CrossMark

A novel optimization algorithm for recommender system using modified fuzzy *c*-means clustering approach

C. Selvi[1] · E. Sivasankar[1]

Abstract Recommender system (RS) is an emerging technique in information retrieval to handle a large amount of online data effectively. It provides recommendation to the online user in order to achieve their correct decisions on items/services quickly and easily. Collaborative filtering (CF) is one of the key approaches for RS that generates recommendation to the online user based on the rating similarity with other users. Unsupervised clustering is a class of model-based CF, which is more preferable because it provides the simple and effective recommendation. This class of CF suffers by higher error rate and takes more iterations for convergence. This study proposes a modified fuzzy *c*-means clustering approach to eliminate these issues. A novel modified cuckoo search (MCS) algorithm is proposed to optimize the data points in each cluster that provides an effective recommendation. The performance of proposed RS is measured by conducting experimental analysis on benchmark Movie-Lens dataset. To show the effectiveness of proposed MCS algorithm, the results are compared with popular optimization algorithms, namely particle swarm optimization and cuckoo search, using benchmark optimization functions.

Keywords Recommender system (RS) · Collaborative filtering (CF) · Clustering approach · Optimization algorithm · Modified fuzzy *c*-means (MFCM) · Modified cuckoo search (MCS)

✉ C. Selvi
  selvichandran.it@gmail.com

  E. Sivasankar
  sivasankarelango@gmail.com

[1] Department of Computer Science and Engineering, National Institute of Technology, Tiruchirappalli, India

1 Introduction

The advancement of technology and growth of population make a wide usage of informal social websites that result in an exponential expansion of web contents. Due to these enormous and variety of web contents confuses the online user to making correct decision for their requirements. Personalized recommender system (RS) is an effective tool to take care of information overloaded problem (Linden et al. 2003). RS accumulates information according to the user's preferences for various activities such as movies, shopping, tourism, TV and taxi through two ways, either implicitly or explicitly. Explicit ratings are numerical ratings and that are directly given to RS model, whereas user opinions are given in text format for implicit rating. An e-commerce site uses the RS to predict user's purchase behaviors using the user's past preference information and suggests products of user's potential interest in order to enhance their profit. Algorithms are used to predict the user preference by interacting with a user for both to acquire data and to provide recommendations. The development of RS is based on the domain of recommendation and agencies such as Amazon and Netflix.

The main categories of RS are collaborative filtering (CF), content-based filtering (CBF) and hybrid filtering techniques. CF is widely used mechanism for RS filtering, which provides suggestion by analyzing the rating details of items or the users. CBF based RS works by matching the features of items that are purchased in the past with features of new item that is ready to be recommended (Lops et al. 2011). If the matches are found between the features of an items, then the corresponding item is recommended to the target user. CF-based model is domain independent and provides better accuracy than the CBF. If the content of the items is not available or challenging to acquire, CF rec-

ommends item to the target user based on the other user's feedback values (neighbors). Due to its easiness, productivity and ability to produce accurate and personalized recommendations, CF is considered as a prominent method in RS.

CF is categorized into memory-based and model-based methods. In memory-based CF, initially, the similarity between the target user and all other users is calculated. Further, the nearest similar users are identified based on the calculated similarity value. Using the identified nearest users, the target item values are predicted and the recommendations are provided accordingly. Memory-based CF technique is further classified into user-based and item-based methods. The user-based method predicts the value of the new item with respect to the nearest neighbors of a target user, whereas in an item-based method, a prediction is done with respect to the nearest neighboring item to the item of a target user. In memory-based CF, on each attempt, the entire user–item rating matrix has to be loaded into memory to calculate similarity among the users. This leads to the scalability issue and also takes longer computational time. To address this issue, the model-based CF is preferred by the researchers.

The model-based technique uses the supervised or unsupervised approaches to learn a model from training user–item rating matrix. A learned model is further used for predicting the rating of target user's item. In this method, once the model is built, then the user–item matrix is not required, which enables the prediction to be done in offline and faster even the number of users and items increases. Many of the research work has been done by using model-based supervised learning techniques (Braida et al. 2015). The main shortcoming in using supervised learning technique is creating a training dataset with the class label. Thus, the accuracy of the RS model depends on the labeled training dataset (Braida et al. 2015). To avoid this limitation, model-based methods with unsupervised clustering approaches are mostly preferred for RS-related applications.

The model-based unsupervised clustering approach groups the similar data points together and finds the pattern for unlabeled data points. Among many clustering, k-means clustering is widely used partition-based clustering approach in RS that allocates data points to the clusters with highest similarity value (Al Mamunur Rashid et al. 2006; Sarwar et al. 2002; Xue et al. 2005). That is, it allocates data point which belongs to exactly one cluster (Ghosh and Dubey 2013). But the data points have the probability to reside two or more clusters. This is measured by fuzzy membership function that used in fuzzy c-means (FCM) clustering approach. Many research has been conducted by using FCM method and its variants (Ghosh and Dubey 2013; Li and Kim 2003; Nasser et al. 2006; Wu and Li 2008; Kuo-Lung and Yang 2002). The main limitation of FCM is that it pro-

vides considerably large error rate and takes more iterations to get well-framed clusters. To overcome the above issues, this study proposes a modified FCM (MFCM) clustering approach, which reduces the error rate and increases an accuracy of recommendation significantly.

After getting well-framed clusters, an optimum data points in each cluster have to be identified and the recommendations made accordingly. Optimization algorithms are used to identify the optimum data points from the each cluster. Many optimization algorithms depend on the applications perfective. Genetic algorithm (GA) (Ar and Bostanci 2016; Bobadilla et al. 2011; Demir et al. 2010), particle swarm optimization (PSO) (Katarya and Verma 2016; Ujjin and Bentley 2003), cuckoo search (CS) (Liu and Fu 2014; Raja and Vishnupriya 2016; Roy and Chaudhuri 2013; Tran et al. 2015), water cycle algorithm (WCA) (Pahnehkolaei et al. 2017), memetic algorithm (MA) (Arab and Alfi 2015; Banati and Mehta 2010; Mousavi and Alfi 2015), artificial bee colony (ABC) (Mernik et al. 2015) and harmony search (HS) (Ameli et al. 2016) are some of the popular optimization algorithms provided in the literature.

The main objective of this study is to select the optimum data points from each cluster to improve the recommendation accuracy. To achieve this, a novel optimization algorithm is proposed based on modified cuckoo search (MCS) algorithms, which eliminate the limitation of traditional CS algorithm and increase the recommendation performance significantly. To compare the efficiency of the proposed MCS algorithm, it is compared with popular optimization algorithms such as PSO and CS.

The main contribution of this study includes,

– Proposed a new CF-based similarity measure that preprocesses the MovieLens dataset by removing the movies with differed characteristics.
– Proposed a new MFCM clustering approach on processed dataset to get a cluster with reduced error.
– The effectiveness of proposed MFCM clustering approach is validated experimentally with existing FCM method.
– The optimal user in each cluster is obtained effectively by proposing a new MCS techniques.
– The effectiveness of proposed MCS is tested with standard benchmark optimization functions and compared with PSO and CS.

The rest of this paper is organized as follows: Section 2 reviews some of the popular model-based CF methods. Preprocessing, proposed methodologies and details of performance measures are elaborated in Sect. 3. Results and discussion of various combinations of clustering and optimization algorithms are explained in Sect. 4. Finally, Sect. 5 concludes the work.

2 Literature survey

An improvement in web 2.0 and social network resulted in an increasing number of individuals who are coming as a main source of data producer rather than being just data consumers. Web data are mainly produced by users that yield result in data overloading. This will not help a user to analyze and take better, precise choices effectively. Handling such a high volume of information by a human being is incomprehensible. But these informations have to be processed to help a user to come up with better decision through recommendation.

Numerous RS algorithms are produced in different applications, for example, online business, media, computerized library and online promoting (Linden et al. 2003). The main classes of RS include CBF (Lops et al. 2011), CF (Patra et al. 2015), and hybrid techniques. CF is categorized as memory-based and model-based methods (Lu et al. 2015; Zhang et al. 2016).

Memory-based CF works were performed by recommending items to an end user using similarity measures (Patra et al. 2015) such as Pearson, cosine, adjusted cosine, constrained Pearson, mean squared difference (MSD), Jaccard MSD and Bhattacharyya. Though many similarity measures are used to predict the value of an item for a recommendation, it has many drawbacks as it considers only co-rated items and leaves some user rating while calculating similarity. So this method is not suitable for a sparse dataset. To overcome these drawbacks, researchers concentrated on model-based CF techniques, which works in building a model from user–item rating matrix using supervised or unsupervised learning algorithms. Model-based CF with supervised algorithms needs a history of labeled data before building a model (Braida et al. 2015). The accuracy of RS is based on the correctness of labeled training data. Labeling training data needs more human effort, and the results are erroneous. To handle these issues, the model is built with unsupervised learning algorithms.

The extra impact on which some RSs are based on clustering approach divides a dataset into groups based on some similarity measures (Mukhopadhyay et al. 2015). Clustering groups the similar users based on their preference in order to enhance the recommendation performance. The k-means algorithm is an efficient clustering approach compared to other clustering approaches in relation to time, complexity or effectiveness for a specific number of clusters (Kanungo et al. 2002). The k-means clustering approach and its variants are applied to various issues of RS applications such as sparsity, scalability and cold-start problem (Zanardi 2011). Bisecting k-means clustering approach was introduced to distinguish and analyze a lot of data points available in privacy preserving application (Bilge and Polat 2013) and web-based movie RS (Sarwar et al. 2002). The new centroid selection algorithm was introduced using k-means algorithm to provide accu-

rate recommendations and also reduces the cost of training the cluster (Zahra et al. 2015). This algorithm improves an accuracy of a model than the random selection of centroids. The CLUSTKNN clustering approach was implemented to improve the quality of RS and also handles the data points in large-scale application (Al Mamunur Rashid et al. 2006). The k-means cluster-based smoothing CF was applied in the large dataset which provides efficient recommendations (Xue et al. 2005). It also eliminates data sparsity and scalability issues. A new clustered social ranking algorithm was developed to support new web user for finding the content interest (Zanardi and Capra 2011; Zanardi 2011). The k-means algorithm and its variants form a cluster where the data points belong to only one cluster at a time (Tsai and Hung 2012). Also, it does not guarantee convergence of clustering process. But some of the data points have a probability to fall into two or more cluster at a time.

This issue was tackled by FCM clustering approach, which calculates a membership value from 0 to 1. This membership value denotes the maximum chance that the data point can place into a particular cluster. The cold-start problem of RS was handled by FCM clustering approach (Li and Kim 2003). FCM generates group rating matrix, which is combined with item ratings to provide recommendations to the cold-start user. The fuzzy k-means expectation maximization (FKEM) algorithm was introduced, which combines weighted fuzzy k-means algorithm with expectation maximization algorithm to attain guaranteed convergence in less number of k value (Nasser et al. 2006). In order to get better interpretability the modified FCM was used, which combines FCM clustering and matrix factorization techniques (Wu and Li 2008). This achieves better recommendation accuracy in Netflix Prize dataset. Alternative FCM (AFCM) and alternative hard c-means (AHCM) algorithm was proposed where the AFCM method differentiates abnormal tissues from the normal tissues better than the AHCM method (Kuo-Lung and Yang 2002). The FCM is used to cluster the data points, and GA is used to find the optimal solution by computing the similarity among the data points to provide the effective recommendation (Gupta et al. 2015). The FCM algorithm was employed to calculate a weight of the web page to provide web page recommendation (Katarya and Verma 2016). The FCM was also used to analyze the user behavior and provides recommendations to the target user for the web log files collected from National Museum of History (Fang and Liu 2003). FCM and its variants are used in various applications such as online learning (Bodyanskiy et al. 2017), image processing (Wen and Celebi 2011; Li and Shen 2010) and biomedical application (Kuo-Lung and Yang 2002). Though FCM and its variants have been used in many application, it yields significantly large error and converges in a large number of iterations to get well-formed clusters (Guo et al. 2015; Koohi and Kiani 2016; Merialdo 1999; Nilashi et al.

2015; Thong 2015). Finding an effective clustering approach to design a well-framed cluster is still a challenge. To overcome the above-mentioned issues, this paper proposes new MFCM clustering approach to improve the RS accuracy.

The aim of this paper is to provide an efficient recommendation to the target user by extracting an optimized user from the processed clusters. Various optimization algorithms have been used in many application. They include GA (Ar and Bostanci 2016; Bobadilla et al. 2011; Demir et al. 2010), PSO (Katarya and Verma 2016; Ujjin and Bentley 2003), CS (Liu and Fu 2014; Raja and Vishnupriya 2016; Roy and Chaudhuri 2013; Tran et al. 2015), WCA (Pahnehkolaei et al. 2017), ABC (Mernik et al. 2015), HS (Ameli et al. 2016) and MA (Arab and Alfi 2015; Banati and Mehta 2010; Mousavi and Alfi 2015). Optimization algorithms are modified and hybridized for the convenience of problem based on an application. GA was used to obtain an optimal similarity function, which performed better than conventional similarity measure (Bobadilla et al. 2011). An innovative music RS was introduced, which combines CBF and interactive GA (Kim et al. 2010). The k-means clustering based on GA was proposed to segment electronic commerce market effectively (Kim and Ahn 2008). PSO and FCM are combined to map the values to an attribute effectively (Brouwer and Groenwold 2010). Also, PSO was used to prioritize individual fuzzified features of a user to get an effective recommendation (Wasid and Kant 2015). A web-based RS was proposed by hierarchical PSO-based clustering, which combines properties of hierarchical and partitional clustering (Alam et al. 2012). Data point is clustered using k-means algorithms and is optimized using CS techniques to provide an efficient recommendation (Katarya and Verma 2016). Clustering with a bio-inspired CS algorithm (Yang and Deb 2014) delivers optimized outcomes. As far the RS is considered, the solutions are scattered in GA due to the randomness nature (crossover and mutation). Exploration and exploitation of search space in PSO are not efficient due to the lesser step size random walk. In CS algorithm, the step size to explore the search space is very efficient than PSO due to the random walk through Levy flight. Also, the number of parameter used to configure initial search is less. CS also balances the search space exploitation and exploration effectively than other optimization algorithms. Owing to these advantages, the CS is preferred in RS, where every nest represents the individual solution that makes the effective global solution that is acceptable (Hatami and Pashazadeh 2014).

To get an optimal solution, the search space exploration and exploitation should be performed in an effective way with solution convergence on less number of iterations. The search space is covered in CS with step size calculation using Levy flight distribution, which is not finite. To make a finite step size, the Levy exponent in CS is replaced by proposing a Gaussian exponent function. The above-mentioned solution
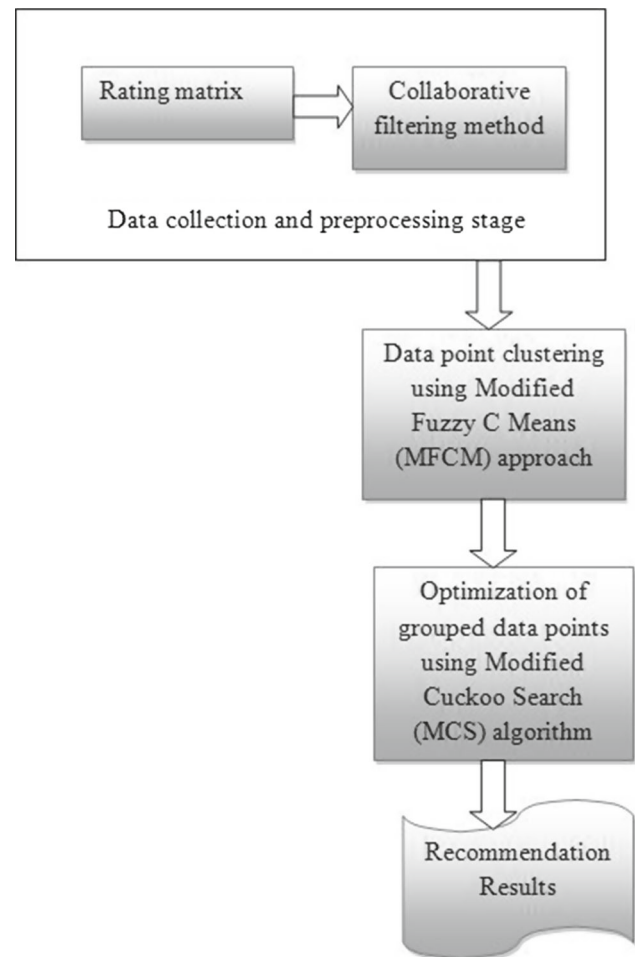
Fig. 1 Flow diagram of proposed RS

is implemented with proposed MCS algorithm to provide an efficient recommendation using proposed MFCM clustering approach.

3 Proposed methodology

The main objective of this study is to apply MCS optimization algorithm to provide an efficient recommendation to the target user using MCFM clustering approach. The following section represents the steps to be carried out for proposed RS as represented in Fig. 1.

3.1 Data collection and preprocessing

The user's rating on the collection of movies is obtained from the benchmark MovieLens dataset.[1] Movies with differed characteristics, that is the movie is not having similar characteristics with other movies, should be removed. Because when these movies are clustered into some groups, they

---

[1] MovieLens 100k dataset. https://grouplens.org/datasets/movielens/100k/.

will, in turn, affect the recommendation accuracy. So, these movies are considered as outliers and should be removed from the dataset before applying the clustering approach.

The movie-based CF method is proposed to find the similarity among the movies, which is the ratio between multiplication and addition of ratings for all the users of co-rated movies. Let X be the set of users $\{u_1, u_2, u_3, \ldots, u_x\}$ and Y be the set of movies $\{m_1, m_2, m_3, \ldots, m_y\}$. Let u be any arbitrary user in a group X who has rated both the movies m_p and m_q ($m_p, m_q \in Y$). Then, the proposed similarity measure is given in Eq. 1.

$$sim(m_p, m_q) = \frac{\sum_{u \in (m_p \cap m_q)} r_{u,m_p} \times r_{u,m_q}}{\sum_{u \in (m_p \cap m_q)} r_{u,m_p} + r_{u,m_q}} \quad (1)$$

where r_{u,m_p} and r_{u,m_q} represent the rating of user u on movies p and q, respectively. Finally, the movie with less similarity value (below 0.3 by experimentation) is removed from the dataset, and the dataset with remaining movies is considered for further processing.

3.2 MFCM clustering approach

FCM clustering approach was developed by Dunn (1973) and further improved by Bezdek (1981). In FCM clustering, each data point has a probability of having a place with each cluster, as opposed to totally having a place with only one cluster as it is the situation in the conventional k-means. FCM places the data points in multidimensional space into the finite number of clusters. The aim of FCM is to find the centroid that increases the similarity between data points within the cluster.

Generally, data points do not exactly belong to only one cluster rather it may have a chance to enter into more than one cluster. FCM uses the membership function that assigns the data points to multiple clusters with different degrees of membership value (η) between 0 and 1. The goal of FCM is to reduce error, and the minimization of error objective function is given in Eq. 2.

$$Objective\_function_{FCM} = \sum_{i=1}^{l} \sum_{j=1}^{n} (\eta_{ij})^m \parallel t_i - c_j \parallel^2 \quad (2)$$

where η_{ij} represents the membership value of data point i to the cluster centroid j and m is the fuzzy index value between 1 to ∞. Initially, the cluster centroid c_j is taken randomly. The membership value η_{ij} and cluster centroid c_j values are updated till the optimal error value of objective function is reached.

Each cluster method run individually with a various number of group set so as to discover the best clustering approach that can provide the most noteworthy rate of recommenda-

tion accuracy. FCM has turned as a well-known clustering approach and performs well in the certain type of applications. FCM is highly suffered by large error rate and more number of iterations to get well-framed clusters. It reduces the performance of RS considerably. To overcome this limitation, MFCM clustering approach that modifies the objective function of FCM using a novel weighted mean distance function (ρ) is proposed as given in Eq. 3. The objective function of MFCM is coined as,

$$Objective\_function_{MFCM} = \sum_{i=1}^{l} \sum_{j=1}^{n} (\eta_{ij})^m \frac{\parallel t_i - c_j \parallel^2}{\rho_j} \quad (3)$$

where ρ_j is the weighted mean distance function for cluster j. It is formulated as in Eq. 4.

$$\rho_j = \left(\frac{\sum_{j=1}^{n} \eta_{ij}^m \parallel t_i - c_j \parallel^2}{\sum_{j=1}^{n} \eta_{ij}^m} \right)^{1/2} \quad (4)$$

Considering ρ in the objective function of MFCM reduces the error rate and a number of iterations significantly. The formula for calculating membership function η_{ij} is given in Eq. 5.

$$\eta_{ij} = \frac{1}{\sum_{k=1}^{n} \left(\frac{\parallel t_i - c_j \parallel}{\parallel t_i - c_k \parallel} \right)^{\frac{2}{m-1}}} \quad (5)$$

The new cluster centroid c_j is calculated as given in Eq. 6.

$$c_j = \frac{\sum_{i=1}^{l} \eta_{ij}^m \times t_i}{\sum_{i=1}^{l} \eta_{ij}^m}, \quad \forall j = 1, 2, \ldots, n \quad (6)$$

Repeat the calculation of η_{ij} and c_j till the minimum error value for $Objective\_function_{MFCM}$ is reached or the membership value change between two iterations is less than the given sensitivity threshold φ. Equation 7 represents the stopping criterion based on membership values between consecutive iterations k and $k + 1$.

$$\max_{ij} \left\| \eta_{ij}^{(k+1)} - \eta_{ij}^{(k)} \right\| < \varphi \quad (7)$$

where φ represents the termination criterion which ranges from 0 and 1. Finally, the data points are clustered with minimum error value using proposed MFCM approach.

3.3 MCS algorithm

CS algorithm was developed by Yang and Deb (2009). CS algorithm uses the following rules (Yang and Deb 2014): (1) every cuckoo lays one egg at a time in a randomly chosen nest from the finite number of host nests. (2) The high quality

of egg within the nest will be carried to succeeding iterations. (3) The fresh and better quality solution replaces the nastiest solution in the nest. (4) The available host nest count is fixed, and the host bird finds the cuckoo egg with a probability, $pa \in [0, 1]$. In this situation, the host bird can decide to throw the cuckoo egg away from the nest or it abandons the nest and builds new nest.

The main objective of RS is to suggest the movies to the user based on the interest of best-optimized user on each cluster obtained from Sect. 3.2. The control parameters needed to find the best optimal solution using PSO algorithm include population size, acceleration constants C1 and C2, inertia weight W and maximum number of iterations. The control parameters used in CS include population size, step size and Levy exponent. The random walk through Levy flight in CS is more efficient in investigating the search space as the contrast with PSO in light of the fact that the step size in CS is very long for more number of iterations. Also, PSO needs larger memory space for critical applications and reaches the regional optimal solution easily. Therefore, the CS optimization algorithm is preferred to explore the efficient solution compared to PSO. Also in CS, the cuckoo solutions are updated when it is moved to the new host.

In order to explore the search space in less number of iterations and also to get a finite number of step size, the Levy exponent function in CS is replaced by novel Gaussian exponent function called as MCS. Before updating the solution space using Gaussian exponent at each iteration, the original population should be stored at each nest as a Local best (Lbest). In MCS, it is necessary to retain the Lbest population in each host and Global best (Gbest) population in each iteration.

The steps proposed for MCS algorithm are given as follows,

Step 1: Initialization phase

The first stage in the CS algorithm is the initialization phase where host nest are initialized randomly, i.e., A_l, where $l = 1, 2, 3, \ldots, n$.

Step 2: Generating new cuckoo phase

With the assistance of proposed Gaussian distribution called Levy stable distribution, a cuckoo is chosen arbitrarily to produce novel solutions. Thus, the introduced cuckoo is evaluated by utilizing the fitness function to determine the incredibility of the solutions.

Step 3: Fitness Evaluation Phase

The proposed fitness function is represented in Eq. 8. The fitness value is calculated for every cuckoo in the search space.

$$P_{\max} = \frac{Q_s}{Q_m} \tag{8}$$

where $P_{\max}$ is the maximum probability fitness function, Q_s is the selected population, which is the sum of ratings of user who rated the movie and Q_m represents total number of population.

Step 4: Updation phase

In CS, the solution is enhanced by Levy flight function. By evaluating the predominance of the new solution, a nest is picked randomly among them. On the off chance, the new solution in the chosen nest is better and progressive than the earlier solution, and it is re-established by the new cuckoo solution. Else, the previous solution is considered as the finest solution. The Levy flights utilized for the general CS calculation are represented in Eq. 9 and is demonstrated as follows

$$f_{\text{Levy}}^{(l+1)} = f_{\text{Levy}}^{(l)} + \gamma \oplus \text{Levy}(\lambda), \quad 1 < \lambda \leq 3 \tag{9}$$

where γ represents the step size ($\gamma > 0$).

The step size in the Levy flight distribution is not finite. Generalization of the control limit theorem is must due to the independent and identically distributed random variables whose variances are not finite. This generalization is achieved by proposing the Gaussian exponent function, which replaces Levy exponent. The Lbest Cuckoo solution in each host is identified and stored before updating the solution. The generation of new solution using Gaussian exponent is coined in Eq. 10.

$$f_{\text{Gaussian}}^{(l+1)} = f_{\text{Gaussian}}^{(l)} + \gamma \oplus \sigma_a \tag{10}$$

where $\sigma_a = \sigma_0 \times \exp(-\mu \times C_{\text{gen}})$, σ_0, μ are constant, and C_{gen} represents a current generation.

Step 5: Reject worst nest phase

In this phase, the most exceeding and terrible nest stays surreptitious, considering its probability values and subsequently makes new ones. The best solutions are positioned based on their fitness. From this point, the Lbest Cuckoo solution with maximum fitness values in each host is identified and the solutions spaces are explored using step size with Gaussian exponent function. At each iteration, the Lbest fitness value of Cuckoo is compared with a fitness value of Gbest solution obtained from the previous iterations. Then, the final Gbest solution is considered as an optimal solution.

Step 6: Stopping criterion phase

The system is rehashed until it achieves the end criteria. The result from the optimization process was based on the fitness value through which the recommendation is performed. This result helps to recommend the product to the user with improved accuracy.

The steps for finding an optimal solution using proposed MCS are given in Algorithm 1.

Table 1 Contingency table for RS

		Predicted result	
		Relevant	Irrelevant
Actual result	Recommended	True positive (TP)	False positive (FP)
	Not recommended	False negative (FN)	True negative (TN)

Algorithm 1 Proposed MCS

Let fitness function be P;
1: **begin**
2:    Produce initial population of n host nests $A_l (l = 1, 2, \ldots, n)$;
3:    **while** ($tc < maxgeneration$) **do** or Stopping condition
4:       Obtain a cuckoo l randomly using Gaussian exponent function;
5:       Estimate its fitness value P_l;
6:       Let $Gbest = Lbest = l$;
7:       Randomly choose nest m from n nests;
8:       **if** ($P_l > P_m$) **then return**
9:          Replace solution m by solution l;
10:          $Lbest = l$; // Exploitation: Lbest
11:          Host bird abandon its nest with probability (pa) and builds a new one via Gaussian exponent;
12:       **if** ($P_{Lbest} > P_{Gbest}$) **then return**
13:          $Gbest = Lbest$; // Exploration: Gbest
14:    Final $Gbest$ is taken as optimized result of MCS algorithm.
15: **end**

3.4 Performance measures

The performance of proposed RS is evaluated using the metrics precision, recall, F-measure and accuracy. These metrics are evaluated based on the entries obtained from Table 1.

The evaluation metrics used in the proposed system are explained next.

3.4.1 Precision

Precision is a percentage of recommended items that are relevant. The formula for precision is given in Eq. 11.

$$Precision = \frac{Number\ of\ correct\ recommendation\ relevant\ to\ the\ query}{Total\ number\ of\ recommendations} \tag{11}$$

3.4.2 Recall

The recall is a percentage of the relevant item that is recommended. The definition for a recall is given in Eq. 12.

$$Recall = \frac{Number\ of\ correct\ recommendation\ relevant\ to\ the\ query}{Total\ number\ of\ relevant\ recommendation} \tag{12}$$

3.4.3 F-measure

F-measure is the harmonic mean of precision and recall. It is defined in Eq. 13.

$$F\text{-}measure = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} \tag{13}$$

3.4.4 Accuracy

Accuracy measure is used to evaluate how the proposed algorithm recommends an item accurately. It is determined as the proportion of the number of users with correct recommendations and a total number of users considered for recommendation process. Equation 14 shows the formula for calculating accuracy.

$$Accuracy = \frac{Number\ of\ users\ with\ correct\ recommendations}{Total\ number\ of\ users} \tag{14}$$

4 Experimental results and analysis

In this section, the experimental analysis is performed in order to evaluate the performance of proposed RS by comparing the result with existing algorithms. Initially, the dataset is preprocessed as described in Sect. 3.1. Second, these processed datasets are given for clustering process to group similar users together. Third, the optimal users in each group are identified to provide satisfactory recommendations by using optimization algorithm. Finally, the standard benchmark functions are used to test the effectiveness of proposed optimization algorithm.

Benchmark MovieLens dataset was collected (see footnote 1) and used for experimental analysis of RS. The size of dataset is 100K, which includes 100,000 ratings provided by 1000 users for 1700 movies. The rating scale values vary from 1 to 5. In this study, 100 users are chosen randomly for proving the efficiency of proposed algorithm. The experimental analysis is carried out based on the proposed techniques as given in Algorithm 2.
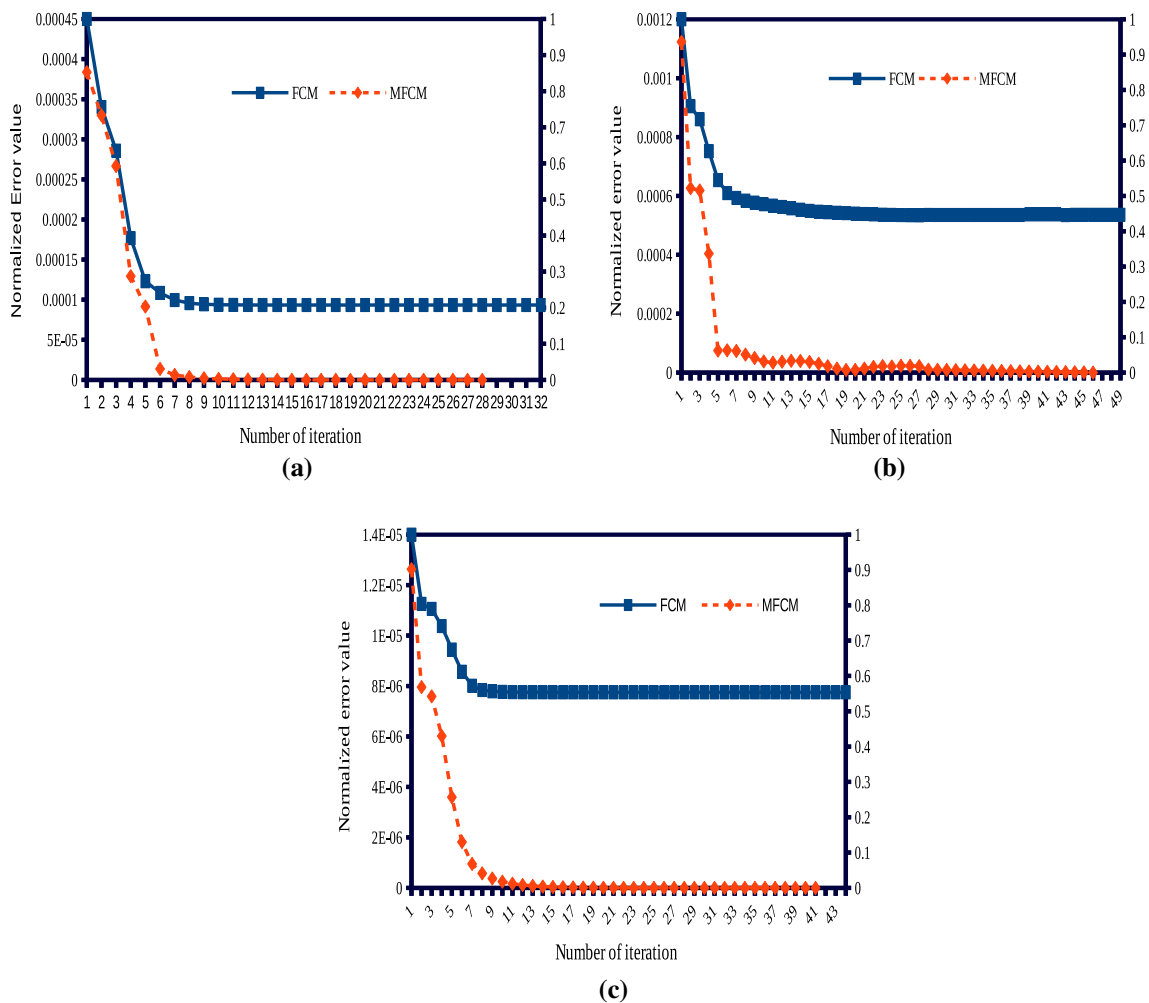
Fig. 2 Normalized error value for clustering techniques. **a** Iris dataset. **b** Glass dataset. **c** Diabetes dataset

Algorithm 2 Proposed RS

Let $[r]_{X \times Y}$ be the rating matrix
1: **begin**
2:     Collect $[r]_{X \times Y}$ dataset;
3:     Calculate $sim(m_p, m_q)$ using Eq. 1;
4:     Remove $min(sim(m_p) = \sum_{m_q \in Y} sim(m_p, m_q))$; Obtain reduced
    dataset.
5:     Initialize n cluster centroids randomly;
6:     **for** $k = 1$ to *MaxIteration* **do**
7:         Update η_{ij} using Eq. 5 and assign the data points to the cluster;
8:         Calculate the new c_j using Eq. 6;
9:         Calculate the new $Objective\_function_{MFCM}$ using Eq. 3;
10:         **if** $\| \eta_{ij}^{(k+1)} - \eta_{ij}^{(k)} \| < \varphi$ **then**
11:             break;
12:         **else**
13:             $\eta_{ij}^{(k)} = \eta_{ij}^{(k+1)}$
14:     **for** each cluster **do**
15:         Call MCS algorithm;
16:         Find *Gbest* user solutions and provide recommendations to
    other users;
17: **end**

4.1 Performance analysis of proposed clustering approach

At first, the dataset is collected and preprocessing is done to eliminate the movies which are having the less correlation with other movies. Then, these processed datasets are given to clustering approach. Clustering plays an important role in providing the recommendation with reduced error. In order to prove the proposed MFCM clustering approach that provides less error rate with few number of iteration, the analysis is performed for standard relational dataset, namely Iris, Class, and Diabetes.[2] The experimentation is done for proposed MFCM to compare with existing FCM method.

Figure 2 represents the error value comparison of clustering approaches FCM and proposed MFCM by referring the clustering parameters given in Table 2. The experimentation is conducted for 10 times, and average error values

_____

[2] Standard relational dataset. http://storm.cis.fordham.edu/~gweiss/data-mining/datasets.html.

Table 2 Control parameters considered for clustering and optimization algorithms

Clustering approach	Optimization algorithms		
MFCM	PSO	CS	MCS
Number of cluster: c	MaxIteration: 100	MaxIteration: 100	MaxIteration: 100
Fuzzy index: $m = 2$	Population size: 100	Population size: 100	Population size: 100
Stopping criteria: $\varphi = 0.3$	Acceleration constants: $C1 = 2$ $C2 = 2$	Discovery rate of alien solution: $p_a = 0.25$	Discovery rate of alien solution: $p_a = 0.25$
Maximum Iteration: *MaxIteration* : 100	Inertia weight: $W = 0.729$	Step size: 0.75	Step size: 0.75
		Levy exponent	Gaussian exponent

Table 3 Recommendation for different cluster sizes on proposed method

Number of cluster	Number of user	Correctly recommended	Wrongly recommended
2	100	71	29
3	100	73	27
4	100	74	26
5	100	75	25

Table 4 Overall F-measure value on different cluster sizes

Number of cluster	Precision	Recall	F-measure
2	0.71	1	0.83
3	0.73	1	0.84
4	0.74	1	0.85
5	0.75	1	0.86

are obtained. The error values are more in initial iterations and getting reduced in further iterations, finally stabilized to the certain error value. The error values are normalized into 0 to 1 scale rating. Figure 2a–c represents the normalized error values of FCM and MFCM clustering for Iris, Glass and Diabetes datasets, respectively. Primary y-axis in Fig. 2 represents the normalized error value of MFCM, whereas the secondary y-axis represents the normalized error value of FCM. In Fig. 2a–c, it is clearly understood the error value of proposed MFCM reduced drastically and converged in earlier iterations than FCM clustering. Therefore, MFCM clustering method is considered as the best clustering method and used for RS to group similar users together with reduced error. So, MFCM is considered to provide well-framed cluster since the objective of RS is focused towards improving the performance by reducing the error rate. As given in Algorithm 2, the MFCM clustering approach is applied to the preprocessed MovieLens dataset and the well-framed clusters are obtained.

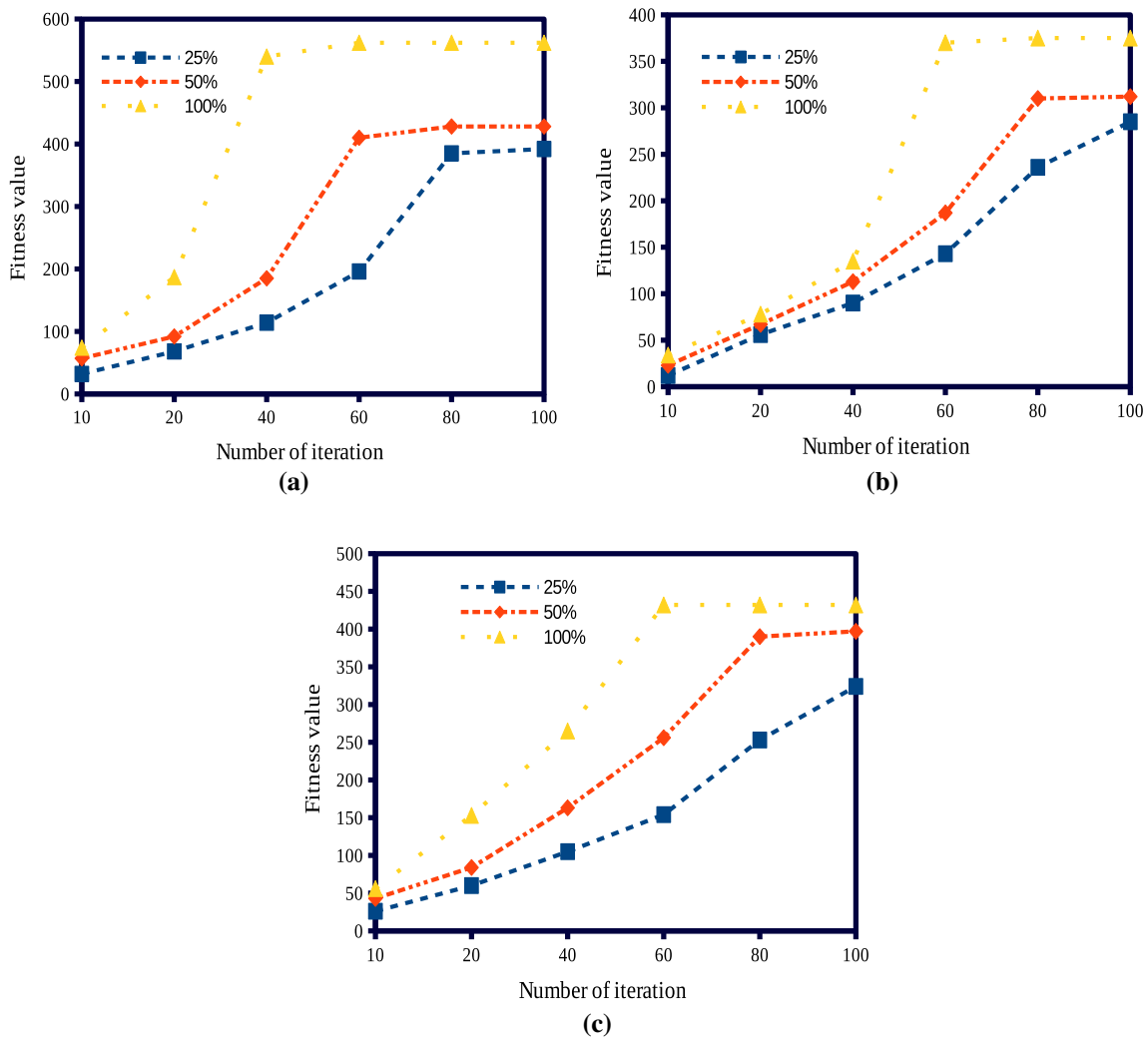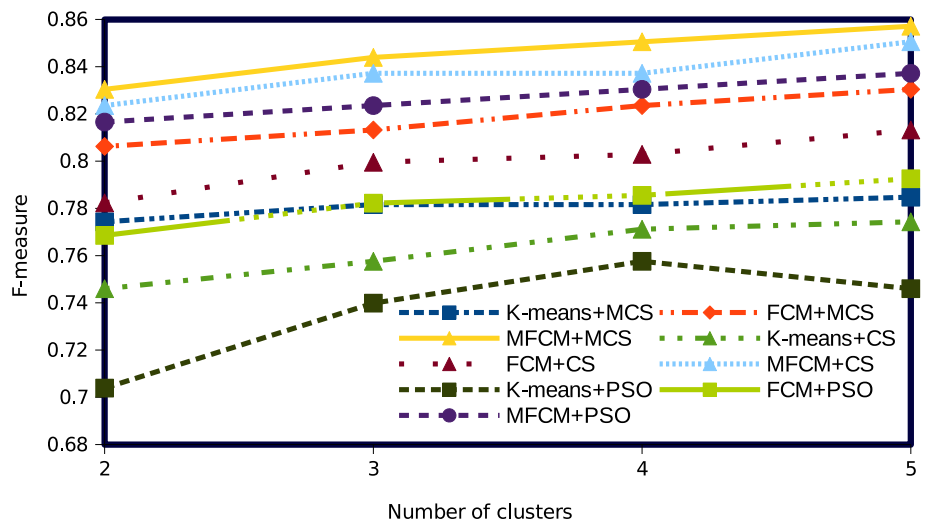## 4.2 Performance analysis of optimization algorithm

Clusters obtained from Sect. 4.1 are optimized using the proposed MCS optimization algorithm. Based on the obtained optimized user from each cluster, the recommendations are provided. In order to validate the performance of proposed MCS, it is compared with the existing PSO and CS. Table 2 shows the parameters to be considered for optimization of users in each cluster.

The performance of proposed RS framework is measured using various metrics such as precision, recall, F-measure and accuracy. To show the effectiveness of proposed method, it is compared with all combinations of existing clustering approaches k-means and FCM and optimization algorithms such as PSO and CS. The results are listed in tables and graphs to clarify an enhanced understanding of the relationship between various parameters such as the number of clusters, number of iterations and number of populations, which help to study the performance of the proposed approach. The recommendations for different cluster sizes for the MovieLens dataset are given in Table 3.

In Table 3, it is clearly understood that the number of correct recommendation is increasing when the cluster size increases. The recommendation results in Table 3 are used to find the precision and the recall value of the proposed system. A high value of precision indicates fewer users getting the wrong recommendation, and a higher recall value indicates the better recommendation for more users. Using these expressions, the F-measure is calculated for different cluster values under the number of iteration as 100 and 100% of population size. The performance results are listed in Table 4.

Figure 3 shows the overall F-measure of the proposed technique with the combination of existing clustering and optimization algorithms by fixing the number of iteration as 100 and 100% of population size. In Fig. 3, it is observed that the proposed MFCM with MCS combination performs better compared to other combinations of clustering and optimization algorithms. Also, it is inferred that the proposed

Fig. 3 *F*-measure for proposed and existing methods of clustering and optimization algorithms

Fig. 4 Fitness value for an optimization algorithms for different population sizes on different iterations. **a** MCS. **b** PSO. **c** CS
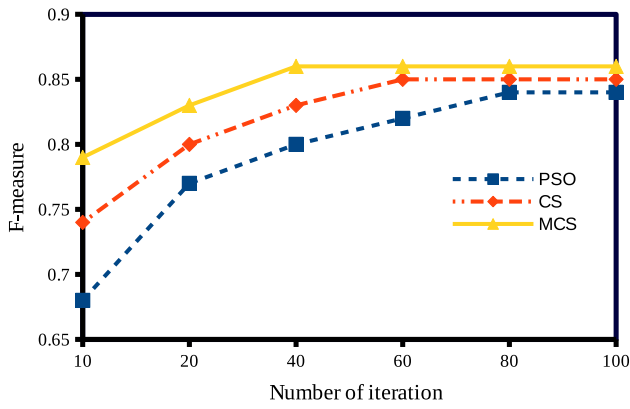
Fig. 5 *F*-measure values of optimization algorithms on different iterations

MFCM clustering yields better performance for all optimization algorithms.

Because of the random nature of optimization algorithm, the performance of RS cannot be judged by a single run. The experimentation should be conducted many times with the different values for each parameter to obtain good performance value. Also, the experiment analysis should be performed for larger population size to obtain a better result in the global space. Processing the data with larger population size may increase the computational time. With the different trials of experimentation, the parameters should be chosen to increase the performance of an algorithm to reduce computational effort. The performance of proposed system is compared with existing PSO and CS on 25, 50 and 100% of population sizes for iterations 10, 20, 40, 60, 80, and 100 using the fitness function given in Eq. 8.

Figure 4 shows the fitness value of proposed MCS, existing PSO and CS algorithms on varying population sizes and iterations. The MFCM clustering is performed by fixing the number of cluster size as 5. Figure 4a–c shows the changes in the fitness value for varying population sizes for MCS, PSO and CS, respectively. In Fig. 4, it is inferred that all the optimization algorithms perform better when the population size is high. That is, the fitness value is high for population size 100%, and also it is getting increased when the number
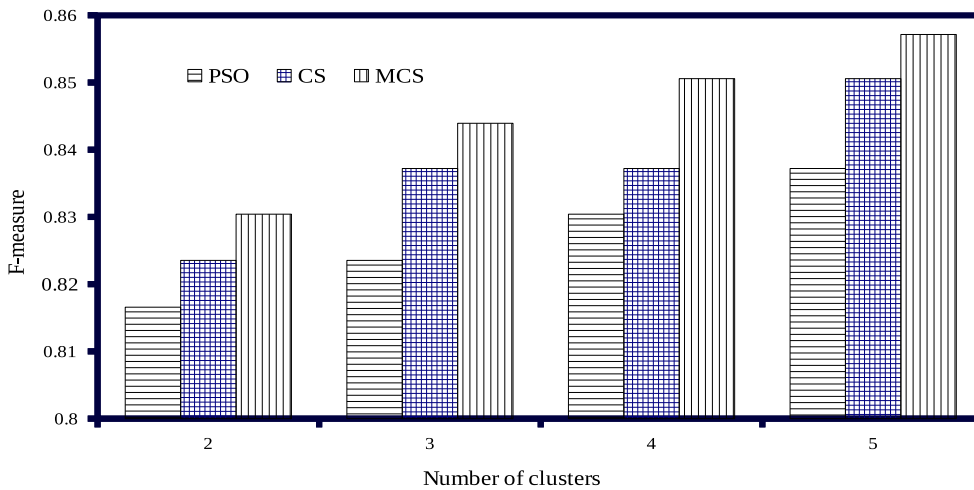
Fig. 6 *F*-measure for optimization algorithms using MFCM clustering approach

Fig. 7 Accuracy for proposed and existing methods of clustering and optimization algorithms
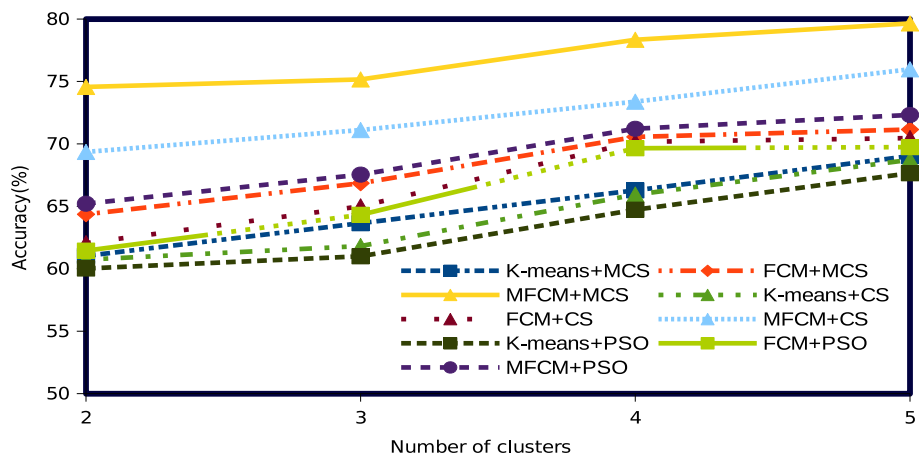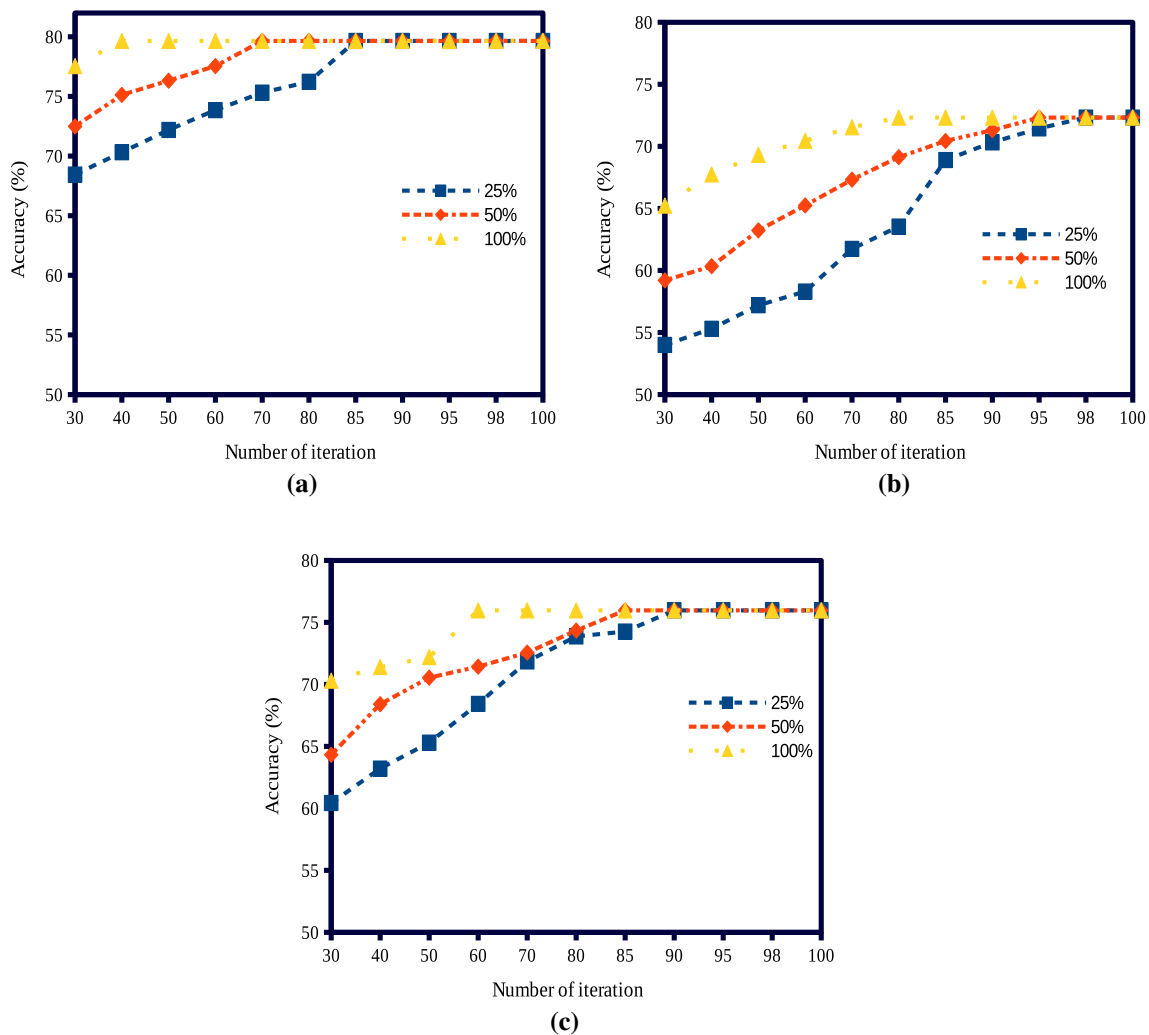
Fig. 8 Accuracy for an optimization algorithms for different population sizes on different iterations. **a** MCS. **b** PSO. **c** CS

of iterations is increased. In Fig. 4, it may be concluded that MCS is having highest fitness value than PSO and CS.

Figure 5 represents the F-measure value of optimization algorithms by varying the number of iterations. Here, the MCS is converged within 40 iterations, CS is between 60 and 70 iterations and PSO is between 75 and 80 iterations. In Fig. 5, it is inferred that MCS converged quickly when compared to other optimization algorithms.

The overall F-measure value of the proposed system with PSO and CS algorithm for MFCM clustering approach is shown in Fig. 6. A number of iteration are fixed as 100, and 100% of the population is considered for experimentation. The F-measure for proposed system is 0.86 for cluster size 5, 0.85 and 0.84 for existing PSO and CS, respectively. It is inferred from Fig. 6 that the performance of RS increases with increasing in number of cluster size. In Fig. 6, it is cleared that the overall performance of F-measure value for

the proposed system performs better than existing PSO and CS techniques.

Accuracy is an another performance metric that measures the overall performance of RS. The accuracy results for various combinations of proposed and existing clustering and optimization algorithms on various cluster sizes are shown in Fig. 7. For experimentation, the number of iteration is taken as 100 and 100% of the population is considered. In Fig. 7, it is concluded that the combination of proposed MFCM with all optimization algorithm performs better than other combinations. Figure 8 shows the accuracy value of optimization algorithms by varying population sizes. Figure 8a–c shows the change in accuracy on different population sizes for MCS, PSO and CS, respectively. It is inferred from Fig. 8 that the accuracy of the RS is increased when the population size is high. Also, the accuracy is increased for the greater number of iterations.
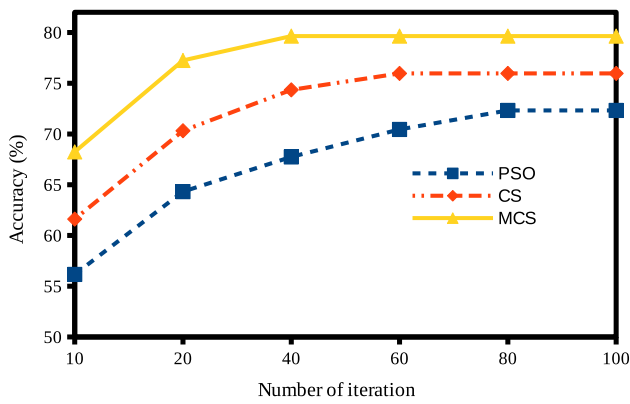
Fig. 9 Accuracy value for optimization algorithms on different iterations

Figure 9 represents the accuracy of optimization algorithms for cluster size of 5 with MFCM clustering and 100% of population size by varying number of iteration. The accuracy value is optimized for proposed MCS at 35 to 40th iterations, existing PSO at 55 to 60th iterations and CS at 75 to 80th iterations. The accuracy of proposed MCS converged earlier with improved value on less number of iterations due to search space exploration by Gaussian exponent, and it performs better than existing PSO and CS.

Figure 10 shows overall accuracy measure for MFCM clustering with existing and proposed optimization algorithms. The proposed method achieves the accuracy value 79.65%, whereas PSO and CS achieve 75.97 and 72.32%, respectively. Figure 10 shows that the overall accuracy is improved for the proposed method than the combination of existing optimization algorithms. It can be concluded that the performance of RS is increasing for increasing number of cluster size, population size and a number of iteration.
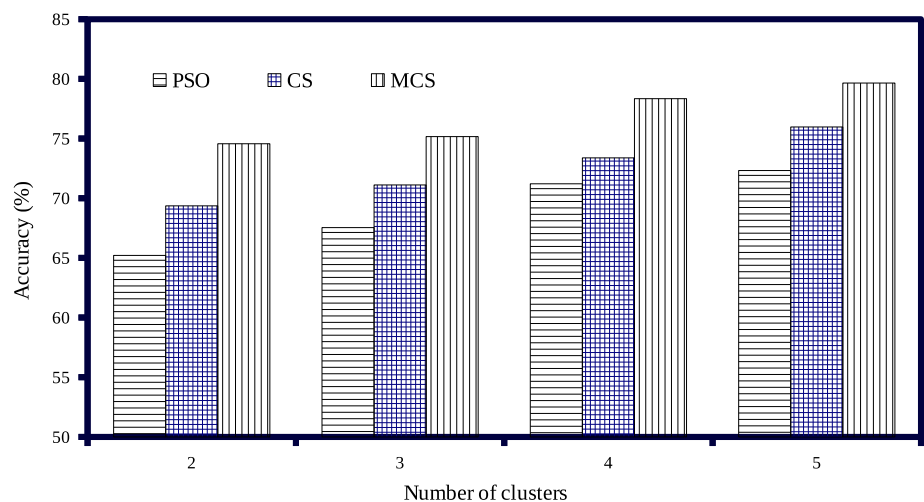
4.3 Performance analysis of proposed MCS with benchmark function

The performance of proposed MCS algorithm is tested with well-defined mathematical-based benchmark function. Among several benchmark function (Mernik et al. 2015), Rastrigin and Rosenbrock are used to test the algorithm for identifying global optimum.

Rastrigin is the multimodal function, meaning that it crosses through many local optimal solutions with one global optimal solution. If the search space is large, then the number of local optimal solutions is also more. A nonlinear Rastrigin function is represented by the following Eq. 15.

$$f_1(y) = \sum_{l=1}^{n} \left[y_l^2 - 10\cos(2\pi y_l) + 10 \right] \tag{15}$$

Rosenbrock is an unimodal function that has only one global maximum value. Equation 16 represents the Rosenbrock function.

$$f_2(y) = \sum_{l=1}^{n-1} \left[100 \left(y_{l+1} - y_l^2 \right)^2 + (y_l - 1)^2 \right] \tag{16}$$

These two benchmark functions are evaluated to get a global optimum solution. So, these functions are converted to a maximization function in order to support the problem considered for this study. The performance is evaluated based on a number of iteration and efficiency. Efficiency measures the success rate of an algorithm to obtain a global maximum solution in less number of iterations. The performance of proposed algorithm is tested on 100% of population size with 100 iterations and compared with other optimization algorithms PSO and CS. The performance of the algorithm is tested by referring the parameters given in Table 2. The performance result for benchmark functions are obtained by

Fig. 10 Accuracy of optimization algorithms using MFCM clustering approach

Table 5 Performance of standard benchmark functions

Function	Mean fitness function			Number of iterations		
	PSO	CS	MCS	PSO	CS	MCS
Rastrigin (f_1)	0.854	0.6568	0.2564	98	88	82
Rosenbrock (f_2)	0.31943	0.243819	0.19382	96	83	74
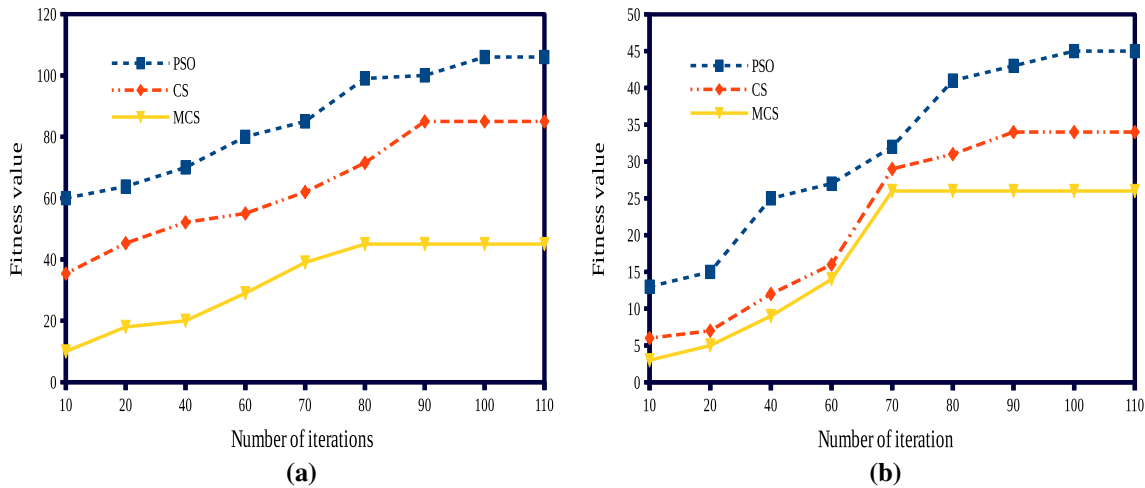
Fig. 11 Convergences of fitness value. **a** f_1 function. **b** f_2 function
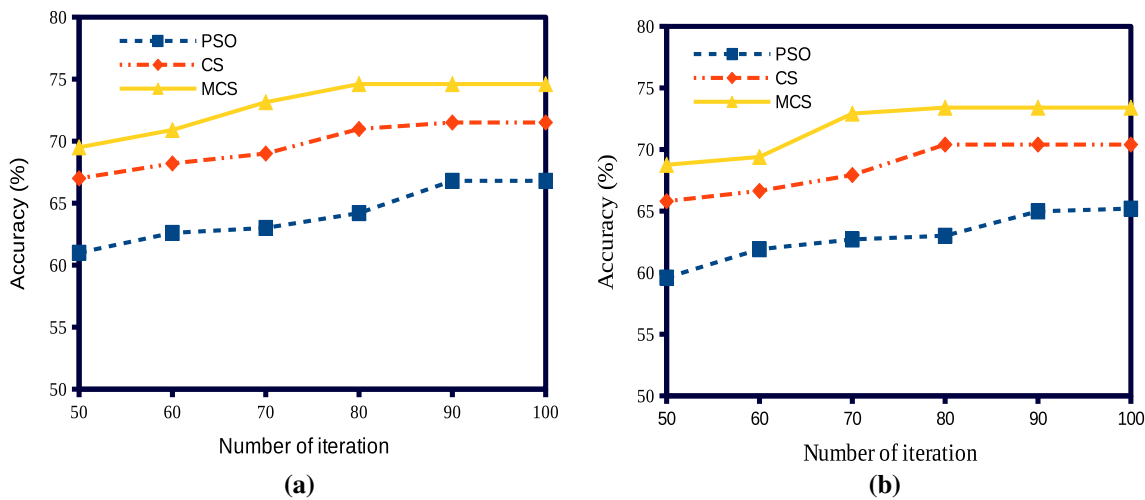
Fig. 12 Convergences of accuracy. **a** f_1 function. **b** f_2 function

converting the maximization function result back to minimization function.

Table 5 represents the mean fitness value of an optimization algorithm over 100 iterations and number of iterations needed to obtain a global optimum solution for 100% of population size. In Table 5, it is inferred that the proposed MCS attains minimum fitness value and converged with less number of iterations than PSO and CS.

Figure 11 shows the convergence of fitness value on PSO, CS and MCS for the test function f_1 and f_2. Figure 11a, b shows the fitness value of optimization algorithm on dif-ferent iterations for f_1 and f_2, respectively. In Fig. 11, it is concluded that the fitness value of MCS converged at 80th iteration for f_1 and 70th for f_2. But other optimization algorithm takes longer iteration to reach an optimal value.

Figure 12 shows the convergence of accuracy on PSO, CS and MCS for the test functions f_1 and f_2. Figure 12a, b shows the accuracy of optimization algorithm on different iterations for f_1 and f_2, respectively. In Fig. 12, it is concluded that the 74.6% of accuracy on MCS converged at 80th iteration for f_1 and 73% of accuracy 70th for f_2. But other optimization algorithms take longer iteration to reach an optimal value,

and also the obtained accuracy for PSO and CS is less than the MCS.

The experimental analysis of benchmark function shows that the proposed MCS optimization algorithm applied on clustered data points obtained from proposed MFCM clustering performs better than other optimization algorithms.

5 Conclusion

This study proposed a new recommender system (RS) that originates the CF approach. The data points are clustered with minimal error rate using a proposed modified fuzzy *c*-means (MFCM) clustering approach. Data points in the cluster are further optimized using proposed modified cuckoo search (MCS) algorithm, which achieves the global optimal solution with less number of iterations. Using the benefits of MCS with MFCM, the proposed RS reduces the recommendation error rate drastically and provides an accurate recommendation. The performance of proposed RS is evaluated using MovieLens dataset, and obtained results show effective when compared with popular optimization algorithms. Also, the superiority of proposed MCS algorithm is tested with standard benchmark optimization functions.

Compliance with ethical standards

Conflicts of interest We declare that we have no conflict of interest.

References

Al Mamunur Rashid SKL, Karypis G, Riedl J (2006) ClustKNN: a highly scalable hybrid model-& memory-based cf algorithm. In: Proceeding of WebKDD

Alam S, Dobbie G, Riddle P, Koh YS (2012) Hierarchical PSO clustering based recommender system. In: 2012 IEEE congress on evolutionary computation (CEC). IEEE, pp 1–8

Ameli K, Alfi A, Aghaebrahimi M (2016) A fuzzy discrete harmony search algorithm applied to annual cost reduction in radial distribution systems. Eng Optim 48(9):1529–1549

Ar Y, Bostanci E (2016) A genetic algorithm solution to the collaborative filtering problem. Expert Syst Appl 61:122–128

Arab A, Alfi A (2015) An adaptive gradient descent-based local search in memetic algorithm applied to optimal controller design. Inf Sci 299:117–142

Banati H, Mehta S (2010) Memetic collaborative filtering based recommender system. In: 2010 second Vaagdevi international conference on information technology for real world problems (VCON). IEEE, pp 102–107

Bezdek JC (1981) Cluster validity. In: Pattern recognition with fuzzy objective function algorithms. Advanced applications in pattern recognition. Springer, Boston, MA pp 95–154

Bilge A, Polat H (2013) A scalable privacy-preserving recommendation scheme via bisecting k-means clustering. Inf Process Manag 49(4):912–927

Bobadilla J, Ortega F, Hernando A, Alcalá J (2011) Improving collaborative filtering recommender system results and performance using genetic algorithms. Knowl Based Syst 24(8):1310–1316

Bodyanskiy YV, Tyshchenko OK, Kopaliani DS (2017) An evolving connectionist system for data stream fuzzy clustering and its online learning. Neurocomputing 262:41–56

Braida F, Mello CE, Pasinato MB, Zimbrão G (2015) Transforming collaborative filtering into supervised learning. Expert Syst Appl 42(10):4733–4742

Brouwer RK, Groenwold A (2010) Modified fuzzy c-means for ordinal valued attributes with particle swarm for optimization. Fuzzy Sets Syst 161(13):1774–1789

Demir GN, Uyar AŞ, Gündüz-Öğüdücü Ş (2010) Multiobjective evolutionary clustering of web user sessions: a case study in web page recommendation. Soft Comput 14(6):579–597

Dunn JC (1973) A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. J Cybern 3:32–57

Fang K, Liu C-Y (2003) Recommendation system using fuzzy c-means clustering. In: Book of information technology and organizations: trends, issues, challenges and solutions, vol 1. Idea group publishing, p 137–139

Ghosh S, Dubey SK (2013) Comparative analysis of k-means and fuzzy c-means algorithms. Int J Adv Comput Sci Appl 4(4):35–39

Guo G, Zhang J, Yorke-Smith N (2015) Leveraging multiviews of trust and similarity to enhance clustering-based recommender systems. Knowl Based Syst 74:14–27

Gupta A, Shivhare H, Sharma S (2015) Recommender system using fuzzy c-means clustering and genetic algorithm based weighted similarity measure. In: 2015 international conference on computer, communication and control (IC4). IEEE, pp 1–8

Hatami M, Pashazadeh S (2014) Improving results and performance of collaborative filtering-based recommender systems using cuckoo optimization algorithm. Int J Comput Appl 88(16):46–51

Jie L, Dianshuang W, Mao M, Wang W, Zhang G (2015) Recommender system application developments: a survey. Decis Support Syst 74:12–32

Kanungo T, Mount DM, Netanyahu NS, Piatko CD, Silverman R, Wu AY (2002) An efficient k-means clustering algorithm: analysis and implementation. IEEE Trans Pattern Anal Mach Intell 24(7):881–892

Katarya R, Verma OP (2016) A collaborative recommender system enhanced with particle swarm optimization technique. Multimed Tools Appl 75(15):9225–9239

Katarya R, Verma OP (2016) An effective collaborative movie recommender system with cuckoo search. Egypt Inform J 18:105–112

Katarya R, Verma OP (2016) An effective web page recommender system with fuzzy c-mean clustering. Multimed Tools Appl 76:21481–21496

Kim H-T, Kim E, Lee J-H, Ahn CW (2010) A recommender system based on genetic algorithm for music data. In: 2010 2nd international conference on computer engineering and technology (ICCET), vol 6. IEEE, pp V6–414

Kim K, Ahn H (2008) A recommender system using GA k-means clustering in an online shopping market. Expert Syst Appl 34(2):1200–1209

Koohi H, Kiani K (2016) User based collaborative filtering using fuzzy c-means. Measurement 91:134–139

Li Q, Kim BM (2003) Clustering approach for hybrid recommender system. In: Proceedings. IEEE/WIC international conference on web intelligence, 2003. WI 2003. IEEE, pp 33–38

Li Y, Shen Y (2010) An automatic fuzzy c-means algorithm for image segmentation. Soft Comput 14(2):123–128

Linden G, Smith B, York J (2003) Amazon. com recommendations: item-to-item collaborative filtering. IEEE Internet Comput 7(1):76–80

Liu X, Fu H (2014) PSO-based support vector machine with cuckoo search technique for clinical disease diagnoses. Sci World J 2014:1–7. https://doi.org/10.1155/2014/548483

Lops P, De Gemmis M, Semeraro G (2011) Content-based recommender systems: state of the art and trends. In: Recommender systems handbook. Springer, pp 73–105

Merialdo AK-B (1999) Clustering for collaborative filtering applications. Intell Image Process Data Anal Inf Retr 3:199

Mernik M, Liu S-H, Karaboga D, Črepinšek M (2015) On clarifying misconceptions when comparing variants of the artificial bee colony algorithm by offering a new implementation. Inf Sci 291:115–127

Mousavi Y, Alfi A (2015) A memetic algorithm applied to trajectory control by tuning of fractional order proportional-integral-derivative controllers. Appl Soft Comput 36:599–617

Mukhopadhyay A, Maulik U, Bandyopadhyay S (2015) A survey of multiobjective evolutionary clustering. ACM Comput Surv (CSUR) 47(4):61

Nasser S, Alkhaldi R, Vert G (2006) A modified fuzzy k-means clustering using expectation maximization. In: 2006 IEEE international conference on fuzzy systems. IEEE, pp 231–235

Nilashi M, Jannach D, bin Ibrahim O, Ithnin N (2015) Clustering-and regression-based multi-criteria collaborative filtering with incremental updates. Inf Sci 293:235–250

Pahnehkolaei SMA, Alfi A, Sadollah A, Kim JH (2017) Gradient-based water cycle algorithm with evaporation rate applied to chaos suppression. Appl Soft Comput 53:420–440

Patra BK, Launonen R, Ollikainen V, Nandi S (2015) A new similarity measure using Bhattacharyya coefficient for collaborative filtering in sparse data. Knowl Based Syst 82:163–177

Raja NSM, Vishnupriya R (2016) Kapurs entropy and cuckoo search algorithm assisted segmentation and analysis of RGB images. Indian J Sci Technol 9(17):1–6

Roy S, Chaudhuri SS (2013) Cuckoo search algorithm using Lévy flight: a review. Int J Mod Educ Comput Sci 5(12):10

Sarwar BM, Karypis G, Konstan J, Riedl J (2002) Recommender systems for large-scale e-commerce: scalable neighborhood formation using clustering. In: Proceedings of the fifth international conference on computer and information technology, vol 1

Thong NT et al (2015) HIFCF: an effective hybrid model between picture fuzzy clustering and intuitionistic fuzzy recommender systems for medical diagnosis. Expert Syst Appl 42(7):3682–3701

Tran CD, Dao TT, Vo VS, Nguyen TT (2015) Economic load dispatch with multiple fuel options and valve point effect using cuckoo search algorithm with different distributions. Int J Hybrid Inf Technol 8(1):305–316

Tsai C-F, Hung C (2012) Cluster ensembles in collaborative filtering recommendation. Appl Soft Comput 12(4):1417–1425

Ujjin S, Bentley PJ (2003) Particle swarm optimization recommender system. In: Proceedings of the 2003 IEEE swarm intelligence symposium, 2003. SIS'03. . IEEE, pp 124–131

Wasid M, Kant V (2015) A particle swarm approach to collaborative filtering based recommender systems through fuzzy features. Procedia Comput Sci 54:440–448

Wen Q, Celebi ME (2011) Hard versus fuzzy c-means clustering for color quantization. EURASIP J Adv Signal Process 1:118

Wu J, Li T (2008) A modified fuzzy c-means algorithm for collaborative filtering. In: Proceedings of the 2nd KDD workshop on large-scale recommender systems and the Netflix Prize competition. ACM, p 2

Wu K-L, Yang M-S (2002) Alternative c-means clustering algorithms. Pattern Recognit 35(10):2267–2278

Xue G-R, Lin C, Yang Q, Xi WS, Zeng H-J, Yu Y, Chen Z (2005) Scalable collaborative filtering using cluster-based smoothing. In: Proceedings of the 28th annual international ACM SIGIR conference on research and development in information retrieval. ACM, pp 114–121

Yang X-S, Deb S (2009) Cuckoo search via Lévy flights. In: World congress on nature & biologically inspired computing, 2009. NaBIC 2009. IEEE, pp 210–214

Yang X-S, Deb S (2014) Cuckoo search: recent advances and applications. Neural Comput Appl 24(1):169–174

Zahra S, Ghazanfar MA, Khalid A, Azam MA, Naeem U, Prugel-Bennett A (2015) Novel centroid selection approaches for kmeans-clustering based recommender systems. Inf Sci 320:156–189

Zanardi V (2011) Addressing the cold start problem in tag-based recommender systems. Ph.D. thesis, UCL (University College London)

Zanardi V, Capra L (2011) A scalable tag-based recommender system for new users of the social web. In: Database and expert systems applications, vol 6860. Springer, pp 542–557

Zhang R, Bao H, Sun H, Wang Y, Liu X (2016) Recommender systems based on ranking performance optimization. Front Comput Sci 10(2):270–280