ECTI Transactions on Computer and Information Technology
Journal homepage: https://ph01.tci-thaijo.org/index.php/ecticit/
Published by the ECTI Association, Thailand, ISSN: 2286-9131

Collaborative Movie Recommendation System using Enhanced Fuzzy C-Means Clustering with Dove Swarm Optimization Algorithm

Saranya S[1] and C. Jeyalakshmi[2]

ABSTRACT

Recommender Systems (RSs) aid in filtering information seeking to envisage user and item ratings, primarily from huge data to recommend the likes. Movie RSs offer a scheme to help users categorize them based on comparable interests. This enables RSs to be a dominant part of websites and e-commerce applications. This paper proposes an optimized RS for movies, primarily aiming to suggest an RS by clustering data and Computational Intelligence (CI). Unsupervised clustering, a model-based Collaborative Filtering (CF) category, is preferred as it offers simple and practical recommendations. Nevertheless, it involves an increased error rate and consumes more iterations for converging. Enhanced Fuzzy C-Means (EFCM) clustering is proposed to handle these issues. Dove Swarm Optimisation Algorithm (DSOA)-based RS is proposed for optimising Data Points (DPs) in every cluster, providing efficient recommendations. The performance of the proposed EFCM-DSOA-based RS is analysed by performing an experimental study on benchmarked MovieLens Dataset. To ensure the efficiency of the proposed EFCM-DSOA-based RS, the outcomes are compared with EFCM-Particle Swarm Optimization (EFCM-PSO) and EFCM-Cuckoo Search (EFCM-CS) based on standard optimization functions. The proposed EFCM-DSOA-based RS offers improved F-measure, Accuracy, and Fitness convergence.

Article information:

Keywords: Recommender Systems, MovieLens Dataset, Data Points, Collaborative Filtering, Enhanced Fuzzy C-Means Dove Swarm Optimisation Algorithm

Article history:
Received: January 7, 2023
Revised: May 1, 2023
Accepted: June 24, 2023
Published: July 22, 2023 (Online)

DOI: 10.37936/ecti-cit.2023173.251272

1. INTRODUCTION

The progress of technology and the development of the populace make huge use of informal social websites, resulting in increased growth of web content. The huge diversity of web content confuses online users for making the right decisions. A personalized Recommender System (RS) is an efficient tool for dealing with this information overload issue (Linden et al. 2003). RS aids in retrieving information by effectively handling a huge quantity of online data. It offers recommendations to users to make correct decisions on items and services. It aids in accumulating information based on user preferences for movies, tourism, TV shows, shopping, and taxi, either implicitly or explicitly.

Explicit ratings are statistical ratings given to the RS model, while implicit ratings include user opinions in text. RS can be employed to envisage users' purchase behaviours based on users' liking-based information and recommends products to improve their profit. Algorithms predict user preferences by interacting with users to acquire data and provide recommendations. The growth of RS depends on a recommendation as well as agencies like Amazon, Netflix, etc.,

The chief classes of RS include Collaborative Filtering (CF), Content-Based Filtering (CBF) as well as Hybrid Filtering (HF) schemes. CF is extensively utilised for providing suggestions by analysing items or user ratings. It helps in generating recommendations to users by rating similarity. CBF matches the features of items bought earlier with the sug-

[1] The author is with Computer Science and Engineering, K.Ramakrishnan College of Engineering, India 621112, E-mail: saranyas.cse@krce.ac.in
[2] The author is with Electronics and Communication Engineering, K.Ramakrishnan College of Engineering, India 621112, E-mail: lakshmikrce.2016@gmail.com

gestions by analysing items or user ratings. It helps in generating recommendations to users by rating similarity. CBF matches the features of items bought earlier with the suggested items' features (Lops *et al.* 2011). In cases where matches are determined amid item features, the conforming item is suggested to the user. CF-based model is not dependent on any domain and offers improved accuracy compared to CBF. If items are not obtainable, CF endorses the item to the user depending on the user's feedback. Owing to easiness, efficiency, and capacity to offer precise and adapted recommendations, CF is a dominant technique in RS.

CF is classified into memory as well as model-based schemes. In the case of memory-based CF, primarily, the similarity between the target and users is computed. In addition, neighbouring comparable users are found depending on computed similarity. Target items are forecasted and recognized adjoining users consequently offer recommendations. CF scheme based on memory is categorized into user and item-based schemes. User-based scheme forecasts the item value based on the neighbours of target users, while in the case of an item-based scheme, prediction is carried out based on the item closer to the target user. In the case of CF based on memory, the whole rating matrix is stored in memory to compute user similarity. This leads to problems related to scalability involving increased computational time.

To handle the above problem, CF based on the model is favoured. This scheme uses supervised and unsupervised methods to learn from the training rating matrix of items. The learned model is employed for forecasting the rating of target user items. Once the construction of the model is over, the user-item matrix becomes unnecessary, thus enabling prediction to be performed offline and quicker, even with the increase in the number of users and items. Much research is carried out by using model-based supervised learning class labels. Thus, the RS model's accuracy depends on the labeled training dataset (Braida *et al.* 2015). These methods can be combined with unsupervised clustering for RS-based applications to deal with this. This groups comparable DPs and determines patterns for unlabelled DPs.

The K-Means algorithm based on partition t is extensively used to assign DPs to clusters with increased similarity (Xue et al. 2005; Al Mamunur Rashid *et al.* 2006). DPs are assigned to a cluster (Ghosh & Dubey 2013). DPs may be in 2 or more clusters. It is determined using Fuzzy Membership Function (FMF) that is employed in Fuzzy C-Means (FCM) clustering. Several types of research are performed using FCM and its variants (Nasser et al. 2006; Wu and Li 2008;). The key drawback of FCM is that it offers a significantly increased error rate involving the increased quantity of iterations to obtain effective clusters. To address the abovementioned issues, this work propounds Enhanced FCM (EFCM) clustering, which decreases the error rate and significantly increases recommendation accuracy.

After forming effective clusters, ideal DPs in every cluster must be determined, and recommendations should be made. DPs in every cluster are found using optimization algorithms. Some optimization algorithms include Genetic Algorithm (GA) (Ar & Bostanci 2016), Particle Swarm Optimization (PSO), Water Cycle Algorithm (WCA) (Pahnehkolaei *et al.* 2017), Memetic Algorithm (MA) (Arab & Alfi 2015; Mousavi & Alfi 2015), Artificial Bee Colony (ABC) (Mernik *et al.* 2015), Harmony Search (HS) (Ameli *et al.* 2016), Cuckoo Search (CS) (Liu & Fu 2014; Raja & Vishnupriya 2016).

Online collective movie recommendations aid users in accessing their favourite movies by accurately determining similar neighbours amid users or movies from historical public ratings. The following are the contributions of this paper:

- Enhanced Fuzzy C-Means-Dove Swarm Optimisation Algorithm (EFCM-DSOA)-based RS that improves recommendation performance is propounded.
- Enhanced Fuzzy C-Means with Dove Swarm Optimization provides a unique solution to the movie recommendations problem. It combines content-based and item-based collaborative filtering techniques to achieve an optimal and personalized movie recommendation for the MovieLens dataset.
- The efficiency of the proposed algorithm is compared with EFCM-PSO and EFCM-CS. This paper uses CF-based similarity measure to preprocess MovieLens Dataset by eliminating dissimilar movies.
- EFCM clustering of the processed dataset is performed to obtain clusters with lessened error.
- The efficacy of the proposed clustering scheme is assessed. The optimum user in every cluster is efficiently determined by using DSOA techniques.

2. RELATED WORKS

Owing to data sparsity, neighbour selection becomes tedious with the increasing number of movies and users. Wang et al. (2014) have propounded a hybrid model-based movie RS that uses enhanced K-means clustering with Genetic Algorithm (GA) to split modified user space. It uses Principal Component Analysis (PCA) to reduce data to remove the sparsity of population space, thus reducing the computation complexity. The proposed model is applied on MovieLens Dataset and offers improved accuracy with reliable and personalised recommendations in contrast to prevalent methods.

Soares & Viana (2015) have compared the performance of collaborative algorithms based on contents for diverse metadata elements. The influence of information on genre, a whole collection of elements or actors/directors on quality, in addition to recommendation accuracy, is examined. The model is applied on Netflix and MovieLens Datasets, and performance is analyzed based on Mean Average Error (MAE) besides Precision. As there is no common metadata mechanism, the granularity of diverse elements varies with applications, and it is vital to determine how these variances influence the resulting quality. The use of various stages of sub-elements impacts the quality of RS.

Katarya & Verma (2016) have designed a hybrid model for movie RS that uses a type division scheme and classifies movies based on users, reducing computation complexity. K-Means offers primary parameters to PSO, thereby improving performance. It also provides primary seed and improves FCM aiding in soft rather than strict clustering of items in K-Means. Type division is applied to reduce dense multi-dimensional data space. The model used for MovieLens Dataset offers better performance in veracity, delivering predictable and personalized recommendations.

Katarya & Verma (2017) developed an RS that applies K-Means and CS on MovieLens Dataset. The approach is systematically explained, and outcomes are discussed. The performance is measured using Standard Deviation (SD), MAE, Root Mean Square Error (RMSE), and t-value for movie RS. The proposed model offers improved reliability, efficiency, and precise recommendations for MovieLens Dataset.

Singh & Solanki (2019) have propounded the movie RS using data clustering and a nature-stimulated scheme. K-means clustering is extensively used owing to its simplicity, thus dealing with huge quantities of data with reduced running time. It drops into local optimum due to arbitrarily produced primary centroids. It achieves a globally optimal solution if integrated with the nature-stimulated algorithm. K-Means is combined with bat, CS, MCS, and firefly algorithms and applied on MovieLens Dataset. The proposed K-Means MCS offers better outcomes in contrast to other algorithms.

Cui et al. (2020) provide users with accurate and fast recommendations with time changes. A recommendation model is proposed depending on Time Correlation Co-Efficient (TCC) with enhanced K-Means with CS (CSK-means)—clustering groups users for quick and precise recommendations. Furthermore, an efficient and personalised RS based on preference pattern is proposed. It offers improved recommendations by investigating user's behaviours. The investigations are performed on MovieLens as well as Douban datasets. The proposed model provides improved precision when compared to the proposed MCoC model.

Kumar & Prabhu (2020) have proposed a cooperative movie RS which involves K-Means clustering with Ant Colony Optimisation (ACO-KM) for the movie dataset. The proposed RS offers better results based on Precision, MSE, accuracy, and Recall. The propounded scheme offers improved outcomes based on speed for MovieLens Dataset. It provides improved scalability as well as efficiency by lessening cold start issues.

Roy & Dutta (2022) have designed an Improved Deep Ensemble Learning Model (ID-ELM) for various RSs related to varying application-based datasets. Primarily, datasets of diverse applications are collected from benchmarked sources, and data is split into numerous groups. Data is pre-processed to make it suitable by stop word removal, punctuation removal, and stemming. Feature extraction is performed using the Continuous Bag of Words Model (CBOW), and they are reduced using PCA. The features are passed on to ID-ELM, wherein an optimised Convolutional Neural Network (CNN) is used for extracting essential features from the pooling layer. The Fully Connected (FC) layer is substituted by a collection of classifiers called Neural Network (NN), AdaBoost, along with Logistic Regression (LR). Lastly, Ensemble Learning (EL) model is ranked depending on reviews, which extends recommendation results. Optimised CNN is combined with Adaptive Seeking Range-based Cat Swarm Optimisation (ASR-CSO) to achieve improved results.

3. PROPOSED METHODOLOGY

This paper proposes an Enhanced Fuzzy C-Means-Dove Swarm Optimisation Algorithm (EFCM-DSOA)-based RS. DSOA is applied to offer an effectual recommendation to users using EFCM clustering. The steps followed for implementing the proposed Recommender System (RS) are shown in Figure 1

Fig.1: *Proposed RS.*

3.1 Data Gathering and Pre-processing

User ratings on movies are obtained from benchmarked MovieLens Dataset. Movies with varying characteristics are to be eliminated. As movies with similar features are clustered into groups, they will impact recommendation accuracy if non-similar ones are included. These movies that do not fit into clusters are called outliers and are to be eliminated before the clustering process. Movie-based CF technique is propounded to determine the similarity between movies (i.e.) proportion of multiplication and addition of user ratings of co-rated movies.

Let 'U' represent the collection of users and 'M' be the collection of movies.

$$A-\{U_1, U_2, U_3 \dots U_A\}$$
$$B-\{M_1, M_2, M_3 \dots M_B\}$$

Consider 'U_i' a random user in 'A' who rates movies 'M_a' and 'M_b' ($M_a, M_b \in B$). The Similarity Measure (S) is determined as shown below.

$$S(M_a, M_b) = \frac{\sum_{U \in M_a \bigcap M_b} R_U^{M_a} \times R_U^{M_b}}{\sum_{U \in M_a \bigcap M_b} R_U^{M_a} + R_U^{M_b}} \quad (1)$$

Where
$R_U^{M_a}$, $R_U^{M_b}$ -'U' s' rating on movies 'a' and 'b' respectively

Lastly, a movie with the least similarity (below 0.4) is disregarded from the dataset, and a reduced and pre-processed dataset is taken for processing.

Ordering of subsection headings uses Arabic numerals with decimal notation. Leave a single space between the order number and the sub-heading caption. As noted above before sub-heading captions, skip one line. The words in the sub-heading caption are font size #10 and the first letter in each word is capitalized. Texts in the following section follow the Body_text style. Leave one line after the body text if it is followed by another section heading or subsection heading. The position of a sub-heading caption and the body text is similar to the section heading.

3.2 Clustering based on EFCM

FCM-based clustering (Dunn 1973) was enhanced by Bezdek (1981). DPs have a high probability of being placed in every cluster in contrast to being placed only in one cluster as in conventional K-Means. It positions DPs in multi-dimensional space into a fixed amount of clusters. The main focus is determining the centroid that rises similarity amid DPs in a cluster.

Usually, DPs do not fit into a particular cluster. Instead, it has a chance of entering into more clusters. FCM involves a membership function that allocates DPs to numerous clusters with varying membership degrees (φ) in the range [0, 1]. The main aim of EFCM is error reduction, along with the objective error function shown in Equation 2.

$$Obj\_Fun_{FCN} = \sum_{i=1}^{l} \sum_{j=1}^{n} (\varphi_{ij})^m ||T_j - C_j||^2 \quad (2)$$

Where

$\varphi_{ij}\varphi_{ij}$ - Membership of 'i' to 'C_jC_j'
i - DP
C_jC_j - Centroid of a cluster
m - Fuzzy index in the range [1, ∞]

Firstly, 'C_j' is taken at random. Values of 'φ_{ij}' and 'C_j' are updated until the ideal error value is reached.

Every cluster scheme runs with several groups to determine efficient clustering, providing the most exceptional recommendation accuracy level. FCM is a recognised clustering scheme performing well in particular applications. Nevertheless, it faces a huge error rate with increased iterations to obtain well-formed clusters. This significantly drops the performance of RS. To deal with this issue, EFCM is proposed wherein the objective function is transformed using weighted mean distance (ω) as shown in Equation (3).

$$Obj\_Fun_{EFCN} = \sum_{i=1}^{l} \sum_{j=1}^{n} (\varphi_{ij})^m \frac{||T_j - C_j||^2}{\omega_j} \quad (3)$$

Where
ω_j - Weighted mean distance for cluster 'j'

$$\omega_j = \left(\frac{\sum_{j=1}^{n} \varphi_{ij}^m . \frac{||T_j - C_j||^2}{\rho_j}}{\sum_{j=1}^{n} \varphi_{ij}^m} \right)^{\frac{1}{2}} \quad (4)$$

The parameter 'ω' significantly drops the error rate and the number of iterations. The membership function 'φ_{ij}' is given by,

$$\varphi_{ij} = \frac{1}{\sum_{k=1}^{n} \left(\frac{||T_j - C_j||}{||T_i - C_k||} \right)^{\frac{2}{m-1}}} \quad (5)$$

'C_j' is modified as shown in Equation (6).

$$C_j = \frac{\sum_{i=1}^{n} \varphi_{ij}^m \times T_i}{\sum_{i=1}^{n} \varphi_{ij}^m}, \forall j = 1, 2, \dots n \quad (6)$$

The values of 'φ_{ij}' and 'C_j' are repeatedly computed until the minimum error for the '$Obj_F un_{EFCM}$' is reached or membership variation amid iterations is less than the threshold of sensitivity (τ). Equation (7) signifies stopping criterion depending on membership values amid successive iterations 'm' and 'm + 1'.

$$\max_{ij} ||\varphi_{ij}^{m+1} - \varphi_{ij}^m|| < \tau \quad (7)$$

Where
τ - Termination criterion in the range [0, 1]

Lastly, DPs are grouped with minimum error using an enhanced EFCM scheme.

Algorithm: Proposed RS

$[R]_{X\times Y}$-Rating matrix

begine

Gather $[R]_{X\times Y}$ dataset

Compute $S(M_a, M_b)$ using Equation (1)

Remove $(S(m_a) = \sum_{m_a \in Y} S(M_a, M_b))$to get reduced dataset

Set random cluster centroids (n)

for (m=1 to MaxIter), do

Modify 'φ_{ij}' using Equation (5) and assign DPs to cluster

Compute 'C_j' using Equation (6)

Compute new $Obj_{Fun_{EFCM}}$ using Equation (6)

if $(||\varphi_{ij}^{m+1} - \varphi_{ij}^{m}||)$ then

break

else

$\varphi_{ij}^{m} = \varphi_{ij}^{m+1}$

for (every cluster) do

Invoke DSOA algorithm

Determine 'G_{Best}' user solutions and offer user recommendations

end.

3.3 Dove Swarm Optimization Algorithm (DSOA)

Doves hunt at places where crumbs are seen, and they search for them. Only some doves may get filled [26]. Unfulfilled doves hover in spots in search of crumbs. It is seen that the filled doves must have occupied spots with more amount of crumbs. The foraging behaviour of doves has inspired the proposal of an optimal algorithm. Let the optimised objective function be 'f(P) .'Every dataset data pattern (P) is considered a location containing crumbs. The quantity of crumbs in 'P' is given by 'f(P) .'The place with more quantity of crumbs is the best solution. The steps of DSOA are shown in Flowchart in Figure 2.

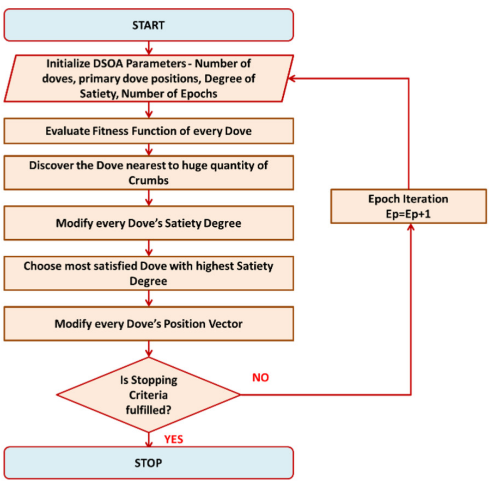


Fig.2: *DSOA.*

Step 1: Choose the number of doves and locate them in the solution space. Let 'N' be the number of doves. The doves are arbitrarily distributed in space. Nevertheless, they are deployed uniformly in rectangular regions.

Step 2: Initialise the amount of Epochs (Ep=0) and Satiety degree ($\mathrm{S}_{\mathrm{D}}^{\mathrm{E}}\mathrm{S}_{\mathrm{D}}^{\mathrm{E}}$) for dove 'D'. The Position Vector (PV) $\mathrm{P_D} \subset \mathrm{R^M} \mathrm{P_D} \subset \mathrm{R^M}$ of 'D' initialization can be performed in 2 ways.

The simplest way is to initialize '$\mathrm{W_D W_D}$' around the solution space randomly. The other way is to initialize the lattice initialized method. The steps are shown as follows: Two effectual weight initialization mechanisms are proposed to set weight vectors to quicken training for building a topologically ordered Feature Map (FM). An initialization scheme is designed for the algorithm. The smallest hyper-rectangle for parameter space includes valid values of factors represented as $[\mathrm{Lo_1,Up_1}]\ldots[\mathrm{Lo_M,Up_M}]$.

Where,

$\mathrm{Lo_a}$ and $\mathrm{Up_a}$ - Lower and upper bounds of dimension 'a' in the solution space

The basic idea of the proposed initialization mechanism is to reduce hyper-rectangle with n-dimensions into a 2-dimensional plane such that a 2-dimensional net may efficiently cover the solution space. For clarity, 'i' and 'j' index cells from 1 to X×Y. The steps are given below.

Step 2.1: Initialise cells on 4 corners. Also, initialise the weight vectors of 4 neurons on network corners as shown in Equations (8-11).

$$W_{1,1} = (Lo_1, Lo_2, \ldots Lo_M)^T \tag{8}$$

$$W_{X,Y} = (Up_1, Up_2, \ldots Up_M)^T \tag{9}$$

$$\mathrm{W}_{1,Y} = (Lo_1, Lo_2, \ldots Lo_{\lfloor \frac{M}{2} \rfloor}, Up_{\lfloor \frac{M}{2} \rfloor + 1}, \ldots Up_M)^T \tag{10}$$

$$\mathrm{W}_{X,1} = (Up_1, Up_2, \ldots Up_{\lfloor \frac{M}{2} \rfloor}, Lo_{\lfloor \frac{M}{2} \rfloor + 1}, \ldots Lo_M)^T \tag{11}$$

Step 2.2: Initialise cells on 4 edges in addition to cell value along edges as shown in Equation (12), j= 2,. . .Y-1.

$$\begin{aligned} W_{1,j} &= \frac{W_{1,Y} - W_{1,1}}{Y-1}(j-1) + W_{1,1} \\ &= \frac{j-1}{Y-1} \cdot W_{1,Y} + \frac{Y-j}{Y-1} W_{1,1} \end{aligned} \tag{12}$$

$$\begin{aligned} W_{X,j} &= \frac{W_{X,Y} - W_{X,1}}{Y-1}(j-1) + W_{X,1} \\ &= \frac{j-1}{Y-1} \cdot W_{X,Y} + \frac{Y-j}{Y-1} W_{X,1} \end{aligned} \tag{13}$$

$$\begin{aligned} W_{i,1} &= \frac{W_{X,1} - W_{1,1}}{X-1}(j-1) + W_{1,1} \\ &= \frac{i-1}{X-1} \cdot W_{X,1} + \frac{X-j}{X-1} W_{1,1} \end{aligned} \tag{14}$$

$$\begin{aligned} W_{i,Y} &= \frac{W_{X,Y} - W_{1,Y}}{X-1}(i-1) + W_{1,Y} \\ &= \frac{i-1}{X-1} \cdot W_{X,Y} + \frac{Y-i}{Y-1} W_{1,Y} \end{aligned} \tag{15}$$

Step 2.3: Initialise residual cells from top to bottom and left to right. Further, initialise weight vectors of 4 neurons on network corners. The description of initialization for residual neurons is given below.
for(j=2 to Y-1)
for(i=2 to X-1)

$$\begin{aligned} W_{i,j} &= \frac{W_{X,j} - W_{1,j}}{X-1}(i-1) + W_{1,j} \\ &= \frac{i-1}{X-1} \cdot W_{X,j} + \frac{X-i}{X-1} \cdot W_{1,j} \\ &= \frac{i-1}{X-1}\left(\frac{j-1}{Y-1} \cdot W_{X,Y} + \frac{Y-j}{Y-1} \cdot W_{X,1}\right) + \\ &\quad \frac{X-i}{X-1}\left(\frac{j-1}{Y-1} \cdot W_{1,Y} + \frac{Y-j}{Y-1} \cdot W_{1,1}\right) \\ &= \frac{((j-1)(i-1) \cdot W_{X,Y} + (j-1)(X-i) \cdot W_{1,Y} + (Y-j)(i-1) \cdot W_{X,1} + (Y-j)(X-i) \cdot W_{1,1})}{(Y-1)(X-1)} \end{aligned} \tag{16}$$

Different sizes are selected, and results are assessed. To assess training outcomes, neuron winners are counted, and the difference value is determined. Initially, the learning rate (L) is initialised to 0.1, and the reduction rate of 'L' is given by:

$$L_n = L_0 \times \left(1 - \frac{iter}{100}\right) = 0.1\left(1 - \frac{iter}{100}\right) \tag{17}$$

Where,
$iter$ - Iterative number
L_0- Initial learning rate

Step 3: Calculate fitness function f(W_j^{Ep}), j =1,. . .,N at Epoch (Ep) as the total quantity of crumbs at the 'D^{th}' dove location.

Step 4: Position 'D_j^{Ep}' near huge quantity of crumbs based on maximum criterion at 'Ep', j=1, 2, . . . N.

$$D_j^{Ep} = \arg\arg\max\{f(W_j^{Ep})\} \tag{18}$$

Step 5: Modify every dove's satiety level using the following equation, j=1, 2, . . . N.

$$S_j^{Ep} = \gamma \cdot S_j^{Ep-1} + e^{\left(f(W_j) - f(W_{D_f})\right)} \tag{19}$$

Step 6: Choose the most content dove, 'D_s^{Ep}' with maximum satiety degree using ensuing maximum criterion, j=1, 2, . . .N.

$$D_S^{Ep} = \arg \max_{1 \le j \le N} \left\{S_j^{Ep}\right\} \tag{20}$$

'D_S' chosen by Equation (20) presents the best foraging performance and may be followed by other doves.

Step 7: Modify every D's PV using the following maximum criterion.

$$W_j^{Ep+1} = W_j^{Ep} + L\beta_j^{Ep}(W_{d_s}^{Ep} - W_j^{Ep}) \tag{21}$$

Where

$$\vartheta_j^{Ep} = \left(\frac{S_{b_s}^{Ep} - S_j^{Ep}}{S_{b_s}^{Ep}}\right)\left(1 - \frac{||W_j^{Ep} - W_{D_s}^{Ep}||}{max\_Dist}\right) \tag{22}$$

$$\max\_Dist : \max_{1 \le j \le N} ||W_j - W_i||\,| \tag{23}$$

'L' is used for modifying D's PV. Equations (21)-(23) are detailed in the next step.

Step 8: Move to step 3 and raise the number of epochs by 1 (Ep=Ep+1) till the termination condition is satisfied. The termination condition is given below.

$$\left|f_{d_s}^{Ep} - T(Ep)\right| \le \in \ or \ Ep \ \le \ Max_E poch \tag{24}$$

DSOA has a complexity of order $O(N.N_D.Ep)$

Where,
N_D - Quantity of DPs
N - Quantity of doves
Ep - Quantity of epochs

If the best solution is to determine minimum ($\mathrm{W}_{\mathrm{j}}^{\mathrm{Ep}}$) ($\mathrm{W}_{\mathrm{j}}^{\mathrm{Ep}}$), then Equations (18) and (19) are modified as (25) and (26), j=1, 2, . . . N.

$$D_j^{Ep} = \arg \ \min\{f(W_j^{Ep})\} \quad (25)$$

$$S_j^e = \{\gamma \cdot S_j^{Ep-1} + e\left(f(W_j) - f(W_{D_f})\right),\ if\ f(W_{D_f}) \neq 0\gamma \cdot S_j^{Ep-1} + 1,\ if\ f(W_{D_f}^{Ep}) = 0 \quad (26)$$

For better understanding, the updating rules in Equations (21)-(23) are interpreted as shown below:

✓ An individual is impacted by the triumph of the best one in the flock and attempts to emulate its behaviour. Doves move in the direction of a dove with a maximum satiety degree to get more amount of food. This kind of social learning is exhibited by modifying PV ($\mathrm{W}_{\mathrm{j}}^{\mathrm{Ep}}$) to be like dove's PV with the highest satiety degree ($\mathrm{W}_{\mathrm{D_s}}^{\mathrm{Ep}}$).
i.e. $W_{\mathrm{j+1}}^{\mathrm{Ep}} = \mathrm{W}_{\mathrm{j}}^{\mathrm{Ep}} + \mathfrak{L} \cdot \vartheta\ _{\mathrm{j}}^{\mathrm{Ep}}(\mathrm{W}_{\mathrm{D_s}}^{\mathrm{Ep}} - \mathrm{W}_{\mathrm{j}}^{\mathrm{Ep}})$

✓ With increased satiety, the dove becomes conservative and hesitates to modify its current foraging policy. In contrast, a dove with reduced satiety degree desires to alter the current foraging policy and will be eager to emulate the nature of the best individual. The social impact is shown by making modifications comparative to first term value on RHS of Equation (22) $\left(\frac{S_{b_s}^{Ep} - S_j^{Ep}}{S_{b_s}^{Ep}}\right)$

- The social influence slowly declines with spreading; hence, the impact level is inversely proportional to the distance between the dove and the best dove in the flock. This type of social influence is imitated by adjusting quantity proportionate to the value of the third term on RHS of Equation (22)

$$\left(1 - \frac{||W_j^{Ep} - W_{D_s}^{Ep}||}{max\_Dist}\right)$$

4. RESULTS AND DISCUSSION

The performance of the proposed EFCM-DSOA-based RS is assessed based on F-measure, Accuracy, and Fitness convergence. The experimental investigation is performed to determine the proposed RS's performance compared to existing algorithms. Firstly, the dataset is pre-processed. Secondly, processed datasets are clustered to comparable group users. Using optimization algorithms, ideal users in every group are found to offer acceptable recommendations. Benchmark functions are used for testing the efficiency of the propounded algorithm. Experimental investigations are carried out on MovieLens Dataset [27]. The dataset is of size 100 KB that includes 1 lakh ratings offered by 1000 users for around 1700 movies. Scales are in the range 1 to 5. Around 1K users are selected arbitrarily to ensure the proposed algorithm's efficacy.

Precision: Precision is the percentage of recommended items that are relevant.

$$Precision = \frac{Number\ of\ correct\ recommendations\ relevant\ to\ the\ query}{Total\ number\ of\ recommendations} \quad (27)$$

- **Recall:** Recall is the percentage of relevant item that is recommended.

$$Recall = \frac{Number\ of\ correct\ recommendations\ relevant\ to\ the\ query}{Total\ number\ of\ relevant\ recommendations} \quad (28)$$

- **F-measure:** It is a harmonic mean of Precision as well as Recall.

$$F - measure = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} \quad (29)$$

- **Accuracy:** The ratio of users with the right recommendations to the total number of users taken for a recommendation.

$$Accuracy = \frac{Quantity\ of\ users\ with\ Right\ recommendations}{Total\ quantity\ of\ users} \quad (30)$$

The dataset is gathered and pre-processed to remove movies that are least correlated to other movies. The processed datasets are clustered. Clustering plays a significant role in offering recommendations with lessened errors. The proposed EFCM clustering scheme offers a reduced error rate for fewer iterations.

EFCM clustering is applied to pre-processed datasets, and well-formed clusters are obtained. Clusters are optimized using the proposed EFCM-DSOA optimization scheme. Depending on the acquired enhanced user from every cluster, recommendations are offered. The performance of EFCM-DSOA is compared with the present EFCM-PSO and EFCM-CS. Table 1 shows parameters for user optimization in every cluster.

Due to the arbitrary nature of algorithms used for optimisation, the performance of RS cannot be de-

Table 1: *Minimum Delay Model Parameter Values.*

Clustering Approach (EFCM)		**Optimization Algorithms (DSOA)**	
Number of cluster	c	Maximum Number of Iterations	100
Fuzzy index (m)	2	Size of Population	100
Stopping criteria(τ)	0.3	Acceleration Constants	C1 = 2 C2 = 2
Maximum Number of Iterations	100	Inertia Weight (W)	0.729
		Alien Solution's (pa) Discovery rate	0.25
		Step size	0.75
		Levy exponent	

termined in a run. Investigations are carried out several times for varying parameters to get better performance. Experimental analysis is performed for huge populations to get improved outcomes in global space. Processing data from a huge population may raise computational time. Parameters are chosen to enhance the algorithm's performance to decrease computational complexity. The performance is compared with EFCM-PSO and EFCM-CS for varying iterations based on a fitness function.

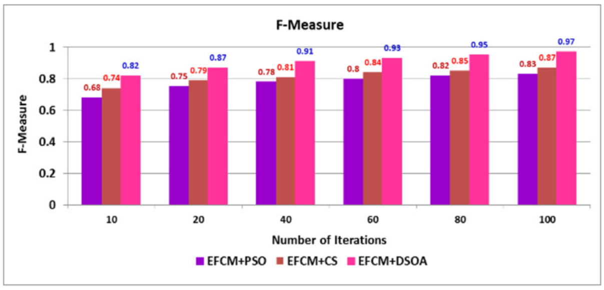


Fig.3: *F-measure of optimization Algorithm for different Iterations.*

Fig., 3 shows the F-measure of different algorithms optimizing varying iterations. EFCM-DSOA, EFCM-CS, and EFCM-PSO converge within 35, 60-70, and 75-80 iterations, respectively. EFCM-DSOA combines rapidly in contrast to other optimization schemes. EFCM-DSOA offers 17% and 11% better F-Measure in contrast to EFCM-PSO and EFCM-CS, respectively.

The F-measure of the proposed system with EFCM is illustrated in Figure 4. The number of iterations is considered to be 100. EFCM-DSOA offers 8% and 6.5% better F-Measure when compared to EFCM-PSO and EFCM-CS, respectively. The performance of RS improves with cluster size. The proposed scheme provides an improved value of F-measure in contrast to EFCM-PSO and EFCM-CS schemes.

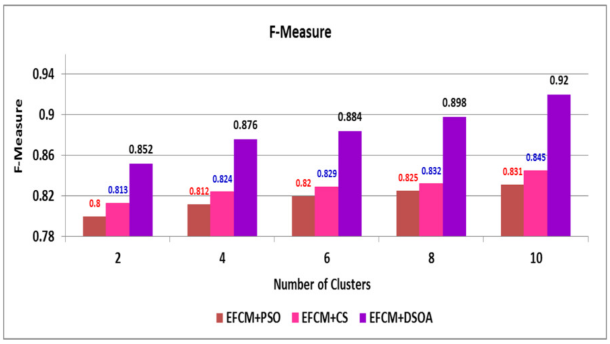


Fig.4: *F-measure of Optimisation Algorithms using EFCM.*

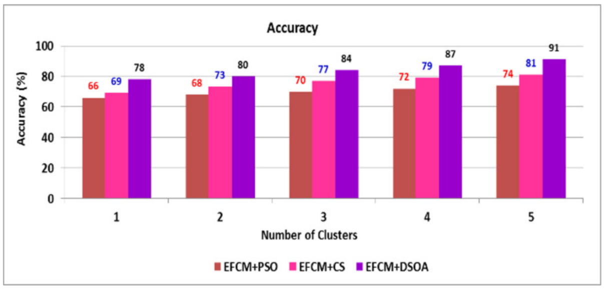


Fig.5: *Accuracy of Clustering and Optimisation Algorithms.*

Accuracy is determined for several combinations of clustering and optimization algorithms on diverse cluster sizes (Figure 5). The proposed EFCM offers better Accuracy in contrast to other combinations. EFCM-DSOA offers 14% and 8.2% better Accuracy than EFCM-PSO and EFCM-CS, respectively.

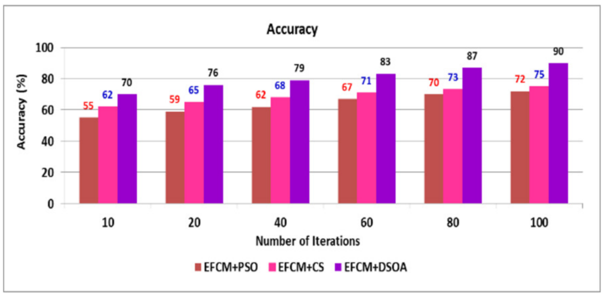


Fig.6: *Accuracy of optimization Algorithms for different iterations.*

Figure 6 represents the accuracy of algorithms with EFCM clustering for varying quantities of iterations for a cluster size of 10. Optimised accuracy is obtained for proposed EFCM-DSOA, existing EFCM-PSO, and EFCM-CS between the 40^{th} and 45^{th}, 60^{th} and 65^{th} iterations, and 80^{th} and 85^{th} iterations, respectively. The accuracy of EFCM-DSOA converges quickly with enhanced value on the reduced quantity of iterations owing to exploring search space by involving the Gaussian exponent. It offers better performance in contrast to existing EFCM-PSO and EFCM-CS algorithms. EFCM-DSOA offers 17% and 12.6% better accuracy when compared to EFCM-PSO and EFCM-CS, respectively.

The proposed EFCM-DSOA algorithm is analysed using several benchmark functions. Rastrigin and Rosenbrock are used for analysing and finding the global optimum.

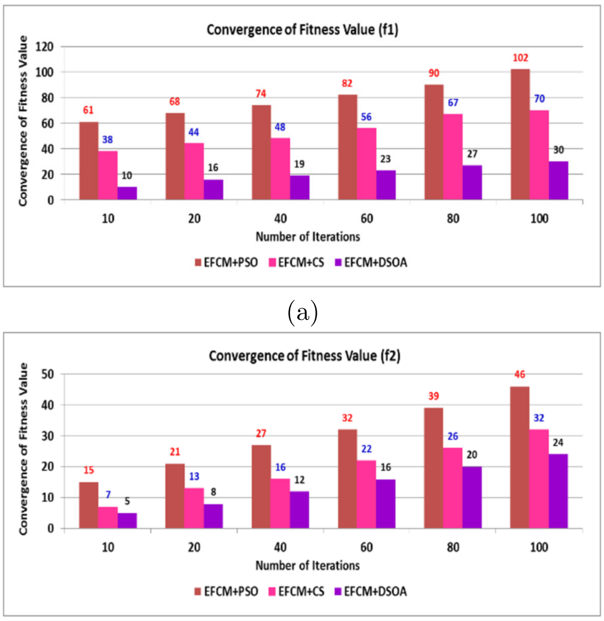
- **Rastrigin** is a multi-modal function that crosses several local optimal solutions with a globally optimal solution. In case of huge search space, there will be more quantity of locally optimal solutions. A non-linear function is shown in the following Equation (31).

$$f_1(y) = \sum_{l=1}^{n} [y_l^2 - 10\cos \ \cos(2\pi y_l) + 10] \qquad (31)$$

- **Rosenbrock:** It is a unimodal function with a global maximum value.

$$f_2(y) = \sum_{l=1}^{n-} [100(y_{l+1} - y_l^2)^2 + (y_l - l)^2] \qquad (32)$$

The benchmark functions are assessed to obtain a universal optimum solution. The functions are transformed into a maximization function. The performance is assessed depending on the quantity of iteration and efficiency. Efficiency gives the algorithm's success rate to get a universal maximum solution in fewer iterations. The benchmark functions are assessed by transforming the maximization to the minimization function.



(a)

(b)

Fig.7: *Fitness Convergences (a)f_1(b)f_2.*

Figure 7 shows fitness convergence on the algorithms for the 'f_1' and 'f_2' test functions. Figure 7(a), (b) shows the fitness for varying numbers of iterations for 'f_1' and 'f_2' correspondingly. The fitness convergence of EFCM-DSOA happens at the 82nd iteration for 'f_1' and the 75th for 'f_2'. It is seen that EFCM-DSOA on clustered DPs obtained from proposed clustering offers improved results in contrast to optimization algorithms. EFCM-DSOA offers 3.8 and 2.5 times better results for 'f_1' and 2.1 and 1.36 times improved results for 'f_2' in contrast to EFCM-PSO and EFCM-CS, respectively.

5. CONCLUSIONS

This paper proposes an Enhanced Fuzzy C-Means-Dove Swarm Optimisation Algorithm (EFCM-DSOA)-based RS. EFCM clustering is proposed to handle the challenges of CF. Data Points (DPs) in every cluster are optimised, offering efficient recommendations. It positions DPs in multi-dimensional space into a fixed amount of clusters. EFCM clustering is performed to obtain clusters with reduced error. The optimum user in every cluster is efficiently determined by using DSOA techniques. The performance is analysed by experimenting on MovieLens Dataset. EFCM-DSOA-based RS offers better F-measure, Accuracy, and Fitness convergence than EFCM-PSO and EFCM-CS-based on standard optimization functions.

References

[1] G. Linden, B. Smith and J. York, "Amazon.com recommendations: item-to-item collaborative filtering," in *IEEE Internet Computing*, vol. 7, no. 1, pp. 76-80, Jan.-Feb. 2003.

[2] P. Lops, M. D. Gemmis and G. Semeraro, "Content-based recommender systems: State of the art and trends," *Recommender systems handbook*, pp. 73-105, 2010.

[3] F. Braida, C. E. Mello, M. B. Pasinato and G. Zimbrão, "Transforming collaborative filtering into supervised learning," *Expert Systems with Applications*, vol. 42, no. 10, pp. 4733-4742, 2015.

[4] G.-R. Xue, C. Lin, Q. Yang, W. Xi, H. J. Zeng, Y. Yu and Z. Chen, "Scalable collaborative filtering using cluster-based smoothing," in *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 114-121, 2005.

[5] A. M. Rashid, Shyong K. Lam, Adam LaPitz, G. Karypis and J. Riedl, "ClustKNN: a highly scalable hybrid model-& memory-based CF algorithm," *8th International Workshop on Knowledge Discovery on the Web (WebKDD 2006)*, Philadelphia, Pennsylvania, USA, pp.1-10, 2006.

[6] S. Ghosh and S. K. Dubey, "Comparative analysis of k-means and fuzzy c-means algorithms,"

International Journal of Advanced Computer Science and Applications, vol. 4, no. 4, 2013.

[7] S. Nasser, R. Alkhaldi and G. Vert, "A Modified Fuzzy K-means Clustering using Expectation Maximization," *2006 IEEE International Conference on Fuzzy Systems*, Vancouver, BC, Canada, pp. 231-235, 2006.

[8] J. Wu and T. Li, "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*, no.2, pp. 1-4, 2008.

[9] Ar, Y., & Bostanci, E. (2016). A genetic algorithm solution to the collaborative filtering problem. Expert Systems with Applications, 61, 122-128.

[10] R. Katarya and O. P. Verma, "A collaborative recommender system enhanced with particle swarm optimization technique," *Multimedia Tools and Applications*, vol. 75, pp. 9225-9239, 2016.

[11] S. M. A. Pahnehkolaei, A. Alfi, A. Sadollah and J. H. Kim, "Gradient-based water cycle algorithm with evaporation rate applied to chaos suppression," *Applied Soft Computing*, vol. 53, pp. 420-440, 2017.

[12] A. Arab and A. Alfi, "An adaptive gradient descent-based local search in memetic algorithm applied to optimal controller design," *Information Sciences*, vol. 299, pp. 117-142, 2015.

[13] Y. Mousavi and A. Alfi, "A memetic algorithm applied to trajectory control by tuning of fractional order proportional-integral-derivative controllers," *Applied Soft Computing*, vol. 36, pp. 599-617, 2015.

[14] M. Mernik, S. H. Liu, D. Karaboga and M. Črepinšek, "On clarifying misconceptions when comparing variants of the artificial bee colony algorithm by offering a new implementation," *Information sciences*, vol. 291, pp. 115-127, 2015.

[15] K. Ameli, A. Alfi and M. Aghaebrahimi, "A fuzzy discrete harmony search algorithm applied to annual cost reduction in radial distribution systems," *Engineering Optimization*, vol. 48, no.9, pp. 1529-1549, 2016.

[16] X. Liu and H. Fu, "PSO-based support vector machine with Cuckoo search technique for clinical disease diagnoses," *The Scientific World Journal*, vol. 2014, 548483, 2014.

[17] N. S. M. Raja and R. Vishnupriya, "Kapur's entropy and cuckoo search algorithm assisted the segmentation and analysis of RGB Images," *Indian Journal of Science and Technology*, vol. 9, no. 17, 89936, 2016.

[18] Z. Wang, X. Yu, N. Feng and Z. Wang, 'An improved collaborative movie recommendation system using computational intelligence," *Journal of Visual Languages & Computing*, vol. 25, no. 6, pp. 667-675, 2014.

[19] M. Soares and P. Viana, "Tuning metadata for better movie content-based recommendation systems," *Multimedia Tools and Applications*, vol.74, pp. 7015-7036, 2015.

[20] R. Katarya and O. P. Verma, "An effective collaborative movie recommender system with cuckoo search," *Egyptian Informatics Journal*, vol. 18, no.2, pp. 105-112, 2017.

[21] S. P. Singh and S. Solanki, "A Movie Recommender System Using Modified Cuckoo Search," *Emerging Research in Electronics, Computer Science and Technology*,Springer, Singapore, 2019.

[22] Z. Cui et al., "Personalized Recommendation System Based on Collaborative Filtering for IoT Scenarios," in *IEEE Transactions on Services Computing*, vol. 13, no. 4, pp. 685-695, 1 July-Aug. 2020.

[23] M. S. Kumar and J. Prabhu, "A hybrid model collaborative movie recommendation system using K-means clustering with ant colony optimisation," *International Journal of Internet Technology and Secured Transactions*, vol. 10, no. 3, pp. 337-35, 2020.

[24] D. Roy and M. Dutta, "An Improved Cat Swarm Search-Based Deep Ensemble Learning Model for Group Recommender Systems," *Journal of Information & Knowledge Management*, vol. 21, no. 3, 2250032, 2022.

[25] M. -C. Su, J. -H. Chen, A. M. Utami, S. -C. Lin and H. -H. Wei, "Dove Swarm Optimization Algorithm," in *IEEE Access*, vol. 10, pp. 46690-46696, 2022.

[26] `https://grouplens.org/datasets/movielens/100k/`

Saranya S received the B.Tech. degree in Information Technology from Dhanalakshmi Srinivasan Engineering College in 2006.She received the Master degree in Software Engineering from Jayaram College of Engineering and Technology, Tamilnadu, India in 2012.She is currently working as an Assistant Professor in the Department of Computer Science and Engineering at K.Ramakrishnan College of Engineering, Trichy, India.Her area of interests includes Machine Learning,Webservices,Deep Learning and Internet of Things.She can be contacted at email: saranyas.cse@krce.ac.in.

C. Jeyalakshmi received B.E degree in Electronics and Communication Engineering from Bharathidasan University in 2002 and M.E. degree in Communication systems from Anna University, Chennai in 2008. She served as a faculty for 11 years in the Department of ECE,Trichy Engineering college, Tamilnadu. Currently she has been with K.Ramakrishnan college of Engg., where she is working as Professor in ECE department. She has obtained PhD degree from Anna University, Chennai in the field of Speech recognition of hearing impaired people in 2015. Her research interest also includes speech processing, Image processing, Machine learning. She has published 35 papers in Reputed International journals and presented papers in more than 10 International Conferences.She can be contacted at email: lakshmikrce.2016@gmail.com