Research Journal of Applied Science

MEDWELL PUBLICATIONS SINCE 2000

# Cuckoo Search Based Personalized View for Movie Recommendation over Social Networks

[1]S. Uma Shankari and [2]M. Chidambaram

[1]*Department of Computer Science, Bharathiar University, 641046 Tamil Nadu, Coimbatore, India*

[2]*Department of Computer Science, Rajah Serofiji College, Thanjavur, India*

**Key words:** Recommender system, movie, features, clustering, context and probabilistic matrix factorization

**Corresponding Author:**
S. Uma Shankari
*Department of Computer Science, Bharathiar University, 641046 Tamil Nadu, Coimbatore, India*

Page No.: 102-111
Volume: 15, Issue 3, 2020
ISSN: 1815-932x
Research Journal of Applied Sciences


**Abstract:** Owing to the exponential growth of information in online social networks, the users of such networks demand the recommendation systems to deliver significant results. A recommendation system rightly suggests the personalized movies that are desirable to the users predominantly from large information storage. Notably, the current research works in movie recommendation system focus on determining the most relevant features from the user profile information and shared contents in the social network. Even though the existing research works recommend the movies that are in proximity to the user preferences, there is a profound need for further exploring the features of the movie and thus ensure the highly desired movies to the users. Hence, this paper targets on recommending the movies with the knowledge of analyzing the movie features along with the data clustering and computational intelligence methods. This article proposes the Cuckoo search based MOst personalized VIEw in item recommendation (C-MOVIE) model, incorporating the missing rating prediction and contextual movie recommendation phases. At first, the C-MOVIE approach explores the features of the movies to recognize the interest of the users in terms of inherent features after reducing the feature dimensionality by Principal Component Analysis (PCA) method. Then, it clusters the users based on the recognized features by K-means clustering and Cuckoo search optimization methods with the intention of grouping the users with similar interests which eases the missing rating prediction when using Probabilistic Matrix Factorization (PMF). In the end, the C-MOVIE approach contextually recommends the movies to the users by mapping the features of the new movies with the features of the clustered users. The experimental results yielded on Douban movie which data set demonstrate that the C-MOVIE approach distinctively delivers the personalized movie recommendation than the existing HPSO method.

## INTRODUCTION

In recent years, Social networks have become increasingly popular Internet services among the people as their top option among the people. The reason being, the social networking sites are enabling the users to easily share the text, image, audio and video with their neighbors on the social network site. The popular social tagging sites include Delicious and micro-blogging sites such as Twitter, LinkedIn and Facebook. The social networking sites often have to deal with the information overload problem while providing the suggestions to the users. The explosive growth of the volume of information in the social networks urges the recommender systems to provide significant results. Recommender system[1] becomes an indispensable component in different e-commerce applications that attempt to suggest the items including music, books, movie, web pages, news and images according to the user interests. Recommending the desired items from analyzing the relevant items of the users is the ideal way to ease the process and enhance the user satisfaction level. Most recommender systems significantly utilize the rating information as the key source to identify the user preferences[2]. Recommendation system entirely relies on the historical rating information about the target user on other items and the rating information of their similar users or items. Many users are usually reluctant to provide the rating on the items. Hence the recommender systems have to confront with the data sparsity issue.

Several commercial systems such as Netflix.com, IMDb.com and Amazon.com predict the user preferences for new items based on the past rating information on other items. These systems employ the two types of recommendation algorithms such as content-based and collaborative filtering methods. The conventional recommendation algorithms[3] fail procedurally in finding the similar users or similar items due to the minimum density of the available ratings in user-item rating matrix. Also, the assumption of similar user's preferences based recommendation misleads the accuracy of the personalized movie recommendation[4]. Thus, there is a need for contextually analyzing the prior rating information becomes crucial even when there are missing rating values. Even though, the existing recommender systems exercise significant efforts on predicting the missing rating values, analyzing the several items rated by an individual adds an extra burden on the prediction system. Since, a single user usually is likely to rate a multitude of items which possibly alleviate the strong feasibility of identifying the user preferences. Thus, the C-MOVIE approach focuses on identifying the inherent features of each user behind the movie ratings and mitigating the number of features before analyzing the user-item rating matrix while predicting the missing rating value. Moreover, the C-MOVIE approach recommends the appropriate movies to the users by the contextual analysis using clustering and factorization algorithms.

The main contributions of the C-MOVIE approach are as follows. By analyzing which inherent features, the C-MOVIE approach predicts the missing movie ratings and enhances the personalized recommendation of movies to the social users. The C-MOVIE approach explores the desired features of the movies rated by the user and selects the optimal features of each movie for every user using the PCA method. By exploiting the k-means clustering and cuckoo search optimization algorithm, the C-MOVIE approach effectively clusters the social users based on the inherent features of each user. Instead of predicting the movie ratings from the similar social user preferences, the C-MOVIE along with the PMF method efficiently manages to predict the missing movie ratings from the neighbors in the clusters.

The C-MOVIE approach maps the optimal features of the all the users with the features of the movie to be recommended to determine a set of users interested in that specific movie which facilitates the highly desired movie recommendation. The experimental results show that the C-MOVIE approach outperforms the existing system by accomplishing a high accuracy of movie recommendation.

**Literature review:** Recommender systems are widely based on a variety of approaches that include content based[5], collaborative[6, 7] and hybrid filtering approach. Among the item recommendation, notably, movie recommendation systems employ the collaborative filtering and clustering methods. Collaborative filtering is being used as the most effective technique for the movie recommendation which is based on the nearest-neighbor mechanism.

**Missing rating prediction techniques:** Most recommendation systems are widely based on the probabilistic matrix factorization to improve the prediction accuracy and resolve the data sparsity issues. SoRec predicts the missing rating value and recommends the items by employing the PMF method which exploits the rating records and the social network information of the users. Rating prediction algorithm initially predicts the personalized utility of a specific item for each user and then, converts the utility into a rating value using the matrix factorization framework. An approach applies the PCA method on the demographic data and ratings of the users to improve the recommendation process in various aspects. Collaborative filtering based recommendation model[8] explores the bilateral role of the user interactions in the social network to allow one on one recommendation between the people. However, user interaction based recommendation is inappropriate in several areas such as community network system. Also, similar profile information and interaction based target user's interest identification is likely to mislead the personalized recommendation.

**Clustering based recommendation techniques:** In recommendation system, machine learning techniques play a crucial role in handling the enormous amount of data. The recommendation systems employ the different machine learning techniques such as decision tree, Naive Bayes classification and k-means clustering[9, 10]. In movie recommendation system, the existing efforts[11, 11] broadly focus on the clustering based techniques to resolve the scalability issue and facilitate the accurate recommendation. The main intention of clustering in the recommendation system is to group the like-minded users into a single cluster based on a specific feature rather than assuming the social network friends as the similar users. A genetic algorithm based clustering method[12] enables the overlapping clusters to ensure the personalized recommendation. k-Nearest Neighbor ( kNN) algorithm[13] based recommender system ensured the reliable and precise recommendation and used as the orientation algorithm in the collaborative filtering. Context-aware movie recommendation presents the two approaches that comprise the assessment of contextual factors and time for the hybrid recommender and the identification of the users in a household which submits a given rating using machine learning algorithms. Several recommender systems exploit the optimization algorithms to provide quality recommendations to the users. Hybrid recommendation approach[14] employs the Particle Swarm Optimization (PSO) based clustering (HPSO) in the movie recommendation system which develops a model for predicting the rating value with the help of ensemble supervised machine learning. Cuckoo search bio-inspired algorithm[15] has been used to facilitate the evolution of the continuing factors based quality recommendation which is accomplished by generating the dynamic variation based clusters within a reasonable time. Cuckoo Optimization based recommender system[16] obtains the optimized weight vector by generating the different weight vectors based on the interest of the users which ensures the significant improvement in the quality and performance of the recommender system. Even though there are numerous efforts on the movie recommendation systems, the approaches mentioned lack in precisely handling the data sparsity issues when dealing with the entire abundant data. Also, these approaches are likely to mislead the accurate movie recommendation due to their presumption of the social link based user preference identification.

**An overview of the c-movie approach:** The explosive growth of online information sharing in the Social networks rapidly escalates the importance of the social recommendation systems. To recommend the appropriate items to the social users, understanding the user's preferences on a specific item in two ways is necessary such as through explicit analysis and implicit analysis in which item refers shopping, movie, tourism and TV. Hence, the social recommender systems focus on exploring the explicit rating information on the user-item relationship. Several users are reluctant to provide the ratings on the items even though they have the interest on the items this in turn, leads the cold-start issue creating complexity in recognizing the users' preferences. To overcome this problem, the existing recommender systems employ the collaborative filtering method to predict the missing rating values based on the possible ratings of the similar users or items. Even though numerous social network based movie recommender systems are available, extracting the tweets regarding the movie rating prediction is arduous. Thus, the C-MOVIE approach targets on predicting the missing rating values on a specific item by exploiting the information from the similar user preferences and the inherent features of the rating value. Moreover, predicting the missing rating value from the vast amount of rating information is a challenging task. The C-MOVIE approach addresses this difficulty by mitigating the number of features while determining the rating value. The probability of the user preference prediction varies by many factors such as popularity, recent and personalized interests. For instance, most of the users provide high ratings for some popular movies such as The God Father, Avatar and Titanic than mediocre movies. Therefore, the social recommender system needs to consider the user's interest in each movie instead of utilizing the rating information of similar users. The C-MOVIE approach focuses on both the movie category and the similar users' preferences on a specific item to predict the missing rating values and consequently to recommend the desired items to the users. It employs both the explicit rating information and implicit features behind the rating information to predict the unknown ratings and manipulate the movie recommendation.

Figure 1 shows the overall process of the proposed C-MOVIE system. The proposed model incorporates the two phases such as selecting the optimal features based on the rating value and predicting the missing rating value based on the context features and recommending the movie to the users.

**Rating based optimal features selection:** The C-MOVIE approach applies the proposed algorithm on the influential user-item matrix regarding the which movie. It exploits the user-movie preference matrix using explicit

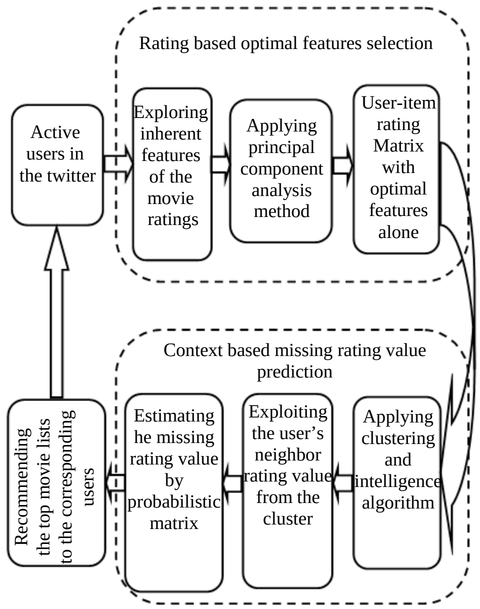


Fig. 1: An Intelligent movie recommendation model for the social network users

rating information and inherent feature information with the intention of identifying the user's preference on each movie. The inherent feature information on each item depends on the inherent analysis related to the contextual features of the movie such as music director, actor, director and so on which facilitates the accurate unknown rating prediction method. Then, it employs the Principal Component Analysis (PCA) to identify the optimal features alone. PCA based optimal feature selection eases the burden of processing the entire rating information while resolving the cold-start issue. It which results in the user-item matrix relationship that embraces a set of optimal features alone by ignoring the less important features.

**Context based movie recommendation:** The C-MOVIE approach applies the clustering and an intelligent algorithm to form the group with similar users with the comparable interests of a set of items. The C-MOVIE approach predicts the missing rating values based on the movie-specific influential user relations in the social network.

Moreover, it employs the probabilistic matrix factorization to predict the unknown rating of an item by a specific social user. The probabilistic model exploits only the movie-specific influential network and the corresponding rating information on the target item to precisely identify the missing rating values. Also, the C-MOVIE approach retains a set of social users with their similar preferences on a movie category. It dynamically assigns the weight to a list of movies for each user using the dynamic weighting scheme according to the rating values and the movies for each user are sorted out in descending order. Eventually, the C-MOVIE approach recommends a desired set of movies to the social user based on the weight of each movie.

## MATERIALS AND METHODS

**Rating based optimal feature selection:** The C-MOVIE approach employs the explicit rating values submitted by the users from the movie reviews to recommend it as their desired movies to them. Owing to the existence of numerous items rated by the users, the C-MOVIE approach targets to reduce the dimensionality of the rating values on the various movie items to ease the precise missing rating prediction. Initially, the C-MOVIE approach analyzes the inherent features behind the movie ratings to accurately identify the user preferences on movies. Then, it generates the weighted user-item relationship matrix comprising the features of the movie based scores using PCA method. Eventually, this phase retains the user-item relationship matrix only with a set of optimal features.

**Recognizing the inherent features behind the movie ratings:** The C-MOVIE approach targets to recognize the inherent features in the contents of movies by the rating values that refer to the content-based recommendation. $U_i \times M_j$ denotes the user-movie rating matrix in $U_i$ and $M_j$ represent the ith user and jth movie, i = {1, 2, ..., M} and j = {1, 2, ..., N}. To find the content of the features that is the primary intention to provide a high rating value on the movie, the C-MOVIE approach applies the similarity measure between the items rated by a user. The similarity measure is based on the content of the items such as actor, director and genre. Every item ($M_j$) comprises a feature vector ($f_K$) as numeric or nominal values. The C-MOVIE approach intends to identify the significant features of the items rather than computing the similarity between the two elements. The importance score or optimal feature selection process is based on the high feature weight that measures the similarity between the items. To identify the exact feature behind the preference, the C-MOVIE approach analyzes the reviews from the comments along with the rating value. The importance score of each feature his based on several factors such as a feature vector, explicit rating and the textual review information.

The C-MOVIE system applies the correlation measurement between the two movies through the features. It categorizes the rating parameters into two categories such as low and high rating value that represents the low rating values are 1 and 2 and high rating values are 3-5 out of 5 ratings. Initially, the

C-MOVIE approach explores the textual reviews by correlating the context of the keywords with a set of features of the item. For instance, 'movie' item comprises a set of features such as director, writer, cast, country, language and company. Let, $U_1$ provides 2 rating value for $M_3$ with the comment as 'the directing seems to pretentiou's. The C-MOVIE approach employs the WordNet ontology source to extract the relevant terms such as 'direction, director and 'directing' while mapping the comment with a set of features. By utilizing this information, the C-MOVIE approach maps the comment with a set of features to identify the importance of the movie's feature in the user point of view. According to this factor, the comment of the $U_1$ is mapped with a 'director' feature. In the resulting feature, the preference of the $U_1$ relies on the 'direction. Similarly, the C-MOVIE approach recognizes a set of inherent features of the movies rated by the users.

**Generating weighted matrix by PCA method:** The C-MOVIE model exploits the PCA method to select the optimal features with the intention of achieving the goal of dimension reduction. To construct a new matrix, the PCA method involves the following steps such as calculating the covariance matrix, computing eigenvalues and eigenvectors of the covariance matrix, selecting principal components and forming the new matrix. The C-MOVIE approach selects the high weighted feature as the principal components for each user.

Initially, the C-MOVIE approach computes the score of each feature corresponding to the user by analyzing the user's comments on the items. This computation is carried out by measuring the frequency of mapping features for each user. A feature has a high score when the user repeatedly discusses the review of the multiple items regarding that specific feature than the other features. As a result, the C-MOVIE approach computes the user's preference on each feature based on the occurrence of the review related to the feature among the rated items or movies. In the user-item rating matrix, there is a feasibility for numerous empty slots that denotes the missing ratings. To determine these missing ratings, analyzing all the features of the items is an arduous task. Hence, the C-MOVIE approach intends to select a minimum number of features with high importance to each user with the help of PCA method which assists to reduce the missing rating prediction time and improve the prediction accuracy.

$$W_f\left(U_i\right) = \sum_{j=1}^{n}\left(\frac{\left(M_j^i\right) * a\left(Re_f^i\right)}{Max\_R * n}\right) \quad (1)$$

In Eq. 1, R ($M^i_j$) refers the rating of the ith user on the jth item. Max_R represents the maximum rating value of the movie that is '5'. ɑ $(Re^i_f)_j$ denotes the review of the ith user regarding fth feature which either refers the value as 1 or 0. If the jth item has the review about the fth feature by the ith user, ɑ $(Re^i_f)_j$ is '1'; otherwise, ɑ $(Re^i_f)$ is '0'. To obtain the principal components, the C-MOVIE approach forms the covariance matrix and finds the eigenvalues and eigenvectors of the covariance matrix. By utilizing the score of each feature along with the PCA method, the C-MOVIE approach mitigates the number of features used in the missing rating prediction process for each user. The resulting, each user has the different number of features or principal components according to their preferences in the numerical ratings and the textual reviews. The number of principal components is less than the total number of features about the item. The C-MOVIE approach observes that the features have the stable and unstable weight values. For instance, a user has several positive weight features that are stable weight and several negative weight feature values referring unstable weight. The positive weight features alone are taken into account to obtain the recommendation for each user.

**Context-based movie recommendation:** The proposed movie recommendation relies on the presumption of the feature weights of the different users on the movies. Initially, the C-MOVIE approach clusters the users based on the weight of each feature to group the similar users with similar preferences using hybrid computation intelligence method. With the aim of increasing the robustness and improving the quality of movie recommendation, the C-MOVIE approach employs the K-means clustering and the cuckoo search optimization algorithm to cluster the users optimally. Then, it employs the PMF method to predict the missing rating value based on the target user's neighbors in the corresponding cluster. After that, it recommends a list of movies to the corresponding user based on the feature preferences.

**Clustering the users based on the features of the items and Predicting the missing rating values using PMF method:** The main objective of the proposed algorithm is to ensure the users with accurate like-minded neighbors fall into one cluster. The proposed clustering method facilitates the rating prediction of a specific movie based on the computation in a specific cluster instead of searching the whole user space. The C-MOVIE approach improves the performance of the personalized movie recommendation through the k-means data clustering and cuckoo search optimization algorithms which averts the premature convergence while determining the optimal solutions. Initially, it performs k-means clustering on the input user-item rating information in which clustering is based on the feature weight of the users. From this point forward, the k-means clustering algorithm randomly selects the centroid points of the feature weights and clusters the users into different clusters. The C-MOVIE approach intends to determine the similar users by measuring the similarity through the Euclidean distance measure between the users. Every user has either one or more than one feature in their preference lists. The k-means clustering method iteratively clusters the users and inspects the users residing in the clusters one by one by exploiting the feature weights and the centroid points. It only retains the users in the cluster, if the distance between the users is minimum otherwise, it adds the user to another cluster which has the closest distance than the current one. However, there is the possibility of mismatching of the users with clusters. Hence, the C-MOVIE approach matches the feature weight of each user with its current cluster mean and also with other cluster's mean to relocate the users based on the minimum distance with the cluster mean. Equation 2 computes the cluster score of each user ($U_i$) on the cth cluster based on the closeness between the $U_i$ and centroid point through all the features:

$$\text{Cluster\_score}\left(U_i\right) = \sum_{f=1}^{n}\left(\left|W_f\left(U_i\right) - W_f\left(CP_c\right)\right|^2\right)^{1/2} \qquad (2)$$

Where $W_f(U_i)$ refers the feature weight of the ith user and $W_f(CP_c)$ denotes the centroid point of cth cluster. Then, the C-MOVIE approach employs the cuckoo search optimization algorithm to optimize the clustering results which are generated by the k-means clustering algorithm. It reforms the clusters in terms of changing the previous centroid points based on the fitness function and retains the number of users in the cluster:

$$\text{Fitness function}\left(U_i\right) = \sum_{N(U_i)_c=1}^{N_c}\left(W_f\left(U_i\right) - \left(\frac{W_f\left(N\left(U_i\right)_c\right)}{N_c}\right)\right) \qquad (3)$$

where, $N(U_i)_c$ denotes the neighbors of the ith user in the cth cluster and $N_c$ represents the total number of neighbors in a specific cth cluster. If the fitness function has the minimum value than the previous centroid point based fitness score, the C-MOVIE approach replaces that specific user as the centroid points in that cluster. By applying Eq. 2 and 3, the C-MOVIE approach clusters the users which are based on the preferences regarding main features such as direction, cast and music and then, divides the clusters into sub-clusters such as the person participating in the categories if direction, cast and music. For instance, Harrison Ford, Brad Pitt and Engin Günaydın are the sub-features of the cast and James Cameron and Woody Allen are the sub-features of the director feature. By utilizing the resultant of the clusters and the PMF, the C-MOVIE approach predicts the missing rating value of ith user on the jth movie is computed by the following Eq. 4. Equation 4 finds the rating value for the target user ($U_{iT}$) on a specific movie:

$$R\left(M_j\right) = \left[1/2\right] * \left(\left(\sum_{f=1}^{k}\left[\frac{W_f}{k}\right] * \sum_{i=1}^{m}\left[\frac{R\left(U_i^f\right)}{m}\right]\right) + \sum_{j=1}^{r}\left[\frac{R\left(M_j^t\right)}{r}\right]\right) \qquad (4)$$

Where $R(M_j)$ refers the rating of the jth movie. $W_f$ denotes the weighted score of a kth feature of the ith user. 'k' represents the number of features which have high preferences of the ith user who has the missing rating value on the jth movie. 'm' denotes the number of users those belonging to kth cluster on the jth movie. R ($U^f_i$) indicates the rating values of the ith user on a jth movie on the fth feature. 'r' refers the number of rated items by the ith user and R ($M^r_j$) denotes the rating value of rth movies by the ith user:

$$p\left(R \mid U, V, \sigma^2\right) = \prod_{i=1}^{m}\prod_{j=1}^{n} N\left(R_{ij} \mid U_i^T M_j, \sigma^2\right)^{I_{ij}} \qquad (5)$$

The C-MOVIE approach predicts the missing rating value by submitting the Eq. 4 according to the Eq. 5 which denotes the probabilistic matrix factorization. The multiplication factor of feature based user and movie rating value assists in determining the probability score of the rating value.

**Recommending the movies to the corresponding users:** The C-MOVIE approach recommends the appropriate movie to the clustered users by exploring the features included in every upcoming movie. To determine a set of users for a new movie recommendation, the C-MOVIE approach maps the features of a new movie with the features tagged on the every sub-clusters. The sub-cluster comprises a set of users under several features, for instance, James Cameron and Brad Pitt. The C-MOVIE approach applies the n-gram similarity measurement between a set of features about each new movie ($M^n_j$) and the features of each sub-cluster ($S^k_C$) is defined in Eq. 6. It selects the corresponding clustered users for each movie based on a highly correlated score of the cluster. The high matching score is based on the frequency of n-gram mapping with respect to the features. The C-MOVIE approach takes into the consideration of n-gram mapping in binary levels such as either '0' or '1' to improve the accuracy of mapping:

$$\text{Sim}\left(M_j^n, S_C^k\right) = \sum_{f=1}^{M_k}\text{avg}\left(R_U\left(S_C\right)\right)_f * \left(\frac{2 * \left|\text{n-grams}\left(M_j^n\right)_f \bigcap \text{n-grams}\left(S_C\left(U_f\right)\right)\right|}{\left|\text{n-grams}\left(M_j^n\right)_f + \text{n-grams}\left(S_C\right)_f\right|}\right) \qquad (6)$$

where, $M_k$ represents the number of features in jth new movie and $U_f$ denotes the feature of any movie is rated by the user in the sub-cluster. avg $(R_U(S_C))_f$ is the average rating value of all the users in fth sub-cluster in users provide the rating on any movie with the interest of fth feature. The resulting, the C-MOVIE approach recommends that new movie to the users who are all involving in either one or more than one clusters, based on the top feature similarity score which reflects the interest or preference on a new movie ($M^n_j$) by the clustered users. Algorithm 1 illustrates the steps involved in the C-MOVIE approach

**Algorithm 1:** C-MOVIE algorithm

**Input:** Social user reviews on movies
**Output:** Recommended movies
**//Rating based optimal features selection**
Let, User i = Hub; user-movie rating matrix, $U_i \times M_j$; $O_K$ЄK
**while** U ={$U_1$, $U_2$, ..., $U_i$}; M={$M_1$, $M_2$, ..., $M_j$}; f ={$f_1$, $f_2$, ..., $f_K$} **do**
**for** each user 'i' and movie 'j' **do**
**for** each feature 'K' **do**
Compute $W_f(U_i)$ for all features using equation (1)
Select the optimal features($O_K$) using PCA, i.e., $O_K$ <K
$O_K(U_i)$ is based on the $W_f(U_i)$
**endfor**
**endfor**
**endwhile**
**//Context based movie recommendation**
**while** U ={$U_1$, $U_2$, ..., $U_i$}; M={$M_1$, $M_2$, ..., $M_j$}; f ={$O_1$, $O_2$, ..., $O_K$} **do**
**for** all the users **do**
Apply the k-means clustering and clusters the users using equation (2)
Calculate fitness function based on $N(U_i)_c$ using equation (3)
**if** (fitness function ($U_i$) < $W_f(CP_c)$) **then**
Replaces the corresponding $U_i$ as the centroid point
**elseif**
Retain $W_f(CP_c)$ as the centroid point
**endif**
Find the missing rating ($R(M_j)$) for ($U_{iT}$) using equation (4) and (5)
**endfor**
**for** all the users ($U_i$) and clusters (C) **do**
Create the sub-clusters ($S_{Ck}$) based on the inherent features
Consider, a new movie with a set of features ($M^{nj}$)
Measure Sim($M^{nj}$, $S_{Ck}$) using equation (6)
**if** ($U_i$ with high Similarity is presented in $S_{Ck}$) **then**
C-MOVIE recommends a new movie to the corresponding $U_i$
**endif**
**endfor**
**endwhile**

## RESULTS AND DISCUSSION

**Experimental evaluation:** This study evaluates the C-MOVIE approach with the conventional HPSO approach[14] to assess the performance improvement of the C-MOVIE approach in terms of prediction and recommendation accuracy. The evaluation process implements the PSO and K-means clustering algorithm in the HPSO approach to compare the performance of the K-means clustering and Cuckoo search optimization algorithm based C-MOVIE approach over a Douban Movie dataset.

**Experimental setup:** The evaluation framework conducts the experiments on Linux Ubuntu 12.04 LTS 64-bit machine with a 2.9 GHz Intel CPU and 32 GB memory. It employs the Java version 1.8.0 from Open JDK to implement the proposed algorithms. To evaluate the C-MOVIE approach, the evaluation framework employs the Douban Movie dataset. This data set comprises a user name, Movie name, the content of the comment and star rating on the movies provided by the users where in rating in the range from 1-5. Moreover, the evaluation model exploits the Infobox information about each movie from the Wikipedia source and employs the textual review information of the corresponding movie by the users from the Douban movie short comments. This dataset consists of >2 million short comments on 28 movies. The user name and their short comments are in the Chinese language that is translated into the English language by the translator tool.

### Evaluation metrics

**Precision:** It is the ratio between the number of users obtaining the relevant movie recommendations and the number of users who are all suggested to a specific movie recommendation.

**Recall:** It is the ratio between the number of users obtaining the relevant movie recommendations and the total number of relevant users with respect to that specific movie recommendation.

**F-measure:** It is the harmonic mean of precision and recall.

$$\text{F-measure} = 2*\frac{\text{Precision}*\text{Recall}}{\text{Precision}+\text{Recall}}$$

**Normalized Mean Absolute Error (NMAE):**

$$\text{Name} = \frac{(1/n)\sum_{i,\,j}\left|P_{i,j}-R_{i,j}\right|}{R_{max}-R_{min}}$$

Where, $P_{i,j}$ and $R_{i,j}$ refers the predicted and actual rating score of the ith user on the jth movie. NMAE score represents the normalized score of NMAE values is irrespective of the rating score. $R_{max}$ and $R_{min}$ denote the maximum possible rating value and the minimum possible rating value respectively.

### Evaluation results

**Unrated movie ratio vs. precision:** Figure 2 illustrates the precision of both the C-MOVIE approach and the HPSO approach when increasing the Unrated movie ratio from 0.753-0.817 and varying the Number of Features (NF) per movie as NF = 4, 7 and 10. The Unrated Movie ratio is defined as the ratio between the number of Unrated movies and the total number of movies. The precision value decreases while increasing the number of Unrated movies per user. When NF = 10, the C-MOVIE

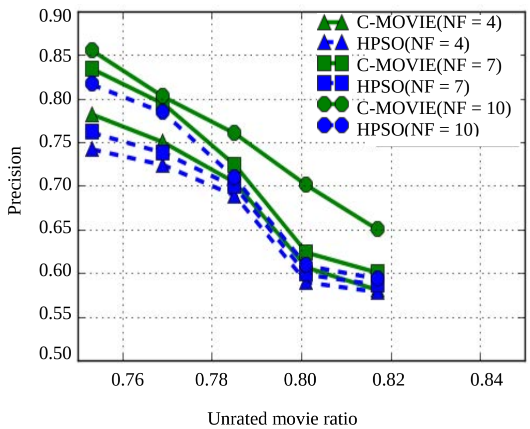


Fig. 2: Unratedm movie ratio vs. precision

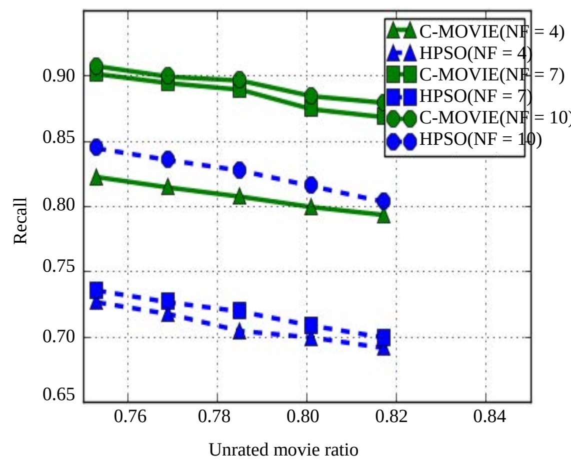


Figure 3: Unrated movie ratio vs. recall

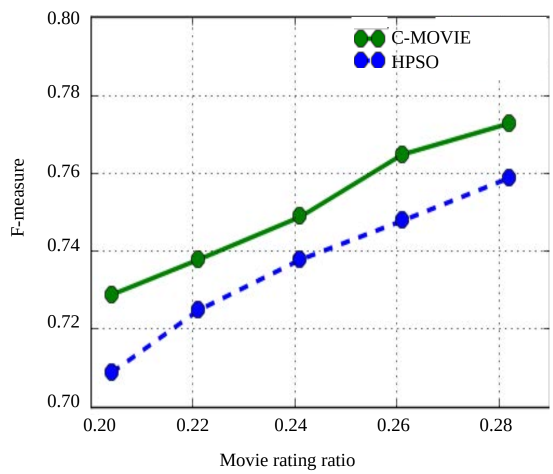


Fig. 4: Movie rating ratio vs. f-measure

marginally decreases its precision value by 23.9% even when increasing the number of Unrated movies. However, at the same scenario, the existing HPSO approach suddenly degrades its performance by 27.17% due to the lack of accurate missing rating prediction from the clustered data. Moreover, the C-MOVIE approach accomplishes a remarkable precision value when there are a huge number of features for each movie. Accordingly, it obtains 3.57% of higher precision value than the HPSO approach when the NF is 7 and Unrated movie ratio is 0.785. The C-MOVIE approach optimally reduces the number of movie's features for each user by the PCA method. Thus, averts the misclassification of rating prediction and inaccurate movie recommendation.

**Unrated movie ratio vs. recall:** The recall value of both the C-MOVIE approach and the HPSO approach is depicted in Fig. 3 with the variation of Unrated Movie Ratio per user and NF per movie. The overall recall value degrades with the increasing Unrated Movie Ratio and the C-MOVIE approach generates higher recall rate indicates that it recommends the desired movies much more effectively to the users than the existing HPSO approach. When NF = 4, the C-MOVIE approach attains the recall value as 80.8% and it is quite close to the recall rate of the HPSO approach when NF = 10 at the point of the Unrated movie ratio is 0.785. It is because, the C-MOVIE approach clusters the users based on the inherent features using k-means clustering as well as a cuckoo search optimization algorithm facilitates the aggregation of like-minded users in a single cluster. Moreover, the C-MOVIE approach explores the intended features behind the rating value and selects the significant features for each movie with the target of dimensionality reduction.

**Movie rating ratio vs. F-measure:** The F-measure value of the C-MOVIE approach and the HPSO approach is shown in Fig. 4 when varying the Movie Rating Ratio from 0.204-0.282. The movie rating ratio is the ratio between the number of ratings provided by the user and the total number of movies. The F-measure value increases with the variation of the number of ratings per user on the movies. The C-MOVIE approach achieves 72.9% of F-measure value even when the movie rating ratio is minimum as 0.204. However, the HPSO approach manages to reach only 70.9% of F-measure value. The C-MOVIE approach utilizes both the preferences of the feature based similar users and preferences on the movies rated by a specific user. The C-MOVIE approach maintains the optimal range of F-measure value from 72.9% to 77.3% when there is a fluctuation in the number of ratings provided which is obtained by effective utilization of the feature based neighborhood information. Moreover, the C-MOVIE approach effectively clusters and sub-clusters the users based on the inherent features using k-means clustering and the cuckoo

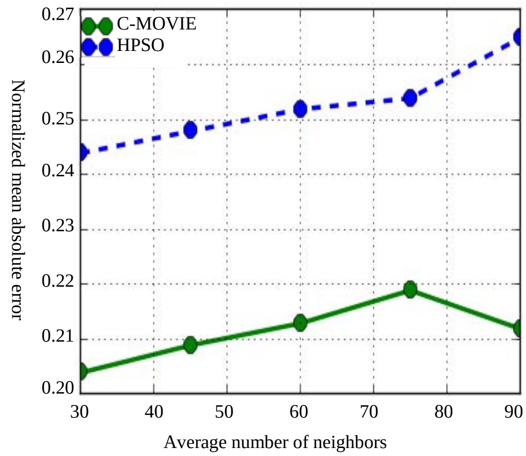


Fig. 5: Average number of neighbors vs. NMAE

search optimization algorithm. Cuckoo search optimization algorithm enhances the performance of the clustered results by the k-means algorithm through the dynamic identification of the best fitness solution for each user with the neighbors of the cluster.

**Average number of neighbors vs. NMAE:** Figure 5 illustrates the rating prediction accuracy of the C-MOVIE approach and the existing approach with the variation of an average number of neighbors in the clusters from 30 to 90. The minimum NMAE score denotes the high performance of the personalized movie recommendation system. When the neighborhood size varies from 30 to 60, the C-MOVIE approach efficiently handles the prediction value to reach the optimum level and then on, the C-MOVIE approach becomes relatively stable around 60 neighbors of the target user in the cluster since, the PCA enabled proposed clustering algorithm performs better accuracy than the existing system. The C-MOVIE approach considers that the PCA-based clustering algorithm is necessary to populate the original user-rating space densely. When the neighborhood size is 75, the C-MOVIE approach and the HPSO approach provide a similar variation of NMAE value seems that the HPSO approach only predicts the accurate rating until it reaches an optimal number of neighbors in the cluster and then, it tends to degrade the performance. Accordingly, the HPSO rapidly increases the NMAE by 4.33% after reaching the optimal number of neighbors from 75-90 is shown in Fig. 5. However, the C-MOVIE approach maintains the optimal prediction rate even after reaching the optimal number of neighbors per cluster due to the consideration of inherent feature based clustering.

## CONCLUSION

This study has proposed a feature based movie recommendation system for social network users. The C-MOVIE approach suggested a desired movie to the users by the proposed algorithm. The recommendation algorithm involves the feature dimensionality reduction, optimal clustering, missing rating prediction and contextual movie recommendation steps. These are accomplished by the various methods such as PCA-based feature reduction, k-means clustering and cuckoo search optimization based optimal clustering, PMF-based rating prediction and movie features based recommendation. The evaluation framework tests the C-MOVIE approach with the Douban movie data set to illustrate the improved results of the proposed personalized movie recommendation system which in terms of both the accuracy and performance. The implementation results of the C-MOVIE approach significantly outperforms the existing HPSO method by accomplishing 23.61% higher recall value.

## REFERENCES

01. Bobadilla, J., F. Ortega, A. Hernando and A. Gutierrez, 2013. Recommender systems survey. Knowl. Based Syst., 46: 109-132.
02. Chen, L., G. Chen and F. Wang, 2015. Recommender systems based on user reviews: The state of the art. User Model. User-Adapted Interact., 25: 99-154.
03. Isinkaye, F.O., Y.O. Folajimi and B.A. Ojokoh, 2015. Recommendation systems: Principles, methods and evaluation. Egypt. Inf. J., 16: 261-273.
04. Desrosiers, C. and G. Karypis, 2011. A Comprehensive Survey of Neighborhood-based Recommendation Methods. In: Recommender Systems Handbook, Ricci, F., L. Rokach, B. Shapira and P.B. Kantor (Eds.). Springer, New York, USA., ISBN-13: 9780387858197, pp: 107-144.
05. Lops, P., D.M. Gemmis and G. Semeraro, 2011. Content-Based Recommender Systems: State of the Art and Trends. In: Recommender Systems Handbook, Ricci, F., R. Lior, S. Bracha and B.K. Paul (Eds.). Springer, Berlin, Germany, ISBN:978-0-387-85819-7, pp: 73-105.
06. Shi, Y., M. Larson and A. Hanjalic, 2014. Collaborative filtering beyond the user-item matrix: A survey of the state of the art and future challenges. ACM. Comput. Surv. (CSUR.), 47: 1-45.
07. Xu, Y. and J. Yin, 2015. Collaborative recommendation with user generated content. Eng. Appl. Artif. Intell., 45: 281-294.
08. Cai, X., M. Bain, A. Krzywicki, W. Wobcke and Y.S. Kim et al., 2010. Collaborative Filtering for People to People Recommendation in Social Networks. In: Advances in Artificial Intelligence. Jiuyong, L. (Ed.). Springer Berlin Heidelberg, Berlin, Germany, ISBN: 978-3-642-17431-5, pp: 476-485.

09. Lin, F.C., H.W. Yu, C.H. Hsu and T.C. Weng, 2011. Recommendation system for localized products in vending machines. Expert Syst. Appl., 38: 9129-9138.
10. Aher, S.B. and L.M.R.J. Lobo, 2013. Combination of machine learning algorithms for recommendation of courses in E-learning system based on historical data. Knowledge-Based Syst., 51: 1-14.
11. Nilashi, M., D. Jannach, B.O. Ibrahim and N. Ithnin, 2014. Clustering-and regression-based multi-criteria collaborative filtering with incremental updates. Inf. Sci., 293: 235-250.
12. Georgiou, O. and N. Tsapatsoulis, 2010. Improving the Scalability of Recommender Systems by Clustering Using Genetic Algorithms. In: Artificial Neural Networks, Diamantaras, K., W. Duch and L.S. Iliadis (Eds.). Springer, Berlin, Germany, pp: 442-449.
13. Soares, M. and P. Viana, 2014. Tuning metadata for better movie content-based recommendation systems. Multimedia Tools Appl., 74: 7015-7036.
14. Behera, R.N., P.L. Saha, A. Chakraborty and S. Dash, 2017. Hybrid movie recommendation system based on PSO based clustering. Int. J. Control Theory Appl., 10: 41-49.
15. Gandomi, A.H., X.S. Yang and A.H. Alavi, 2013. Cuckoo search algorithm: A metaheuristic approach to solve structural optimization problems. Eng. Comput., 29: 17-35.
16. Hatami, M. and S. Pashazadeh, 2014. Improving results and performance of collaborative filtering-based recommender systems using cuckoo optimization algorithm. Int. J. Comput. Appl., Vol. 88, No. 16.