Addressing the Cold-start Problem Using Data Mining Techniques and Improving recommendation By Cuckoo Algorithm: A Case Study of Facebook

Saman Forouzandeh† , Atae Rezaei Aghdam††, Shuxiang Xu††† and Elmira Akbari Pour Dibazar††††

† Department of Computer Engineering, University of Applied Science and Technology, Center of Tehran Municipality ICT org , Tehran, Iran.
†† Faculty of Information Technology , University of Applied Science and Technology, Center of Tehran Municipality ICT org , Tehran, Iran
††† School of engineering and ICT , Faculty of Science, Engineering and Technology , University of Tasmania , Tasmania, Australia.
††††Computer Science Department , Iran University of Science and Technology , Tehran, Iran.

Abstract
The popularity of Social networks, user demands, market realities, and technology developments are driving recommendation systems to explore new models of marketing and advertisements. Each user on social media sites shows a specific behavior such as; ranking interest items or like favorite contents. Hence, by studying user profiles we can explore their interests and preferences. Due to the great bulk of data on social media websites, the process of extracting hidden knowledge from data has become a hectic activity. For achieving this goal data mining techniques were flourished to discover interesting knowledge along with recommendation systems to suggest appropriate items to users based on this extracted knowledge. On of the most common obstacles in recommendation systems is “cold-start” problem, which is related to users who do not indicates any behavior on social medias. This paper aims to propose a solution for tackling this problem by using data mining techniques. Data were collected from facebook user accounts through Netvizz tools. In the next level, we enhance the recommendation method through Cuckoo algorithm to offer minimum number of items to get maximum feedback from users. Results indicate high performance of our propose solution.
Keywords:
Recommendatiuon systems, Data mining, Social networks, Cuckoo algorithm, cold-start

1. Introduction

Social network is a kind of service that focuses on online network, which includes different types of people who are interested in sharing their experiences and activities through the Internet [1]. Social network’s users have particular profiles and these profiles assist researchers to acquire more information about users [2]. Finding similarities among users profile is conducted based on various types of techniques, so the users profiles have become a significant source of information for researchers [3][4]. Social networks contain complementary information and independent data center to help individuals in forming new connections. This type of information plays an essential role in recommendation systems performance [5]. Moreover, Recommendations to users are based on analyzing their online behaviors and information. Due to the great bulk of information, the process of analyzing data has become a hectic and time-consuming activity. Therefore, data mining techniques can be utilized to discover the hidden knowledge from a bulk of data. In fact, Data mining refers to uncovering new, interesting and valuable patterns and knowledge from massive amount of data [6][7].

Extracted knowledge from data sets is used in recommendation systems to suggest goods or services to customers based on their preferences and tendency. Actually, recommendation systems try to find a thinking pattern of users, offering the best option to them [8][9]. On of the most significant challenges in recommendation systems is “cold-start” problem. This problem occurred due to insufficient information of recommender system about novel users or new items. In essence, recommender systems can not recognize users online behavior pattern and can not offer the best option to users as well thus, it frequently causes poor prediction outcomes for a recommender systems. [10][11][12][13]. In this study, we proposed a solution for this problem through clustering techniques and association rules. In the next level, for optimizing suggestions to users and increasing the accuracy of results, Cuckoo algorithm has been carried out. The main aim of this research study is to recommending minimum options with maximum acceptance by users. Our case study is facebook website which is one of the most popular online social networks in the world. It includes more than 100 millions daily users, each of them login to their profile at least one time during a day [14]. This paper is arranged in following sections; section 1; exploration of related works

Manuscript received August 5, 2026
Manuscript revised August 20, 2026
https://doi.org/10.22937/IJCSNS.2026.26.8.19

regarding cold-start problem; section 2; explanation of our research methodology and section 3; evaluation of results.

2. Literature Review

Social networks involve complementary information and independent databases. This type of information and relationship between them can boost the performance of recommendation systems [5]. The purpose of social media is to encourage users to share many attributes with the people close to them. In the other word, users incline to be touched by opinion and recommendation by their friend with similar interests rather than marketing promoters [15]. In this regard, one study analyzed online users behavior's on the facebook website to explore the advantages of using this websites, finding the factors that effect users behaviors. Due to some characteristics of facebook such as; indicating some aspects of individual life e.g. Location, occupation, email address, people tend to use this social network website more than others. These features' lead facebook to become one of the most popular social networks all across the globe [16]. The popularity of facebook dramatically increased in recent years and a plenty of research done in this area as EBSCO contend that there were 634 research papers published regarding facebook until May 2015[17]. In the research done by Petrocchi and Asnaani in 2015, they distributed a survey among 214 students at Boston University to compare facebook users who only used Facebook with the users used both Facebook and Twitter social networks. They analyzed factors like; age, gender, nationality, and number of posts or twits they posted in a particular period of time. The results of this study showed that there is no huge difference between these two groups of users however, users tend to join more than one social network [18]. Hence, social networks become more and more prevalent among people in the world. Increasing popularity of social network sites helps companies to promote their product and services through knowledge discovery and recommendation systems techniques. In addition, customers come across with numerous choices to select. Actually, choosing the best option for users based on their preferences has become a hectic debate. One study analyze some Twitter profiles and suggested a framework to automatically offer real time recommender system, which applied sentiment analysis to explore positive and negative feedback, and achieved data set of users with their preferred products without asking them to enter ratings and feedback's. Data were gathered from Twitter site through twitter API. Further, clustering techniques and association rules have been used to distinguish and recommend appropriate items to users [19]. Another study conducted on Facebook social site to propose a social recommender system with user-oriented approach focusing on relationship between social relations and user interest similarities through sentiment analysis. In the first phase, data were crawled from Facebook website pages, creating a favorite product list of users including cellphone brands and in the second phase, each user was presented regarding specifications of information such as; year of birth, gender, education level, hometown, interests, etc. The results argued that social features outweigh biography characters in expressing the social relations amongst users [20].

In a research done by Forouzandeh and Soltanpanah, they applied data mining techniques and association rules to propose a recommender system that suggest options to purchase to users based on their behavior on facebook website. Data collection performed by Netvizz software, exploring user interests through their likes in other posts then, by similarity of users profiles on the facebook and their interests to different products, they offered a second choice to users [21]. Data mining techniques assist recommendation systems to offer the best options to customers based on their tendency for instance, one research study implemented data mining techniques such as visualization, clustering and association rules to find the best places for tourists based on their favorites. Authors analyzed tourists profiles on tripadvisor website, which is one of the popular travel websites. Data were crawled from 600 users profiles through specific web crawler from tripadviser website then, data mining techniques and qualitative analysis by NVIVO software were carried out to discover the valuable and conclusive outcome. The results of this study used in recommender system to suggest tourist destination for travellers in Malaysia [22]. Another study conducted K-means and Cuckoo search optimization algorithms applied at Movielens dataset to present a novel recommendation system regarding movie suggestion. They used well-known Movielens dataset to analyze a data. Their hybrid model performance was more efficient than current one in terms of Mean Absolute Error (MAE), Standard Deviation (SD), and Root Mean Square Error (RMSE) metrics [23].

In terms of addressing "cold-start" problem, there are several research studies have been published. For example in the research study done by Braunhofer, he suggested a problem solving method by switching hybrid context-aware recommender systems algorithm for solving a poor recommendation to new users. He analyzed two data sets including; 2,422 contextually-tagged rating for places of interests and 2,296 contextually-tagged rating for movies. Results claimed that content-aware recommendation algorithm was more precise than the baseline model. In this model the information of user profiles have been used to recommend an option [24]. Another study offered a graph-based and matrix factorization recommendation model to tackle the cold-start problem regarding movie and music recommendation to new users. Data sets are achieved from structured version of Wikipedia called "DBpedia" and proposed model was evaluated with a facebook "like" dataset. Results of this study mentioned that suggested approach provided relevant recommendation even for users with very few likes. In addition, graph-based technique more efficiently elicited content information [25]. Moreover, Chou Szu-Yu and his colleagues conducted a research in order to addressing cold-start problem for next-song recommendation by using factorization-based algorithm to exploit content features from music's and sequential behavior to guess the next song. They studied on a large-scale music recommendation datasets and found that proposed algorithm achieved high performance that other content-based methods. Additionally, their method assists users to explore new options in next item recommendation [26]. In another research regarding cold start problem, authors presented three-stage solution in order to predict users behavior and interests. The suggested method uses demographic data for identifying other users with similar behavior. Next, each new user is classified in a specific category and consequently a rating prediction technique has been applied at items.

Finally, the experimental results presented high performance of offered system in various experimental scenarios [27]. For diminishing the cold start problem in Recommendation systems Pereira and Hruschka offered a collaborative filtering recommendation with demographic information. Users are classified into clusters based on their similar characteristics. The method is based on current algorithm SCOAL (Simultaneous Co-Clustering and Learning). Providing a hybrid recommendation methodology to tackle the cold start problem. Results indicated that the efficiency of proposed techniques to replace the lack of information for novel users [28]. Another solution for cold start problem presented in paper [29] by combination of association rules and clustering techniques. Movielens dataset has been chosen for data collection for this study. They analyzed a dataset through Weka software to carry out association and clustering approaches on data. Results showed that using association rules can be solved novel user problem and clustering techniques can be utilized to categorize elements based on similarities to predict the items in order to tackle new item problem [29].

3. Proposed Framework

Our suggested framework is depicted in figure 1. As can be seen, it consists of two phases. In the first phase, data were harvested from facebook users profiles and in the second phase, all algorithms were applied at collected dataset.

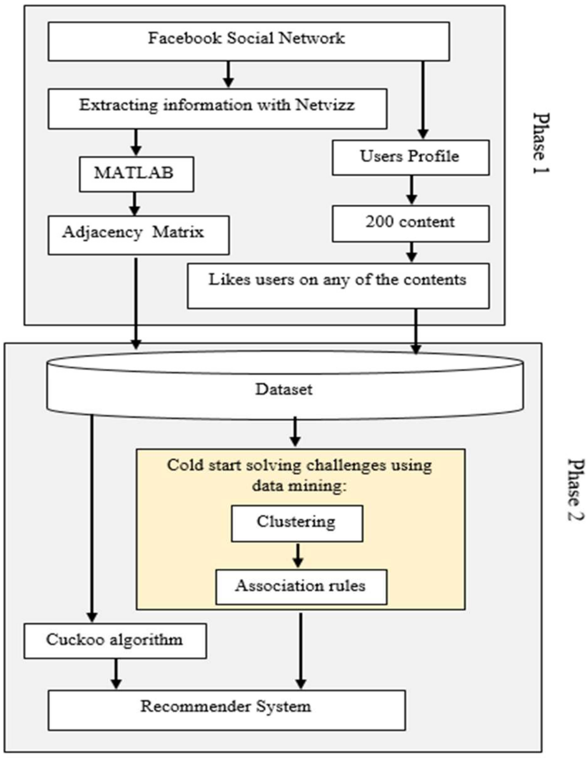

Figure 1 Proposed Framework i

3.1 Gathering data from facebook

For collecting data from users profile on facebook, we used Netvizz software to obtain friendship graph for each user. Moreover, Paper [14][21][30] used Netvizz application for harvesting data to analyze. Data collection includes two sections. Section one consists users profiles information and section two contains adjacency matrix made by MATLAB software to identify the friendship relations. Thus, the profile data and adjacency matrix were added to dataset. To analyze users behavior's we proposed 200 topics related to technology, sport, health, music and book, asking users to like each item if they interested in. These questionnaire were distributed through 5 accounts on facebook contains totally 3650 friends in which 2780 users fulfilled this questionnaire.

3.2 Solving cold-start problem and proposing recommendation system

In the second phase, according to our dataset, data analysis done in two stages. In the first stage, several data mining

Figure 1. Proposed framework

techniques were applied to solve the cold-start problem including: clustering, DBSCAN algorithm and association rules mining. In the second stage, Cuckoo algorithm has been utilized to boost recommendation system accuracy in terms of suggesting items to users.

4. Cuckoo optimization algorithm

Cuckoo algorithm is an optimization algorithm introduced by Xin-she Yang and Suash Deb in 2009 [31]. Like other evolutionary algorithms, Cuckoo Optimization Algorithm (COA) starts by initial population of Cuckoos. It was inspired by Cuckoos species [32]. Cuckoos put their eggs in the nest of other birds. Only the eggs that are highly similar to bird eggs have a chance to grow otherwise, they will throw a way by birds. [33][34][35]. The more eggs survived in the nest the more benefits we can get. Hence, the location in which more eggs lived will be the best point for us in terms of optimizes parameter [36][37][38]. For solving an optimization issue, it is required to convert variables to a form of array. This array known as “Habitat” in Cuckoo algorithm. In a N dimensional optimization issue, a habitat is an array of $1*N_{var}$ indication present living location of Cuckoos. This array defines as below[39]:

$$\text{Habitat} = [x1, x2, \ldots, x_{Nvar}] \quad (1)$$

The habitat profits (f_P) is attained by assessing of profit function based on equation 2 as below:

$$\text{Profit} = fp(\text{habitat}) = f_P\,(x1, x2, \ldots, x_{Nvar}) \quad (2)$$

The cuckoo algorithm maximizes the benefit function. To apply Cuckoo algorithm in optimization problems we can used the profit function defines as below:

$$\text{Profit} = -\,\text{Cost}\,(\text{habitat}) = -\,f_C\,(x1, x2, \ldots, x_{Nvar}) \quad (3)$$

In the beginning of process the habitat matrix creates with N_{pop} * N_{var} arrays. Then, some eggs are dedicated to each habitat. In natural life each Cuckoo lay eggs between 5 to 20 eggs , so that numbers define as a lower bond and upper bond of algorithm. Also they lay eggs within the maximum distance from their habitats, which is known as Egg Laying Radius (ELR). In each optimization problem every variable has upper bond Var_{hi} and lower bond Var_{low}. The ELR formula presented as below:

$$ELR = \alpha \times \frac{\text{Number of current cuckoo's eggs}}{\text{Total number of eggs}} \times (var_{hi} - var_{low}) \quad (4)$$

Where α represents as integer, suppose to handle the maximum value of ELR.

5. Methodology

Our methodology is divided into two phases. In the first phase, we applied density-based clustering and DBSCAN algorithm and in the second phase for improving suggested items to users, Cuckoo algorithm has been employed. Clustering of users done based on their profiles. In essence, the purpose of clustering in this research is to finding users with high similarities with intended user. The DBSCAN algorithm is one of the density-based data clustering algorithms that offers some concepts such as; core point, nearby neighbors, noise, and so on [32][40]. According to the structure of facebook website, selecting clusters were based on users neighbors; it means that each selected user put into the cluster with his/her neighbors. In fact, in social media websites users usually connect to people who are similar with him/her in terms of profile, behavior or interests [41]. Therefore, the length and the number of clusters are depending on user neighbors. Finally, the items that are ranked by users were recommended to intended users. For calculating profile similarity, we used formula defines as below [42]:

$$R_{c,c'} = \sum_{C' \in C} sim\,(C, C') \times r_{c,c'} \quad (5)$$

Where the similarity between user C and C' is estimated by Sim (C,C'). The more similarity between user C and C', the more accuracy for $r_{c,c'}$ regarding select an end user. In the next level, the mutual items among users should be found. For this purpose, we can use Cosine similarity or Pearson correlation coefficient formula. Experimental studies contend that Pearson formula is more effective that Cosine similarity. Thus, The Pearson correlation coefficient formula has been applied in this study. The formula defices as below [42][43][44]:

$$(\quad Simi_{i,j} = \frac{\sum m \in (i \cap j)\,(r_{i,m} - \acute{r}_i)(r_{j,m} - \acute{r}_j)}{\sqrt{m \in \sum_{(i \cap j)} \,(r_{i,m} - \acute{r}_i)^2}\sqrt{m \in \sum_{(i \cap j)} \,(r_{j,m} - \acute{r}_j)^2}}$$

In this formula $Simi_{i,j}$ represents a distance of two items or users. If i and j are two users, then i∩j represent all items that these two users have ranked them. r_i' and r_j' are average ranked of user i and j on the same items i∩j. $r_{i,m}$ is user I ranked to item m and $r_{j,m}$ is user j ranked to item m. Due to the fact that user’s ranking in facebook is a binary, accordingly $r_{i,m}$ and $r_{j,m}$ include likes of user i and j on item m. using formula (1) leads to identify the users that have profile similarity with each other and intended user. Afterward, formula (2) assists us to compare user’s behavior to extract the mutual items that they liked. Finally, the output can be recommended to users.

In the next phase, Cuckoo algorithm has been carried out to select a user randomly in each cluster, comparing this profile with other profiles in the cluster. If there is a similarity between this user and others, we can conclude that Cuckoo will put eggs in this profile due to high level of similarity. It should be noted that the main goal of using Cuckoo algorithm in this study is to find clusters in which we can extend a suggestion to all members of cluster. Finally, it is assumed that the users that have similar profile with each other's will have similar behaviors and if they accept one suggestion, their neighbors might accept this suggestion as well.

6. Algorithm Implementation

In this research, we applied algorithms in two sections. In section one, DBSCAN algorithm is carried out and in section two, Cuckoo optimization algorithm has been implemented. The following sections describe about both algorithms:

6.1 DBSCAN Algorithm

To address the cold-start problem we used data mining techniques. First, DBSCAN algorithm uses to identify users with profile similarity. This process is done by following algorithm [45];

Algorithm 1. DBSCAN algorithm

```
Input: D: data set, e: radius, MinPts: minimum number of points
Output: π: Clustering
clusterId = 0
for all unvisited point p ∈ D do
   mark p as visited
   N = getNeighbors (p,ε)
   if sizeof (N) < MinPts then
     mark p as noise point
   else
     clusterId + +
     add p to cluster clusterId
     for all point p′ ∈ N do
       if p' is not visited then
         mark p' as visited
         N′ = getNeighbors (p′ , ε)
         if sizeof ( N' ) ≥ MinPts then
         N = N ∪ N′
       if p' does not belong to a cluster then
         add p' to cluster clusterId
return π
```

DBSCAN (Density Based Spatial Clustering of Application with Noise) algorithm, which is disclosed in algorithm 1, is an efficient density-based clustering technique, which can explore random shape clusters and noise. In DBSCAN algorithm the following key phases are executed: 1) Labeling all elements as core, border, or noise points. 2) Putting an edge concerning all core points, which are neighbors. 3) Labeling connected core points as a cluster. 4) Involving all border points within the neighbors of a cluster into the similar cluster [45].

DBSCAN algorithm works based on distance scale, considering intended user profile as a point with coordination in (0,0) and for each difference field in neighbor's profile the position of point in axis will be changed whereas, if the similarity of two profiles were high the distance between this profile and intended user will be zero. There were 200 topics regarding sport, health, technology books and music distributed among users to like their favorite subjects. For applying clustering algorithm, due to the structure of our survey, which includes 5 main topics thus, we choose 5 users who liked more than other in each subject and each cluster contains these 5 users neighbors. Given that in this research clustering technique is based on user profiles and DBSCAN algorithm calculated the distance of points, then user profiles formed into binary code as below table:

Table 1. Coding user profiles fields on facebook

Fields	State 1	Code	State 2	Code
Age	Under 30	0	Above 30	1
Gender	Man	0	Woman	1
Education	Diploma	0	Academic degree	1
Marital status	Single	0	married	1

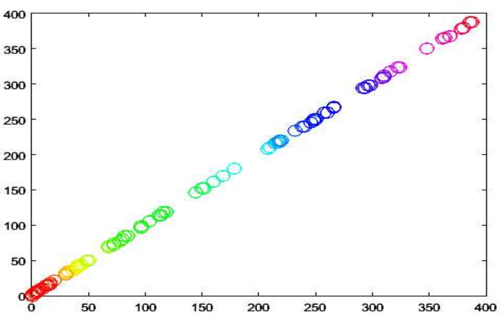

Figure 2. DBSCAN algorithm outcomes

According to table 1, DBSCAN algorithm were carried out for 5 users and the outcome is depicted in figure 2 as below (it is presented a book cluster results):

As it is shown in figure above, the user who has maximum number of likes put into position (0,0) and other users are located into different positions base on their similarities. Red colours circules indicate that these users have a high level of similarities with intended user. Results for 5 users and final outcomes are presented in table 2 as below:

Table 2. Clusters information

Clusters	User				Number of user	Profile matching	Similarit% y
Health	1	1	1	1	94	52	% 55
Technology	0	0	1	1	217	164	% 75
Book	1	1	1	1	234	98	% 42
Film	0	1	0	0	147	74	% 50
Music	0	0	0	1	119	63	% 52

As it is shown in table 2, the highest similarity is related to Technology cluster in which 75% of users have a profile similarity with intended user. Hence, technology cluster is one of the most suitable clusters to study user behavior's to extract that which item should be suggested to the novel user in terms of solving the cold-start problem.

6.2 Association rules

The main aim of association rules is to find the most frequent items among users that have high similarity with intended user containing subjects that gain a plenty of likes to recommend to user in future. Association rules algorithm consists of three elements which are defines as below [6][20]:

1. Support: represents a percentage or total transactions of A and B to total number of records. The more percentage of support, the more relation that A and B have with each other. Support rule is defines as below:

$$\text{supp}(A \Longrightarrow 0) = P(A \cap 0) = P(A)\,P(0 \mid A) = P(0)\,P(A \mid 0) \quad (7)$$

$$\text{supp}(B \Longrightarrow 1) = P(B \cap 1) = P(B)\,P(1 \mid B) = P(1)\,P(B \mid 1)$$
(8)

2. Confidence: is a ratio of the number of transactions, which include all items in the consequent, as well as the support to the number of transactions. Using this criterion besides support can evaluate an outcome of association rules algorithm. Below equations define Confidence criterion:

$$\text{Conf}(A \Rightarrow 0) = P(A \mid 0) \quad (9)$$

$$\text{Conf}(B \Rightarrow 1) = P(B \mid 1) \quad (10)$$

3. Lift: lift known as interest factor or interestingness, presents independency between item A and B. A lift ratio larger than 1.0 implies that the relationship between the antecedent and the consequent is more significant than would be expected if the two sets were independent. Equations (11) and (12) shows lift scale for A and B:

$$\text{Lift}(A \Rightarrow 0) = \frac{P(A \mid 0)}{P(0)} = \frac{P(A \cap 0)}{P(A)\,P(0)} \quad (11)$$

$$\text{Lift}(B \Rightarrow 1) = \frac{P(1 \mid B)}{P(1)} = \frac{P(B \cap 1)}{P(B)\,P(1)} \quad (12)$$

The outcomes of association rules techniques as shown below:

Rule #1: 129 --> 70
Support = 0.66159
Confidenec = 0.93233
Lift = 1.2452

Rule #2: 27 --> 183
Support = 0.63026
Confidenec = 0.91507
Lift = 1.3356

Rule #3: 90 --> 160
Support = 0.65128
Confidenec = 0.89437
Lift = 1.357

Rule #4: 95 --> 117
Support = 0.64103
Confidenec = 0.92616
Lift = 1.2757

As it is shown in figure 2, rules are extracted through association rule techniques and first rule (red color text) indicates a maximum amount of support. It means that the topic number 70 and 129 are the most popular topics among users. Second most popular subjects are 95 and 117 in rule 4 with high dependency ratio. Also, the lift number indicates an attractiveness of these rules. Therefore, extracted items are appropriate options for recommending to users to overcome the cold-start problem.

6.3 Cuckoo Optimization Algorithm

In this section, Cuckoo optimization algorithm has been utilized to find a high level of similarity between users and goal user. First of all, some nests should be created then, Cuckoos lay their eggs on these nests. The similar eggs will be survived in the nests. In this study, user profiles considered as a Cuckoo's egg and the similarity criteria's are based on four factors that are shown in table 1. Cuckoo algorithm has been applied according to equation 13 as below:

$$CS = \sum_{i=1}^{4} \sum_{j=1}^{n} \sum_{k=1}^{4} x_i \oplus m_{jk} \quad (13)$$

To making mathematical model of algorithm, we define two matrixes, comparing them with each other. In the other word, we store user profile information in the form of matrix. This matrix presents as a X_i and use $\sum_{i=1}^{4}$ to read it. As it is mentioned before, user profiles contains four field hence, we define its change interval from 1 to 4 by variable i. Then, this matrix should be compared with other matrixes defines as an m_{jk} matrix in which j represents a number of rows and k indicates the number of columns. Change interval for variable j depends on the number of eggs (users) in nest including from 1 to n. j variable represents number of fields that suppose to compare with X_i matrix. Operator XOR is defined to create output of comparing these matrixes. If the amount of two fields similar to each other, the outcome will be 0 otherwise, it will 1. Cuckoo algorithm defines as below:

Algorithm 2. Cuckoo algorithm

```
Input: D: data set , users profile matrix
Output: μ: Optimal nest
X = [a : b : m : n]
    Size = 1 to p : fields of users profile
  for i= 1 to Size
  function compare(a,b,m,n,name)
  n = 0;
  for i=1:size(a,1)
   for j=1:size(b,2)
   for k=1:size(m,3)
   for s=1:size(n,4)
    if  Isequal (i(1, a), j(2,b) , k(3,m) , s(4,n))
       n = n + 1;
  end
    end
       end
         end
             end
display(name);
 precent (display (n-size(a,1) & size(b,2) & size(m,3) & size(n,4) /100));
return μ
```

In the first place, some habitats should be provided to put user profiles as eggs into nests for identifying the similarity among eggs (other user profiles). In this research, our case study is facebook website and we choose 5 accounts to study. In each account we selected 2 users that have maximum number of neighbors. Neighbors were put in the habitats to compare with intended user in terms of profile similarity. Results of comparison of 10 users and 10 habitats are shown in table 3 as below:

Table3. Results of applying COA at 10 habitats

	Cuckoo Egg				Number of Users	% Similarity
Rest 1	0	1	1	0	704	% 52
Rest 2	0	0	1	1	593	% 59
Rest 3	1	0	1	0	1389	% 57
Rest 4	0	1	0	0	886	% 41
Rest 5	1	0	1	1	1660	% 73
Rest 6	1	1	0	1	1079	% 56
Rest 7	0	0	1	0	917	% 47
Rest 8	1	1	0	0	914	% 61
Rest 9	0	1	1	1	2398	% 66
Rest 10	1	1	1	0	1705	% 64

As it can be seen, each color in the table above, presents specific group of users. Moreover, the number of eggs and user profiles are indicated as well as similarity percentages with other eggs (profiles). According to the results, “Rest 5” is nominated for the best place for Cuckoo eggs because of 73% similarity. Also, a second best place for Cuckoo eggs is “Rest 9” with 66% similarity. Thus, these two habitats are suitable nests for growing Cuckoo eggs. In essence, for users who are located in Rest 5 and Rest 9 habitats, we can recommend similar items due to the high similarity of profiles with each other’s. Actually, we can conclude that they will have similar tendency and preferences.

7. Analysis of Results

In this section, we evaluate proposed algorithms that carried out in this research study. Cuckoo algorithm performance compared with other algorithms. For performance assessment of algorithms we used some standard measurement parameters consist of [42][46]:

7.1 Mean Absolute Error (MAE)

It is a statistical accuracy metric for assessing the average total difference between the anticipated rating and actual rating. A minor MAE value indicates more accurate prediction. It is shown in equation 13 as below:

$$\mathrm{MAE} = \frac{\sum | \acute{r}_{ij} - r_{ij}|}{\mathrm{M}} \qquad (13)$$

Where M is an overall number of predictions. $r`_{ij}$ represents the forecasted value for user i and on item j and r_{ij} is the true rating.

7.2 Accuracy and integrity

For evaluating an accuracy of model, two measurement factors are used including: “Precision” and "Recall”. These metrics explore an accuracy of model regarding user behavior prediction. The equations for both Precision and Recall are define as:

$$\text{Precision} = \frac{|TP|}{|TP+F\quad|} \qquad (14)$$

TP represent a number of records that systems predicted them correctly. FP presents a number of records that are not chosen in TP domain. Recall is a portion of relevant instances that have been retrieved over total relevant instances. The equation 15 defines Recall fraction:

$$\text{Recall} = \frac{|TP|}{|TP+F\quad|} \qquad (15)$$

TP represents a number of records that systems predicted them correctly and FN represents records that are not selected in a whole dataset. Hence, the outcome amount of Recall will be lower than Precision.

In this paper, the main aim of using Cuckoo optimization algorithm is to finding user favorites through minimum recommendation to them, suggesting appropriate items based on their preferences. In this regard, we compare Cuckoo algorithm with traditional Collaborative Filtering (CF), Particle Swarm Optimization algorithm (PSO) and Firefly Algorithm (FA) to verify our results. The CF algorithm proposes recommendation to users based on their similarities regarding profiles, ranking items and so on [42][47][48][49]. In this research study, a similarity criterion is based on two factors; gender and education level. In PSO algorithm, particles tend to move to the best and optimized position in the search-space, which is, updated as better position uncovered by neighbor particles [50][51]. In this study, user’s profile considered as a particle and other users (particles) moved to this profile if they have similarity with this user. The FA algorithm is inspired by flashing behavior of flashing bugs. Randomly produced possible solutions will be measured as fireflies where their brightness is determined by their performance on the objective function. The algorithm works in two phases; updating light intensity and movement. Fireflies are attracted to neighbors with high brightness and after several repetitive procedure, the optimize solution will be come out [52][53]. For applying FA at our data, each user is defines as a firefly and other users (fireflies) move closer to this user based on their similarities and this is the brightness of goal firefly. Hence, all fireflies that have similar profiles come together thus, we can propose specific items to this group of fireflies.

For validating our results of Cuckoo algorithm, we compared this algorithm outcomes with other algorithms mentioned above. As we choose 10 user profiles to study in previous section, the same profiles have been chosen to analyze by CF, PSO and Firefly algorithms. In each algorithm if users have similar interests with each other’s the accuracy of recommender system regarding user behavior prediction will be high. For achieving this goal we analyzed 300 user behaviors through these algorithms carried out by MATLAB software. We executed CF, PSO, FA and Cuckoo algorithms to compare the results based on MAE metric, indicating the number of users who liked same posts. Results are shown in figure 4 as below:

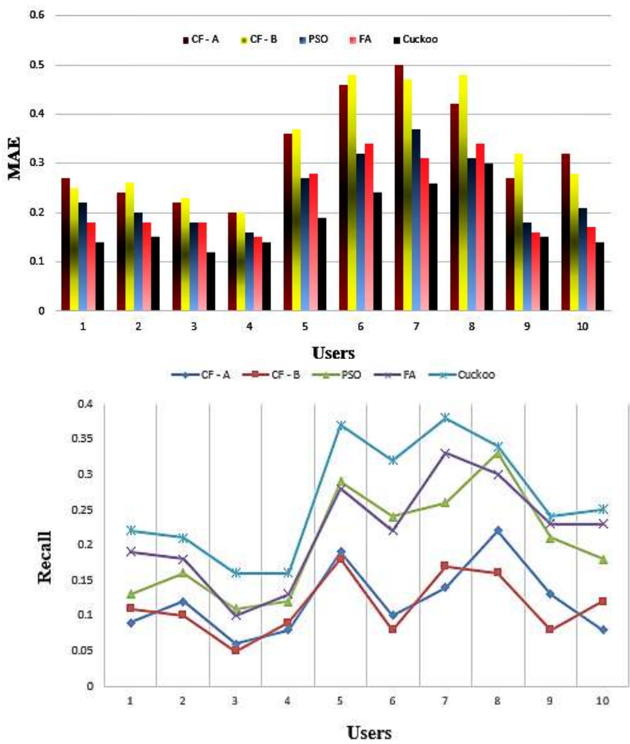

Figure 4. Results of 5 algorithms based on MAE metric

As it is shown, among all algorithms, the error rate for Cuckoo algorithm is less that others. In terms of traditional CF, we divided filtering method based on education level in CF-A and gender in CF-B. Another important point about figure 4 is that the accuracy of each algorithm is different in each position. Evaluation based on Precision and Recall metrics are shown in figure 5 and 6 as below:

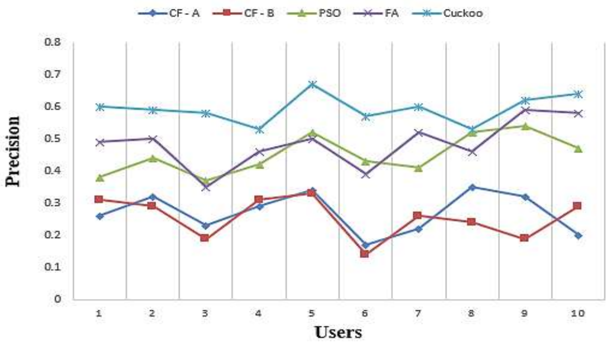

Figure 6. Accuracy of recommendation of 5 algorithms (Recall metric)

Figure 5. Accuracy of recommendation of 5 algorithms (precision metric)

other algorithms therefore, using this algorithm serve as a breeding ground for enhancing recommendation system performance.

8. Conclusion

In this research paper, cold-start problem in recommendation systems has been studies and to tackle this problem density-based clustering has been utilized. Results presented a high performance of proposed method for solving cold-start problem compared to other solutions. Actually, we showed that the relationship between user profiles similarity and their preferences on facebook. It means that users profiles similarities lead to similar requirements and interest. Therefore, user profiles are one of the most significant factors in order to recognize them, recommending interesting items based on their preferences. In the next level, Cuckoo algorithm has been applied to find users that have similar behaviors and profiles. Thus, through minimum number of suggestions we can meet users requirements. The outcomes of our study can be very useful for firms and organizations regarding marketing and advertisements. Before this, they had to study user profile individually to extract their interests, but through Cuckoo algorithm they can find user interests and preferences by groups of users, suggesting them items and services. Consequently, they can reduce advertisement and marketing costs.

References

[1] Y. A. Kim and J. Srivastava; *Impact of social influence in e-commerce decision making* ; Technical report, ICEC 07, vol.258. ACM ; New York 2007.

[2] Jamjuntra Laikhram , Chartsuwan Pantakan , Wonglimsamut Peerapong , Porkaew Kriengkrai , Supasitthimethee Umaporn ; *Social network user identification* ; IEEE 9th International Conference on Knowledge and Smart Technology (KST) ; PP. 132 – 137 ; Chonburi, Thailand - 2017.

[3] Ahmed Kazi Wasif , Hasan Mohammad Zahidul , Mohammed Noman ; *Centric Social Discovery Using Neural Network under Anonymity Constraint* ; IEEE 2017 IEEE International Conference on Cloud Engineering (IC2E) ; PP. 238 – 244 ; Vancouver, BC, Canada – 2017.

[4] Ashok Meghana , Rajanna Swathi , Joshi Pradnyesh Vineet , Sowmya Kamath S ; *A personalized recommender system using Machine Learning based Sentiment Analysis over social data* ; IEEE IEEE Students' Conference on Electrical, Electronics and Computer Science (SCEECS); PP. 1 - 6 ; Bhopal, India - 2016.

[5] Davis, D., Lichtenwalter, R., Chawla, N.V. ; *Supervised methods for multi-relational link prediction* ; Social Network Analysis and Mining ; PP. 1–15 ; 2013.

[6] Jiawei Han and Micheline Kamber ; *Data Mining: Concepts and Techniques* ; Second Edition , Diane Cerra ; Elsevier Inc - 13: 978-1-55860-901-3 ; 2006.

[7] Fang Zhiyuan , Zhang Lingqi , Chen Kun ; *A behavior mining based hybrid recommender system* ; IEEE IEEE International Conference on Big Data Analysis (ICBDA) ; PP. 1 - 5 ; - Hangzhou, China - 2016.

[8] Venugopalan S. , Srinath M. V. , Rodrigues Paul ; *Recommender System for E-Learning through Content and Profile Based Approach* ; ACM Press the Second International Conference - Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies - ICTCS '16 ; PP. 1 - 5 ; Udaipur, India - 2016.

[9] Gavriilidis Vasileios , Tefas Anastasios , Kotropoulos Constantine , Nikolaidis Nikos ; *Enhanced similarities for a music Recommender System* ; IEEE 2016 Digital Media Industry & Academic Forum (DMIAF) - Digital Media Industry & Academic Forum (DMIAF) ; PP. 113 - 116 ; Santorini, Greece - 2016.

[10] Zhao Vicky Na , Moh Melody , Moh Teng-Sheng ; *Contextual-Aware Hybrid Recommender System for Mixed Cold-Start Problems in Privacy Protection* ; IEEE 2nd International Conference on Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing (HPSC) and IEEE International Conference on Intelligent Data and Security (IDS) ; PP. 400 - 405 ; New York, NY, USA - 2016.

[11] Yi Peng , Yang Chen , Zhou Xiaoming , Li Chen ; *A movie cold-start recommendation method optimized similarity measure* ; IEEE 2016 16th International Symposium on Communications and Information Technologies (ISCIT) ; PP. 231 - 234 ; Qingdao, China - 2016.

[12] Wei Jian , He Jianhua , Chen Kai , Zhou Yi , Tang Zuoyin ; *Collaborative Filtering and Deep Learning Based Hybrid Recommendation for Cold Start Problem* ; Collaborative Filtering and Deep Learning Based Hybrid Recommendation for Cold Start Problem ; PP. 874 - 877 ; 2016.

[13] Reshma R , Ambikesh G , Thilagam P Santhi ; *Alleviating data sparsity and cold start in recommender systems using social behaviour* ; IEEE Fifth International Conference on Recent Trends in Information Technology (ICRTIT) ; PP. 1 - 8 ; Chennai, India - 2016.

[14] Forouzandeh Saman, Soltanpanah Heirsh , Sheikhahmadi Amir ; *Content marketing through data mining on Facebook social network* ; Journal of Webology ; Vol. 11(1), Article 118 ; 2014. Available at: http://www.webology.org/2014/v11n1/a118.pdf

[15] Yung-Ming Li , Chun-Te Wu , Cheng-Yang Lai ; *A social recommender mechanism for e-commerce: Combining similarity, trust, and relationship* ; Decision Support Systems 55 - 740–752 ; Elsevier – 2013.

[16] Ashwini Nadkarni , Stefan G. Hofmann ; *Review: Why do people use Facebook?* ; Journal of Personality and Individual Differences 52 - 243–249 ; Elsevier – 2012.

[17] Greitemeyer Tobias ; *Facebook and people's state self-esteem: The impact of the number of other users' Facebook friends* ; Computers in Human Behavior ; Vol. 59 , PP. 182 – 186 ; 2016.

[18] Petrocchi Nicola , Asnaani Anu , Martinez Alejandra Piquer , Nadkarni Ashwini , Hofmann Stefan G ; *Differences Between People Who Use Only Facebook and Those Who Use Facebook Plus Twitter* ; International Journal of Human-Computer Interaction ; Vol. 31(2) , PP. 157 – 165 ; 2015.

[19] Rodavia Maria Rosario D. , Ballera Melvin , Clemente Gina , Ambat Shaneth ; *AutoRec: A Recommender System Based on Social Media Stream* ; IEEE 2017 International Conference on Platform Technology and Service (PlatCon) ; PP. 1 – 6 ; Busan, South Korea – 2017.

[20] Kim Kuinam J. , Joukov Nikolai ; *Sentiment Analysis and User Similarity for Social Recommender System: An Experimental Study* ; Springer-Verlag (book) , [Lecture Notes in Electrical Engineering] Information Science and Applications (ICISA) ; Vol. 376, Chapter 109 , PP. 1147 – 1156 ; 2016.

[21] Forouzandeh Saman, Soltanpanah Heirsh , Sheikhahmadi Amir ; *Application of Data Mining in Designing a Recommender System on Social Networks* ; International Journal of Computer Applications ; Vol . 124 (1) ; 2015.

[22] Aghdam, A. R., Kamalpour, M., & Sim, A. T. H. (2014). Knowledge Discovery for Tourism Using Data Mining and Qualitative Analysis: A Case Study at Johor Bahru, Malaysia. International Journal of Asian Business and Information Management (IJABIM), 5(4), 48-59.

[23] Katarya Rahul , Verma Om Prakash ; *An effective collaborative movie recommender system with cuckoo search* ; Egyptian Informatics Journal ; 2016.

[24] Matthias Braunhofer ; *Hybrid Solution of the Cold-Start Problem in Context-Aware Recommender Systems* ; Springer International Publishing Switzerland ; LNCS 8538, PP. 484–489 ; 2014.

[25] Tomeo Paolo , Fernández-Tobías Ignacio , Di Noia Tommaso , Cantador, Iván ; *Exploiting Linked Open Data in Cold-start Recommendations with Positive-only Feedback* ; ACM Press the 4th Spanish Conference - Proceedings of the 4th Spanish Conference on Information Retrieval - CERI '16 - Granada, Spain ; PP. 1 - 8 ; 2016.

[26] Chou Szu-Yu , Yang Yi-Hsuan , Jang Jyh-Shing Roger , Lin Yu-Ching ; *Addressing Cold Start for Next-song Recommendation* ; ACM Press the 10th ACM Conference - Boston, Proceedings of the 10th ACM Conference on Recommender Systems - RecSys '16- Massachusetts, USA ; PP. 115 - 118 ; 2016.

[27] Blerina Lika, Kostas Kolomvatsos, Stathes Hadjiefthymiades ; *Facing the cold start problem in recommender systems* ; Expert Systems with Applications 41 (2014) 2065–2073 ; Elsevier – 2014.

[28] Andre Luiz Vizine Pereira , Eduardo Raul Hruschka ; *Simultaneous co-clustering and learning to address the cold start problem in recommender systems* ; Knowledge-Based Systems with Applications ; Elsevier – 2015.

[29] Hridya Sobhanam , A.K.Mariappan ; *Addressing cold start problem in recommender systems using association rules and clustering technique* ; International Conference on Computer Communication and Informatics (*ICCCI* - 2013) - Coimbatore, INDIA ; IEEE – 2013.

[30] Bahareh Shams Zamenjani ; Marketing based on user behavior in Facebook social network through recommender system design ; IEEE International Congress on Technology, Communication and Knowledge (ICTCK) ; pp . 181 – 185 ; 2015.

[31] Yang Xin-She , Suash Deb ; *Cuckoo Search via Lévy flights* ; IEEE World Congress on Nature & Biologically Inspired Computing (NaBIC) ; PP. 210 – 214 ; Coimbatore, India – 2009.

[32] Charu C. Aggarwal ; *Data Classification : Algorithms and Applications* ; International Standard Book Number-13: 978-1-4665-8675-8 (eBook - PDF) ; Taylor & Francis Group, LLC ; 2015.

[33] Yang Xin-She ; *Multi-Objective Optimization of a Real-World Manufacturing Process Using Cuckoo Search* ; Springer-Verlag (Book) - [Studies in Computational Intelligence] Cuckoo Search and Firefly Algorithm ; Vol, 516 (9) , PP. 179 – 193 ; 2014.

[34] Sulaiman M. H. , Mohamed M. R. ; *Solving economic dispatch problems utilizing Cuckoo Search algorithm* ; IEEE 2014 IEEE 8th International Power Engineering and Optimization Conference (PEOCO) - Langkawi, Malaysia ; PP. 89 - 93 ; 2014.

[35] Ljouad Tarik , Amine Aouatif , Rziza Mohammed ; *A hybrid mobile object tracker based on the modified Cuckoo Search algorithm and the Kalman Filter* ; Journal of Pattern Recognition – Elsevier ; Vol. 47(11) , PP. 3597 – 3613 ; 2014.

[36] Elkeran Ahmed ; *A new approach for sheet nesting problem using guided cuckoo search and pairwise*

clustering ; European Journal of Operational Research – Elsevier ; Vol. 231(3) , PP. 757 – 769; 2013.

[37] Sanchez Mauricio A. , Castillo Oscar , Castro Juan R. ; *Information granule formation via the concept of uncertainty-based information with Interval Type-2 Fuzzy Sets representation and Takagi–Sugeno–Kang consequents optimized with Cuckoo search* ; Journal of Applied Soft Computing – Elsevier ; Vol. 27 , PP. 602 – 609 ; 2015.

[38] Masoud Asadi , Yidan Song , Bengt Sunden , Gongnan Xie ; *Economic optimization design of shell-and-tube heat exchangers by a cuckoo-search-algorithm* ; Journal of Applied Thermal Engineering ; Vol. 73 (1) , PP. 1032 – 1040 ; 2014.

[39] Ramin Rajabioun ; *Cuckoo Optimization Algorithm* ; Journal of Applied Soft Computing ; Vol. 11 (8) ; PP. 5508 - 5518 ; Elsevier – 2011.

[40] Mohammed J. Zaki , Wagner Meira Jr. ; *Data Mining and Analysis: Fundamental Concepts and Algorithms* ; Printed in the United States of America ; ISBN 978-0-521-76633-3 ; 2014.

[41] Ashwini Nadkarni , Stefan G. Hofmann ; *Review: Why do people use Facebook?* ; Journal of Personality and Individual Differences 52 - 243–249 ; Elsevier – 2012.

[42] Ricci Francesco, Rokach Lior , Shapira Bracha , Kantor Paul B. ; *Recommender Systems Handbook || Introduction to Recommender Systems Handbook* ; Springer-Verlag (Book) , Chapter 1 , PP. 1 – 35 ; 2011.

[43] H Li , D Wu , N Mamoulis ; *A Revisit to Social Network-Based Recommender Systems* ; Publication rights licensed to ACM. ACM - 978-1-4503-2257-7/14/07 ; 2014.

[44] Qi Lianyong , Dou Wanchun , Zhang Xuyun ; *filtering approach for cold-start problem in web service recommendation* ; ACM Press the Australasian Computer Science Week Multiconference , Proceedings of the Australasian Computer Science Week Multiconference on - ACSW '17 ; PP. 1 – 9 ; Geelong , Australia – 2017.

[45] Selim Mimaroglu , Emin Aksehirli ; *Improving DBSCAN's execution time by using a pruning technique on bit vectors* ; Pattern Recognition Letters ; Vol. 32 (13) , PP. 1572 – 1580 ; 2011.

[46] Wang Zan , Yu Xue , Feng Nan , Wang Zhenhua ; *An Improved Collaborative Movie Recommendation System using Computational Intelligence* ; Journal of Visual Languages & Computing ; Vol. 25(6) , PP. 667 – 675 ; 2014.

[47] Yang Xiwang , Guo Yang , Liu Yong , Steck Harald ; *A survey of collaborative filtering based social recommender systems* ; Computer Communications ; Vol. 41 , PP. 1 – 10 ; 2014.

[48] Elahi Mehdi , Ricci Francesco , Rubens Neil ; *A survey of active learning in collaborative filtering recommender systems* ; Computer Science Review ; 2016.

[49] DouYingtong , Yang Hao , Deng Xiaolong ; *A Survey of Collaborative Filtering Algorithms for Social Recommender Systems* ; IEEE 12th International Conference on Semantics, Knowledge and Grids (SKG) ; PP. 40 – 46 ; Beijing , China – 2016.

[50] AlRashidi M.R. , El-Hawary M.E. ; *A Survey of Particle Swarm Optimization Applications in Electric Power Systems* ; IEEE Transactions on Evolutionary Computation ; Vol. 13 (4) ; PP. 913 – 918 ; 2009.

[51] Shahana K , Ghosh Subhajit , Jeganathan C. ; *A survey of particle swarm optimization and random forest based land cover classification* ; IEEE International Conference on Computing, Communication and Automation (ICCCA); PP. 241 - 245 ; Greater Noida, India - 2016.

[52] Kisi Ozgur , Shiri Jalal , Karimi Sepideh , Shamshirband Shahaboddin , Motamedi Shervin , Petković Dalibor , Hashim Roslan ; *A survey of water level fluctuation predicting in Urmia Lake using support vector machine with firefly algorithm* ; Applied Mathematics and Computation ; Vol. 270 , PP. 731 – 743 ; 2015.

[53] Tilahun Surafel Luleseged , Ngnotchouye Jean Medard T. ; *Firefly algorithm for discrete optimization problems: A survey* ; KSCE Journal of Civil Engineering ; Vol. 21(2) , PP. 535 – 545 ; 2017.