, Markov models commonly used to predict web pages. For this purpose, in this research, we use a new Markov model and use the structure of registration files to predict the next pages obtained by the user. The proposed Markov model is based on a matrix of 1 to k and in the form of a Markov model which predicts the next pages. In order to reduce the complexity of the search space, as well as to better navigate to the recommender system, we use k-means clustering to group users. The results of the evaluations of the proposed method on the NASA web server log file and in the F-Measure criterion shows 0.57% superiority over BCF.

***Keywords*—**Markov chains, File structure, Recommendation system, K-means clustering.

I. INTRODUCTION

The amazing advancement of computer technology and the humanization of this tool has led to tremendous advances in the acquisition and storage of digital data as well as the emergence of large databases have been created in various fields. Trade, Agricultural, Traffic, Internet, Astronomical Data, Phone Details, Medical and clinic data are examples of such databases. In fact, database production and collection techniques have grown much faster than our ability to understand and use them [1].

From the late 1980s onwards, human beings began to think about access to the information contained in this vast amount of data, and began efforts to do so that were not possible with traditional database systems. The intensity of competition in the scientific, social, economic, political and military spheres has also doubled the importance of speed or access to information. Therefore, the need to design systems that are able to quickly explore users' favorite information with an emphasis on minimal human intervention on the one hand and turn to analytical methods appropriate to the volume of large data on the other hand was felt.

Web mining is the process of searching in the vast web world in order to find specific information and data [2, 3]. Branch web mining is data mining and data mining in the term knowledge is the discovery of a database. Web mining operations to search for the desired data must be as follows: Make sure the best results are provided to the user in the shortest possible time. With the rapid growth of information in the web world, there is a need to improve fast and secure methods web mining operations are more common. Due to the abundance of information available in the web world, users always expect to see the best answers they want at the beginning of the results. In many cases, the web pages found are against the user's wishes.

Recommending what the audience will welcome is the responsibility of the recommending systems. These systems, many of which we use today, try to be based on our interests and those of others, such as what books we have read and others that have had similar interests to ours, such as what books we have read, giving us a good offer. The advisory system or suggestion system, by analyzing the user's behavior, proposes the most appropriate items (data, information, goods and etc.). This system is an approach that is presented to deal with the problems caused by the large and growing volume of information and helps its user to get closer to their goal faster among the huge volume of information. Some consider the bidder system to be equivalent to collaborative filtering [4].

In the current research using Markov chains of all orders and user registration files, a highly accurate recommendation system is provided. Identifying effective features will be important to create a registration file. The filing of the log file and the identification of the sessions are examined in the current research.

In the continuation of this research, we will examine some of the works done in section II. In Section III, Markov's chain-based recommendation system presented. The results of the evaluation of the proposed method and its discussion are given in Section IV, and finally, Section V conclusion and future work is given.

II. RELATED WORKS

The earliest algorithm presented in this field is the shortest path algorithm proposed by Sisodia et al. [5]. To solve the problem of the shortest path in social network graphs, in which the weights of the edges are independent and predetermined values, there are different algorithms with polynomial time such as Digestra, Floyd and Warshal, such as the work done by Lokeshkumar and Sengottuvelan [6]. But these algorithms are dynamic if the weight of the edges change, they are not able to find the optimal solution because with the slightest change in the graph, the whole graph must be re-examined.

The process of finding the shortest route was then upgraded by Joshi and Kaur [7]. This algorithm can be used to find the shortest path between pairs graph nodes are presented in a random graph. The problem of the shortest possible path with the associated edges is first explored by George [8]. The problem of finding the shortest path of communication link in random graphs with correlated edges is that each edge may be in one of two modes without congestion and at the same time the cost distribution functions of the edges are already known, by Asghari and Navimipour, was raised [9]. In this article for the first time an algorithm for solving the problem of the shortest path of random graphs in situations where there is a correlation between the cost of the edges and also the possible weight distributions sides that are not already known are recommended.

After using the shortest path algorithms, linear programming algorithms were introduced. Navimipour and Milani research has shown that there is no optimal or efficient algorithm to solve the correct linear programming in the field of large graphs [10]. In fact, linear programming is a difficult NP-problem. Charband and Navimipour [11] is that we should not expect to be able to solve a problem in general. To do this, it is necessary to replace a correct linear programming with another optimization problem that can be solved so that the answer to the second problem is approximate to the answer to the first problem [12]. The process of designing and computational algorithms in the field of graphs and with the use of social networks continued until the separation algorithms with maximum click presented by Asghari and Navimipour [13]. These algorithms are designed based on graph subdivision problems and have different approximations according to each sub-problem, parameters and boundary. Tavakoli et al. has shown that these algorithms are based on solving graph problems such as Max-K-Cut, Max-K-Uncut, and so on.

Eirinaki et al. [14] presented an article entitled "Recommendation Systems for Large Social Networks: Examining Challenges and Solutions". In this article, it is stated that social networks are very important for networking, communication and content sharing. Sperli et al. [15], research into the proposed system to improve the approach to social media. In this article, a suggestion system for the program large data designed to provide useful advice is used on online social networks. Parvazeh et al. [16] using a topological combination, introduced the similarity of graph structure and user profile information about Friend Recommendation Systems (FRS) with similar interests, features, and applications. The proposed system consists of two stages. In the first step, all users according to the structural measurement and profile similarity and with the K-Medoids method is clustered. In the second step, the FriendLink algorithm is applied to each of the clusters, and the similarity of each user is calculated with all non-adjacent users. Finally, each user is given a number of users as a friend, given the highest similarity score.

Mahara [17] proposed a new similarity measurement based on the average difference measurement for filtering a partnership in a weak range. Conventional similarities, such as the cousin, Pearson's correlation coefficient, and Jaccard-like resemblance, are less accurate, so a new similarity approach is used in this study. Based on the average measurement, it has been suggested that it takes into account the habits of the users. Wang et al. [18] proposed a new recommendation strategy using expanded neighbor collaborative filtering. In this model, the strategy of the second neighbor's proposal is considered, which is expected to play a role in covering and diversifying the recommendations. Panchal et al. [19] provided a combined method for predicting user web access based on the Markov model. In this model, three clustering methods, Markov model and association rules were used in order to obtain better predictive accuracy, with the difference that first the discovery of the pattern of data by Apriori algorithm with association rules was applied to web sessions and then clustering is done as follows.

Krishnamoorthy et al. [20] presented a method for determining user behavior on the web using the Markov model. An analysis of the Markov model and the Markov model has been conducted at all levels, and a modified Markov model has been proposed to assess the scalability of the number of routes. No'aman et al. [21] presented a hybrid proposal model for web navigation. This hybrid model includes the Markov model and association rules. Markov model two is used to predict results on data. Khalil et al. [22], an integrated model for predicting page access next suggested on the web. They presented a hybrid model, including three models of clustering, Markov, and association rules.

III. THE PROPOSED METHOD

In the last decade, suggestive systems have gained widespread acceptance in research and business forums and are used as a means of intelligent search among the vast amount of information available. One of the most common techniques of recommender systems is to generate suggestions based on user-like resemblance.

In recommender systems, the input data includes a list of m users, $U = \{u_1, u_2, \dots, u_m\}$ and n items (plane), $P = \{p_1, p_2, \dots, p_n\}$. Each row represents items that have been ranked by a particular user, and these rankings are in accordance with the user profile in the registry files, and each column is ranked according to the rankings received by a specific item by all m users. The system section recommends a suggestion for predicting items in the future; we will be more likely to be visited by users. The items offered to each user are according to the behavioral similarities and the user's interest in that item. In this study, we use web pages as suggested items.

In this study, we use data from NASA's web server log file to provide a recommendation system. Pre-processing is

required for NASA web server registration files due to their features. The proposed method of this research is a combination of web application and web structure in the web server log file and uses k-means clustering algorithms and Markov chain. Here is an overview of the suggested recommender system. The suggested recommender system is a combination and consists of 6 steps according to Fig. 1. In the first step, the user log file is pre-processed. This step is done in order to format and access user details more easily by deleting invalid information. Pre-processing a log file includes steps to clear data, identify users, identify user sessions, and remove remote pages.

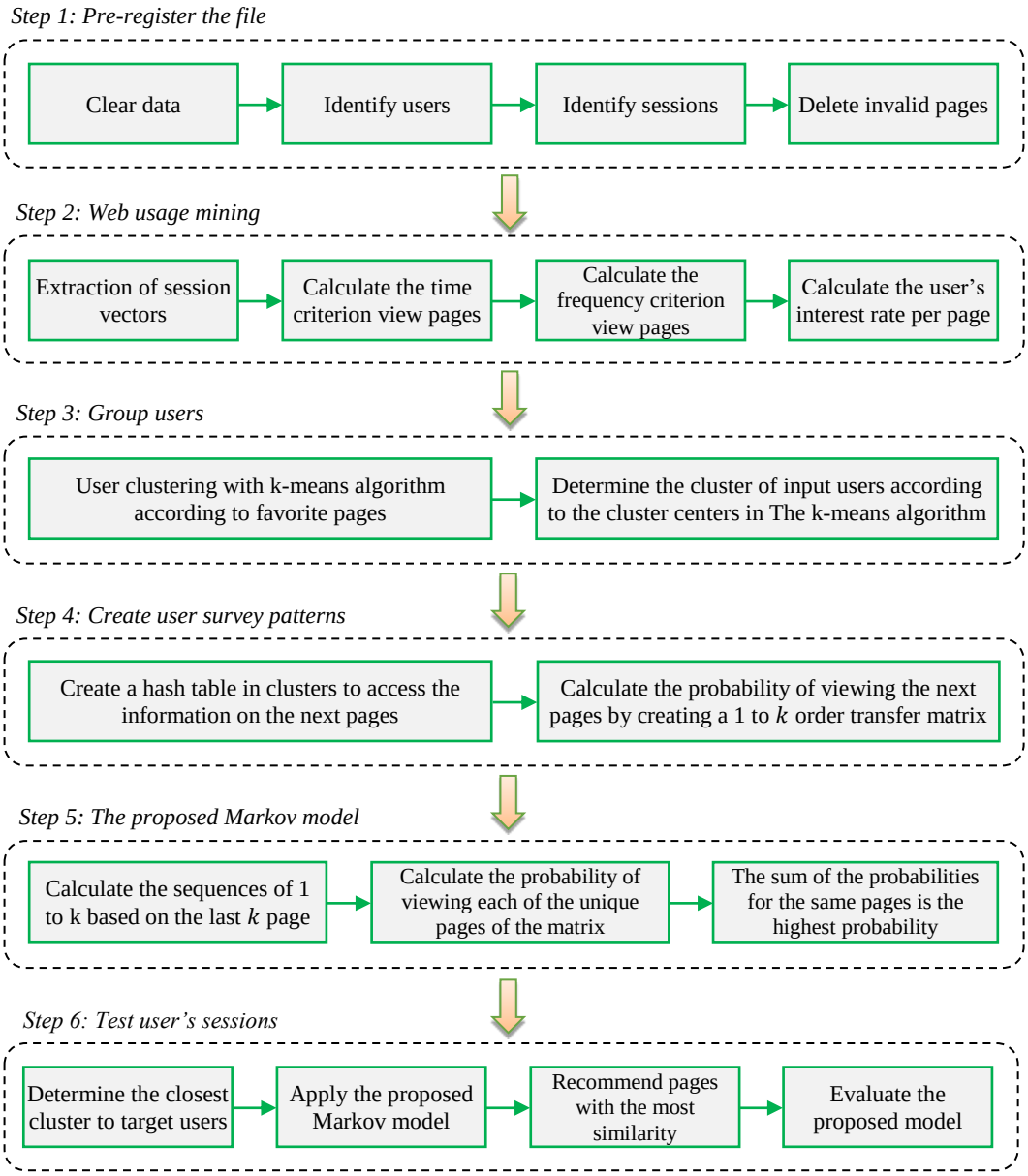

Figure 1. Diagram block of the proposed method

Web application is applied in the second step. At this stage, according to the two criteria, the viewing time of the page and the page frequency for each user of the session vector are extracted. In this step, by assigning weight to each page, we determine the user's interest in the pages. This weight according to the average the harmonics of the screen viewing time and screen frequency are calculated. In the third step, in order to reduce the complexity of the search space and also to better orient the proposed recommender system, we use k-means clustering for grouping users according to the favorite pages. After the application clustering and cluster recognition, we create a model of the closest cluster to determine the cluster of new input users in the test phase.

In the fourth step, in order to use the Markov chain to suggest pages to users, user preview patterns are created in the form of hash tables. This is to access the next pages, the frequency of viewing the next pages and the duration of viewing the next pages for each sequence of pages one.

It is a cluster. In this step, we also use the probability matrix based on the hash table to calculate the probability of viewing the next pages. In this study, we use the 1 to k order transfer matrix to achieve the goals and increase the

accuracy of the suggestions. Finally, in step five, a new Markov model is based Matrix level 1 to k is provided in the form of Markov model of all levels. This model is used to recommend new pages to users. Review of meetings Test users are done in step six. At this point, considering the centers of the corners in the k-means algorithm, the closest cluster to input users Specified, and then based on the probability matrix of order 1 to k and the Markov model of all the proposed levels, the pages of the highest similarity are recommended.

A. Pre-process file registration

The process of analyzing this research is based on the NASA dataset. Therefore, this section is based on the NASA log file. Fig. 2 shows some of the records in this dataset.

```
burger.letters.com - - [01/Jul/1995:00:00:12 -0400] "GET /images/NASA-logosmall.gif HTTP/1.0" 304 0
burger.letters.com - - [01/Jul/1995:00:00:12 -0400] "GET /shuttle/countdown/video/livevideo.gif HTTP/1.0" 200 0
205.212.115.106 - - [01/Jul/1995:00:00:12 -0400] "GET /shuttle/countdown/countdown.html HTTP/1.0" 200 3985
d104.aa.net - - [01/Jul/1995:00:00:13 -0400] "GET /shuttle/countdown/ HTTP/1.0" 200 3985
129.94.144.152 - - [01/Jul/1995:00:00:13 -0400] "GET / HTTP/1.0" 200 7074
```

Figure 2. Part of the records of the NASA log file

Each row shows a record of a user's activity, which consists of 9 different sections. Many of these sections are not required in the application of web mining as well as the proposed research method. So we clear this data by deleting it. Details of the sections from left to right in order of users' IP address, user ID, access date, access time, HTTP request method, source route in web server, protocol used for transfer, status code and number of bytes transmitted.

Our goal is to process, record, and format files for easier access to stored information. Data cleansing is a way to do the pre-processing process in web application applications. This clears other invalid records from the log file. In this study, the following items are examined to determine the validity of each record.

- According to the research approach in examining users' interest in web pages, only addresses related to files with HTML and HTM extensions are considered.
- Records that have used the HEAD application method in HTTP are invalid.
- Records with status codes 301, 403, 404, 500 are invalid.
- Records for creating a new session are invalid. These records are characterized by a lack of a web address.
- Records sent by the robot are also invalid. These requests are often sent by sites such as Google and Yahoo.

In this study, users' IP addresses are used in each record to identify users. This field is displayed in IP and Host form in the log file. Here we create a unique list of users' IP addresses. This data includes m users as $U = \{u_1, u_2, \dots, u_m\}$. Scheduling sessions of new users are based on their IP addresses. For each user u_{i}, a session contains a list of pages that they have viewed in a row. So the list of user sessions u_{i} can be displayed with S_i, as $S_i = \{\langle p_1, p_2, \dots\rangle, \langle p_1, p_2, \dots\rangle, \dots\}$. Each section $\langle p_1, p_2, \dots\rangle$ of S_i is a new session for the user u_i. p_j Shows pages visited by a user in a session. To determine the sessions, we use the two fields, time and history. From time difference (per minute) between the first time you see the IP of a user in the log file to the last less than 30 minutes, pages observed at this time by the user in a new session. Otherwise, we use the 10-minute rule. This law is based on time stay on the screen for 10 minutes (depending on a threshold).

B. Extract features

User sessions can be expressed as vector weights. The weight value for each page indicates how much users are interested in that page. In this study, the weight of each page is calculated according to the two criteria of page frequency and duration of page view and use of harmonic averages. To determine the level of interest in user sessions, we consider $n \times m$ matrix as Eq. (1).

$$interest = \begin{bmatrix} w_{1,1} & \dots & w_{1,m} \\ w_{2,1} & \vdots & w_{2,m} \\ \dots & & \dots \\ w_{n,1} & \dots & w_{n,m} \end{bmatrix}_{n\times m} \qquad (1)$$

Where, w Shows the weights and $w_{i,j}$ indicates the level of interest of the user i^{th} to page j^{th}. Due to the difference between the size of the data in the two criteria D and F, we put the values of these two criteria in the same range as the MAX-MIN normalization method. The total importance of a page (The weight of each page) calculated from the combination of two criteria F and D and with the average harmonic operator. The following equation shows the amount of interest in each page. $interest_{i,j}$ shows the user's interest in this page.

C. Apply clustering on users

After calculating the users' interest in the pages, the k-means clustering algorithm is used to place similar users in the same groups. In the k-means algorithm, Euclidean distance similarity criteria are used to calculate similar users to obtain significant clusters.

Specifically, the reason for user clustering is that the similarity of the target user is calculated only with the users of a cluster to suggest pages, and that cluster will have the highest similarity with the target user. In the test phase, for each input user, all training users should be checked for page recommendations. In this study, to reduce complexity, only one group of clusters (from users who have the most similarities to the target user) are identified and examined to recommend the pages.

D. Markov chain on clusters

In this paper, we use the Markov model of predicting the next pages for users. Markov's model is often used to identify patterns and behaviors of users based on a sequence of previously viewed pages. The idea of this model is to use a suitable data structure such as hash table, through which it can check the probability of each pattern. Using the hash table has a fixed execution time to predict, because the access time of each input in the hash table is indexed and is assumed to be constant. When it comes to predicting the next pages for users, the input data for a Markov model are web sessions that each session follows.

When the Markov model ranks are identified (from 1 to k), the transfer probability matrix is created. In the matrix, the probability of transfer of each input $t_{i,j}$ indicates the number of times the activity a_i mode s_j has been observed. For example, in session $\langle p_3, p_5, p_2, p_1, p_4 \rangle$ and the Markov 1st order model, if we assume that the p_3 page corresponds to the s_3 mode, then the p_5 page follows the p_3 page. For the Markov model, order $\{p_3, p_5\}$ follows the p_2 sequence.

In this study, we extract users' click streams using the hash table. With this technique, information can be accessed to the next pages for each sequence of pages within each cluster. This information includes the name of the next page, the number of observations and the duration of the observation. Therefore, for each field in the hash table, three factors are defined: the name of the next page, the number of observations, and the duration of the observation. The hash table is calculated for levels 1 to k.

To calculate the probability of viewing the next pages, we use the transfer matrix and the amount of interest calculated from the harmonic relationship. For each level of the Markov model corresponding to the hash table, a probability transfer matrix is created.

If $interest_{i,j}$ is the user's interest rate i^{th} user to j^{th} page, $p_{i,j}$ is considered as the probability of viewing the page (probability of transfer) and is calculated as Eq. (2).

$$interest = \begin{bmatrix} w_{1,1} & \dots & w_{1,m} \\ w_{2,1} & \vdots & w_{2,m} \\ \dots & & \dots \\ w_{n,1} & \dots & w_{n,m} \end{bmatrix}_{n \times m} \quad (2)$$

The value of $p_{i,j}$ is considered normalized from the level of interest. In many cases, low-order (first or second) Markov models are not able to accurately predict the next page seen by the user, and this is because these models do not take a deep look at the user's past and only based on observation. They predict the last one or two pages, and as a result, higher-order (three or four) Markov models should be used to get better accuracy. Unfortunately, higher-end Markov models also have limitations, such as high number of modes, low coverage, and sometimes even lower accuracy due to low coverage. One way to solve the problem of low-learning coverage is to use different types of Markov models and then combine them to predict. In this approach, for each sample, the largest Markov model that covers the sample is used for prediction. For example, if this model consists of three levels of the Markov model, the first given example is predicted to be three if possible, and if it is not covered by three orders, this process is repeated for second and first order. This method is called All k-order Markov model. Of course, it should be noted that although this procedure solves the problem of low coverage, it worsens the problem of complexity, because cases are related to all parts of model.

In this study, we use a Markov model at all orders in a new structure to predict pages. First, the matrix of order 1 to k is created. Then, with the arrival of a new session, all sequences 1 to k are calculated based on the last page k. In the next step, the possibility of viewing each of the different pages after the current session sequences is extracted from the transfer matrix. Finally, the probability of the same pages for all sequences is added together. This probability is calculated for all the distinct plates that have a non-zero final probability. After that, the most likely pages (pages that are likely to be viewed after the current session) are recommended to the target user. For example, for a session $\langle s_1, s_2, s_3, s_4, s_5 \rangle$, if $k = 3$, we have to calculate sequences 1, 2, and 3 from the session pages based on the last 3 pages. We then extract the probabilities of viewing the different pages from the transfer probability matrix. Therefore, the probability of viewing distinct pages after sequence 1 of the sequence matrix 1, the probability of viewing distinct pages after sequence 2 of the sequence matrix 2, and the probability of viewing distinct pages after sequence 3 of the sequence matrix 3, and the probability of similar pages for all sequences gather.

$$\begin{aligned} &p(a_{n+1}|a_n a_{n-1} \dots a_{n-k}) \\ &= p(a_{n+1}|a_n) + p(a_{n+1}|a_n a_{n-1}) \\ &+ \dots + p(a_{n+1}|a_n a_{n-1} \dots a_{n-k}) \end{aligned} \quad (3)$$

In the testing phase, to recommend pages to incoming users, their sessions are first extracted. Then, using the nearest cluster center criterion; the corresponding cluster is identified with the target user and the closest users are assigned to the input user. The closest cluster users show the highest similarity to the input user. Then, based on the proposed Markov model, the probability of each page for recommendation is calculated and the pages with the highest probability for recommending to the input user are identified. Finally, in order to check the efficiency of the proposed model and the accuracy of the proposed pages, criteria such as Precision, Recall, F-measure are used.

IV. RESULTS AND DISCUSSION

Extensive experiments have been performed in this section to investigate the superiority of the recommended recommender system. Web access files, simple text files contain valuable information about the events and behaviors of users on the web server in chronological order. In this paper, we use the NASA web server log file to simulate and evaluate the proposed method. The information in this address http://ita.ee.lbl.gov/html/contrib log file can be found at Aug 1995. In this research, MATLAB software version 2019a has been used to

simulate and analyze the proposed recommender system on the NASA data set. Simulation and all tests were performed using a 2.4 GHz Intel Cor i7 CPU and 8 GB of RAM.

In this study, we use k-fold cross-validation to report results according to different evaluation criteria. The NASA record file used in the experiments is divided into two parts, instructional (E^T) and experimental (E^P), at each stage of 10-fold validation. 90% of the records are used for E^T and the remaining 10% for E^P. In this study, we use Precision, Recall, and F-Measure criteria to evaluate performance and compare the results of the proposed recommender system. We use the BCF [23] method to evaluate the performance of the proposed recommender system. BCF (Bhattacharyya Collaborative Filtering) method by Patra et al. [23] was presented. In this method, the Collaborative Filtering (CF) criterion based on the neighborhood, uses all rankings among users, this method uses the Bhattacharyya similarity criterion to calculate the scoring of items.

In our first experiment, we calculated the Precision criterion for the number of pages proposed from 1 to 10 (top-k). The results of the Precision criterion are shown in Fig. 3 for the proposed method and the BCF method. The proposed method is higher than most other methods in terms of Precision criteria. In the best case, the proposed method has reached the Precision criterion of 98.21% with 1 suggested page. The results show that by increasing the number of suggested pages, the standard value Precision is declining in different ways. The reason for this is the method of calculating the Precision criterion, which shows the ratio of the number of correctly predicted pages to the number of suggested pages.

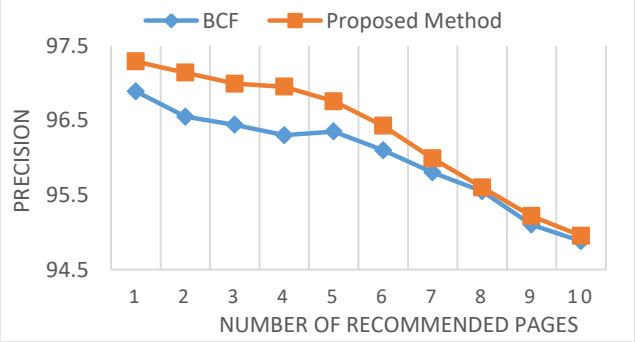

Figure 3. Precision criteria results with a number of different suggestions

In the next experiment, we calculate the Recall criterion for the proposed method and the BCF method. Fig. 4 shows the results of the Recall criterion.

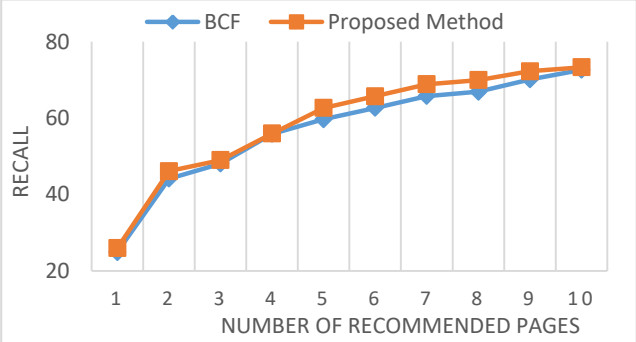

Figure 4. Recall criteria results with a number of different suggestions

The results obtained from the Recall criterion show that in most cases the proposed method performs better than the two methods compared. At best, the proposed method has reached the Recall criterion of 73.34% with 10 suggested pages. The results show that increasing the number of suggested pages increases the value of the Recall criterion in different ways. The reason for this is the method of calculating the Recall criterion, which shows the ratio of the number of correct pages suggested to the number of actual pages viewed by the user.

The Precision criterion value decreases with increasing number of suggested pages, and in the Recall criterion this value decreases. Therefore, it is necessary to use a combination of two criteria, Precision and Recall, for a more detailed study. Here we use the F-Measure criterion to evaluate the methods, which is calculated from the harmonic mean of the two Precision and Recall criteria. Fig. 5 shows the F-Measure criteria results.

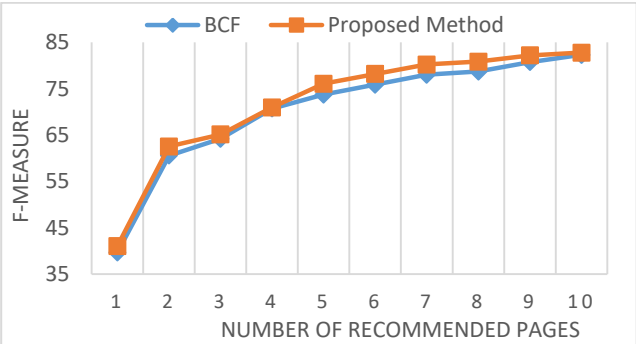

Figure 5. F-Measure criteria results with a number of different suggestions

The F-Measure criteria results show that the proposed method performs better than other methods. At best, the proposed method has reached an accuracy of 82.76% when the $topk = 10$. In general, the results of experiments show the high accuracy of the proposed method compared to other similar methods. For a more detailed examination, Table I shows the numerical results of the various criteria with the proposed 10 pages (due to the superiority of the total of the various experiments). The results of the evaluations show an improvement in the proposed method in the F-Measure criterion compared to the BCF method of 0.57% in the Aug file registration data.

TABLE I. EVALUATION RESULTS OF THE PROPOSED METHOD AND COMPARISON WITH BCF METHOD

Methods	Precision	Recall	F-Measure
BCF	94.88	72.66	82.29
Proposed Method	94.95	73.34	82.76

V. CONCLUSIONS AND SUGGESTIONS

In the last decade, suggestive systems have gained widespread acceptance in research and business forums and are used as a tool for intelligent search among the vast amount of information available. One of the most common techniques of recommender systems is to generate suggestions based on user-like resemblance. The proposed method of this research is a combination of web application and web structure in the web server log file and uses k-means clustering algorithms and Markov chain. In general, the results of experiments show the high accuracy of the proposed method compared to other similar methods. The results of the evaluations show an improvement in the proposed method in the F-Measure criterion compared to the BCF method of 0.57% in the Aug 2.66% recorded file data.

In the course of this research, the proposed systems can be improved in such a way that in addition to offering the user's favorite resources, similar friends can also be offered to him or her. These systems can also be improved by adding user ratings.

REFERENCES

[1] Lu, J., Wu, D., Mao, M., Wang, W., & Zhang, G. (2015). Recommender system application developments: a survey. *Decision Support Systems*, 74, 12-32.

[2] Ahmadian, S., Joorabloo, N., Jalili, M., Ren, Y., Meghdadi, M., & Afsharchi, M. (2020). A social recommender system based on reliable implicit relationships. *Knowledge-Based Systems*, *192*, 105371.

[3] Rezaeipanah, A., Mojarad, M., & Fakhari, A. (2020). Providing a new approach to increase fault tolerance in cloud computing using fuzzy logic. *International Journal of Computers and Applications*, 1-9.

[4] Rezaeipanah, A., Ahmadi, G., & Matoori, S. S. (2020). A classification approach to link prediction in multiplex online ego-social networks. *Social Network Analysis and Mining*, *10*(1), 1-16.

[5] Sisodia, D., Singh, L., Sisodia, S., & Saxena, K. (2012). Clustering techniques: a brief survey of different clustering algorithms. *International Journal of Latest Trends in Engineering and Technology*, 1(3), 82-87.

[6] Lokeshkumar, R., & Sengottuvelan, P. (2014). A Novel Approach to Improve Users Search Goal in Web Usage Mining. *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, 9(2), 624-628.

[7] Joshi, A., & Kaur, R. (2013). A review: Comparative study of various clustering techniques in data mining. *International Journal of Advanced Research in Computer Science and Software Engineering*, 3(3), 55-57.

[8] George, A. (2013). Efficient high dimension data clustering using constraint-partitioning k-means algorithm. *International Arab Journal of Information Technology*, 10(5), 467-476.

[9] Asghari, S., & Navimipour, N. J. (2016). Service composition mechanisms in the multi-cloud environments: a survey. *International Journal* of *New Computer* Architectures and Their Applications, *6*, 40-48.

[10] Navimipour, N. J., & Milani, F. S. (2015). A comprehensive study of the resource discovery techniques in peer-to-peer networks. *Peer-to-Peer Networking and Applications*, 8(3), 474-492.

[11] Charband, Y., & Navimipour, N. J. (2016). Online knowledge sharing mechanisms: a systematic review of the state of the art literature and recommendations for future research. *Information Systems Frontiers*, 18(6), 1131-1151.

[12] Navimipour, N. J., & Asghari, S. (2017). Energy-Aware Service Composition Mechanism in Grid Computing Using an Ant Colony Optimization Algorithm, *ICEIC 2017 International Conference on Electronics, Information, and* Communication, 282-286.

[13] Asghari, S., & Navimipour, N. J. (2016). Review and comparison of meta-heuristic algorithms for service composition in cloud computing. *Majlesi Journal of Multimedia Processing*, 4(4), 28-34.

[14] Eirinaki, M., Gao, J., Varlamis, I., & Tserpes, K. (2018). Recommender Systems for Large-Scale Social Networks: A review of challenges and solutions. *Future Generation Computer Systems*, 78(1), 413-418.

[15] Sperli, G., Amato, F., Mercorio, F., Mezzanzanica, M., Moscato, V., & Picariello, A. (2018). A Social Media Recommender System. *International Journal of Multimedia Data Engineering and Management*, 9(1), 36-50.

[16] Parvazeh, F., Harounabadi, A., & Naizari, M. A. (2016). A Recommender System for Making Friendship in Social Networks Using Graph Theory and users profile. *Journal of Current Research in Science*, (1), 535.

[17] Mahara, T. (2016). A new similarity measure based on mean measure of divergence for collaborative filtering in sparse environment. *Procedia Computer Science*, 89, 450-456.

[18] Wang, B., Gao, Q., Feng, X., & Pan, F. (2017). Recommendation strategy using expanded neighbor collaborative filtering. In *Control Conference (CCC), 2017 36th Chinese* (pp. 1451-1455). IEEE.

[19] Panchal, P. S., & Agravat, U. D. (2013, July). Hybrid technique for user's web page access prediction based on Markov model. In *Computing, Communications and Networking Technologies (ICCCNT), 2013 Fourth International Conference on* (pp. 1-8). IEEE.

[20] Krishnamoorthy, P., Chaki, S., & Verma, N. (2017). Method and apparatus for determining user browsing behavior, *U.S. Patent No. 9,661,088*. Washington, DC: U.S. Patent and Trademark Office.

[21] No'aman, M., Gadallah, A. M., & Hefny, H. A. (2015, December). A hybrid recommendation model for web navigation. In *Intelligent Computing and Information Systems (ICICIS), 2015 IEEE Seventh International Conference on* (pp. 552-560). IEEE.

[22] Khalil, F., Li, J., & Wang, H. (2009). An integrated model for next page access prediction. *International Journal of Knowledge and Web Intelligence*, 1(1-2), 48-80.

[23] Patra, B. K., Launonen, R., Ollikainen, V., & Nandi, S. (2016). A new similarity measure using Bhattacharyya coefficient for collaborative filtering in sparse data. *Knowledge-Based Systems*, 82, 163-177.

AUTHORS PROFILE

Mr. Mousa Mojarad received his PhD in Computer-Software Engineering in 2020. He is currently a lecturer and faculty member of the Islamic Azad University, Firoozabad Branch. His hobbies are big data, cognitive computing, clustering, software engineering, classification models, and cloud computing. He has more than 8 years of teaching experience and 6 years of research experience.

Mr. Mohammad Amin Mohammadshahi received the B.Sc. Degree in Fire and safety services from University of Applied Science and Technology, Bandarsazan Genaveh, in 2011, and the M.Sc. degree in computer software engineering from Liyan Institute of Education, Bushehr, in 2021. His primary research interest is in using meta-heuristic algorithms for routing, although he has concurrent research in robotics, machine learning, optimization algorithms, and artificial intelligence.

Mr. Morteza Saberi Nasab received the B.Sc. Degree in Fire and safety services from University of Applied Science and Technology, Bandarsazan Genaveh, in 2013, and the M.Sc. degree in computer software engineering from Liyan Institute of Education, Bushehr, in 2021. His current research interests include Evolutionary Computation, Optimization Methods, and Swarm Intelligence.

Mr. Abdolrahim Shamsi received the B.Sc. Degree in Fire and safety services from University of Applied Science and Technology, Bandarsazan Genaveh, in 2013, and the M.Sc. degree in computer software engineering from Liyan Institute of Education, Bushehr, in 2021. Abdolrahim's current research interests include the affective computing, virtual reality, human-computer interaction, and computer networks.

