Multimedia Tools and Applications (2019) 78:14303–14330
https://doi.org/10.1007/s11042-018-6790-y

A novel Adaptive Genetic Neural Network (AGNN) model for recommender systems using modified k-means clustering approach

C. Selvi[1] **· E. Sivasankar**[1]

Received: 22 November 2017 / Revised: 10 August 2018 / Accepted: 18 October 2018 /
Published online: 5 November 2018

Abstract

The Recommender System (RS) plays an important role in information retrieval techniques in a bid to handle massive online data effectively. It gives suggestions on items/services to the target online user to ensure correct decisions quickly and easily. Collaborative Filtering (CF) is a key approach in RS providing a recommendation to the target online user, based on a rating similarity among users. Unsupervised clustering approach is a model-based CF, which is preferred as it ensures simple and effective recommendation. Such CFs suffer from a high error rate and needs additional iterations for convergence. This paper proposes a Modified k-means clustering approach to eliminate the above mentioned issues to provide well-framed clusters. The novel supervised Adaptive Genetic Neural Network (AGNN) method is proposed to locate the most favored data points in a cluster to deliver effective recommendations. The performance of the proposed RS is measured by conducting an experimental analysis on benchmark MovieLens and Netflix datasets. Results are compared with state-of-the-art methods namely Artificial Neural Network (ANN) and Fuzzy based RS models to show the effectiveness of the proposed AGNN method.

Keywords Recommender System (RS) · Collaborative Filtering (CF) · Clustering approach · Modified k-means · Genetic Algorithm (GA) · Artificial Neural Network (ANN) · Adaptive Genetic Neural Network (AGNN)

1 Introduction

Growth and use of websites increases daily due to technology advancement and population increase. This leads to a massive growth of web content in terms of user's opinion regarding

✉ C. Selvi
selvichandran.it@gmail.com

E. Sivasankar
sivasankarelango@gmail.com

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

a specific service on the website. Due to this, a new online user has difficulty in deciding on the correct services. Manual analysis of web content by the user is both time-consuming and leads to incorrect decisions the maximum number of times [35]. To overcome such information overload, a personalized Recommender System (RS) is resorted [31]. The RS tool uses user opinion history on services like Movies, Shopping, TV, Tourism, and Taxi use etc. User opinion provided for services are two types: Implicit and Explicit. Explicit opinions are numerical ratings given within the range of certain rating scale values which change from service to service(i.e., one item may be rated on a 1 to 5 scale and another may have a 1 to 10 rating). Implicit opinions are given in unstructured text formats.

An e-commerce website uses a RS tool to predict the on line user's next purchase behavior using his/her past history and recommends the service to them to increase profit. There are several filtering algorithms which provide recommendations to targeted online users. The three RS categories include Collaborative Filtering (CF) [26], Content-Based Filtering (CBF) [1, 3] and hybrid filtering [17, 18]. CF is a widely used RS filtering algorithm which analyses user's explicit rating on an item to provide a recommendation. The CBF algorithm works by matching content features of an already purchased items with features of an item yet to be recommended [33]. Recommendations are provided only when matches are located. CF ensures improved accuracy compared to CBF and it is also domain independent. CF methods are preferred when an item has no content features or its features are hard to extract. Then, CF recommends an item to the targeted online user by utilizing the item's explicit numerical rating and its closest neighbors. Hence, CF is considered effective due to its ease and ability to ensure an accurate recommendation.

CF is further divided into memory-based and model-based methods. In the former, the closeness between the targeted online user and other users is calculated. Then users nearest targeted user are identified based on the similarity values calculated. Finally, the target's item values are predicted based on the obtained nearest users and the recommendations are provided based on the ranked target items predicted value. This method is further classified into user-based and item-based methods. The user-based method predicts a new item's rating value based on targeted online users nearest users. In item-based methods, a target item's rating value is predicted regarding items nearest to the target item. In memory-based CF, the full user-item rating matrix has to be loaded into memory for every similarity calculation of the target user. This leads to scalability issues when the number of users rating on items increases with time. It also consumes additional time to process the huge matrix. To overcome this, many researchers prefer the model-based CF method.

The two main approaches of the model-based method include supervised and unsupervised learning techniques. Many research work in RS was done using supervised learning methods [13]. When the model is developed using supervised learning algorithms on available data history (user-item rating matrix), prediction on the unseen data is possible at anytime. That is, data is prepared in advance and the model learns from it collectively. The drawback of this method is that, there is no straight forward method to transform the user-item rating matrix (CF task) into a training dataset with labelled classes acceptable to supervised learning tasks [13]. This results in inaccurate recommendations but can be overcome through use of a model-based unsupervised clustering based CF approach [13].

Similar data points are grouped together using a model-based unsupervised clustering approach which locates patterns for unlabeled data points. The popular k-means clustering approach is used in RS as it assigns data points with reduced distance to a corresponding cluster [45, 48]. Data points in k-means are assigned to only one cluster at a time. Much research work was undertaken using k-means and its variants. Its shortcomings include considerably high error rate and needing many iterations to ensure well-framed clusters [19, 39].

Also, data points may reside in two or more clusters at a time [19]. To solve the above issues, this paper proposes a new modified k-means clustering approach which reduces error rate and increases recommendation accuracy greatly.

A supervised Artificial Neural Network (ANN) model is used in the proposed RS model as the most favorable data points in each well-framed cluster must be identified to provide effective recommendations while simultaneously lowering the recommendation error rate. Weights in ANN are updated through backpropagation technique. ANN error values are gradually reduced by updating the network's weights using backpropagation based on the error value from earlier iterations. Also, it takes many iterations to ensure a significantly lower error rate. To ensure favorable data points on a lower error rate with reduced computational time, the RS method is proposed with ANN hybridization using a bio-inspired algorithm.

As backpropagation of weight updation consumes computational time (many iterations), a bio-inspired algorithm optimizes ANN weights at each iteration. Various bio-inspired algorithms used in most applications include Genetic Algorithm (GA) [8, 12], Particle Swarm Optimization (PSO)[26] and Cuckoo Search (CS) [32, 44]. GA is preferred due to its randomly multiplying solutions using crossover and mutation operations.

The objective of this paper is to select most favored data points from each cluster to improve RS accuracy for which a novel supervised Adaptive Genetic Neural Network (AGNN) method which eliminates ANN's limitations and increases RS performance significantly is proposed. It also uses an unsupervised modified k-means clustering approach. To show the proposed method's improved efficiency, it is compared to ANN and Fuzzy based RS models.

The main contribution of this paper include,

- Standard benchmark MovieLens and Netflix datasets being preprocessed to identify and remove a movie with different characteristics (outliers) using CF based similarity measure.
- Proposing a new modified k-means clustering approach on a processed dataset to ensure well-framed clusters.
- The proposed modified k-means clustering approach being validated experimentally with the current k-means method.
- The most favored user in each cluster being identified effectively by proposing a new AGNN method to assure an accurate recommendation.
- Effectiveness of the proposed AGNN with new modified k-means clustering approach is experimented on standard benchmark datasets and compared with existing ANN and Fuzzy based RS models.

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 Section 3. Results and discussion of various clustering combinations, AGNN, ANN and Fuzzy based models are explained in Section 4. Finally, Section 5 concludes the work.

2 Literature survey

The growth of web 2.0 and social media resulted in a huge increase in individuals who come up with the resource of the content producer rather than being just content consumers. This leads to information overload due to which a targeted online user finds it difficult to analyze and come up with correct decision effectively most times [35]. Also, handling a huge volume of user-generated web content manually by a person is hard and consumes much time. All these web contents must be processed to arrive at better and precise decisions which is done by the RS tool. Numerous web services like Online business, Media, Computerized library and Online promoting etc, use the RS tool to increase revenue [31]. The main RS classes include: CBF [33] and CF [22, 43] filtering techniques. CF is again classified into memory-based and model-based methods [34, 56].

Memory-based CF models recommend items to the targeted online user based on neighbors using similarity measures like Pearson, Cosine, Adjusted Cosine, Constrained Pearson, Mean Squared Difference (MSD), Jaccard MSD and Bhattacharyya etc [43]. CF has scalability issues and also will not ensure an accurate recommendation in sparse dataset as most similarity measures used for CF based RS work only on co-rated item values.

Model-based CF method overcomes this by building a model using supervised and unsupervised approaches. Many RS methods use model-based supervised approaches to provide recommendations [13]. Its drawback is that the accuracy of RS is based on labeled training dataset's correctness. Also labeling a training dataset needs much effort and could result in erroneous values. The model-based unsupervised CF method tackles this [30]. Most RS based application use clustering approaches which group similar data points together using distance measures [38]. The k-means clustering approach is a widely used method for all clustering approaches regarding time and complexity [25]. The k-means and variants are applied to RS application issues like scalability, sparsity, and cold start problems [46, 54]. Some work on the movie RS was done using variants of k-means clustering approaches as stated below. Bisecting k-means clustering approach was proposed for privacy-preserving applications [11] and web-based movie RS [48]. A new centroid selection algorithm was used to reduce the cluster's training cost using traditional k-means clustering approach [53]. The CLUSTKNN clustering approach was used to handle data points in large-scale RS applications [45]. The clustered social ranking algorithm was introduced to suggest content interest for new web online users [54, 55]. The k-means approach and its variants form a cluster where data points have a fixed chance to fall into one cluster at a time [51]. But data points could fall into more than one cluster. Also, the k-means approach does not guarantee clustering convergence. To achieve guaranteed convergence with lower k-value, Fuzzy K-means Expectation Maximization (FKEM) algorithm was introduced as it combines weighted k-means approach and the Expectation Maximization algorithm [39]. Though k-means and its variants are used in many RS applications [10], it has a bigger error rate and takes more iterations for data point convergence. That is, generating well-framed clusters with lower error rate is still a challenge in RS related applications. To overcome these issues this paper proposes a new modified k-means clustering approach to enhance RS accuracy.

The main goal of this paper is to provide improved RS performance by finding most favorable data points from a processed cluster to suggest an item to the targeted online user. To get such favorable data points, an supervised ANN model is preferred in this paper as it focuses on reducing prediction error rate gradually by updating node weights. Research work on Neurocomputing with ANN model is performed for quite a long time [15, 36]. Personal computers computational capabilities are started to meet requirements with ANN, which increased its popularity from the 1980's [15]. ANN detects complex non-linear relationships among independent variables and features, to detect higher order polynomial features and their interactions in addition to influencing the availability of various optimization algorithms. All these can be done by choosing a suitable non-linear activation function [41, 52]. From this, it is concluded that a model designed with ANN can learn the complicated relationships between items and users effectively and thus ensure improved recommendations. The content and collaborative features are combined to provide recommendations by developing an ANN model to lower log loss and misclassification error rate using a stochastic gradient descent optimization algorithm [14, 29, 42]. A backpropagation ANN based TV RS was proposed using three-layered ANN model [5, 21, 23]. Spiking and multilayer perceptron ANN [6] and a two-stage of Self-Organizing Map (SOM) ANN based RS model [29] were suggested to improve key metrics performance and scalability for movie RS. TV RS was proposed to find the probability that the user can watch a program using his/her implicit and explicit feedback. The proposed model learns implicit input data using Bayesian and decision tree classifier models, and explicit input data using the ANN model [57]. Though lot of work in RS was done using an ANN model, it has certain drawbacks. In every ANN iteration, weights are updated through backpropagation based on the error value from the earlier iteration. Weights updation at each iteration reduces predicted error value gradually. An ANN model attains expected lower error rate only after many iterations due to backpropagation.

To overcome this, a RS model which combines GA optimization algorithm with ANN to optimize node weights at each ANN iteration to attain the expected error value with reduced iterations is proposed. GA suits weight optimization problems with a fitness function subject to soft and hard constraints [28]. GA is used in two types of RS [12] like clustering [28] and hybrid models [2]. GA is used in many applications: with k-means clustering approach for personalized RS in online shopping market segmentation [28], to obtain an optimal similarity function among conventional similarity measures for movie RS [12], combined with CBF for music RS [27], optimizing weights of implicit and explicit features [47], weight and threshold optimization to calculate similarity functions [24], to obtain user-feature space containing most discriminating features from large user-item space [7], to locate an optimal solution on clustered data points obtained by Fuzzy C-Means clustering approach [20] and to improve overall recommendation accuracy by optimizing all combinations of combined recommendation results [16]. A new GA based similarity measure called SimGen was proposed to compute similarity value between users without using traditional similarity measures [4]. As far as RS is considered, solutions are scattered in GA due to the randomness nature (crossover and mutation), GA is used in this paper to optimize ANN weights in each iteration to reduce its computational time.

In this paper, a novel AGNN method is proposed to optimize weight of the ANN model using GA algorithm to recommend items to online targeted users using a new modified k-means approach to improve RS accuracy.

3 Proposed methodology

The main objective of this proposed RS model is to ensure an efficient recommendation to the targeted online user using AGNN model with a modified k-means 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

Ratings of different users on multiple movies are collected from standard benchmark MovieLens [37] and Netflix [40] datasets. In the preprocessing step, movies with dissimilar characteristics are removed i.e. If online movies are rated by considerably lesser number

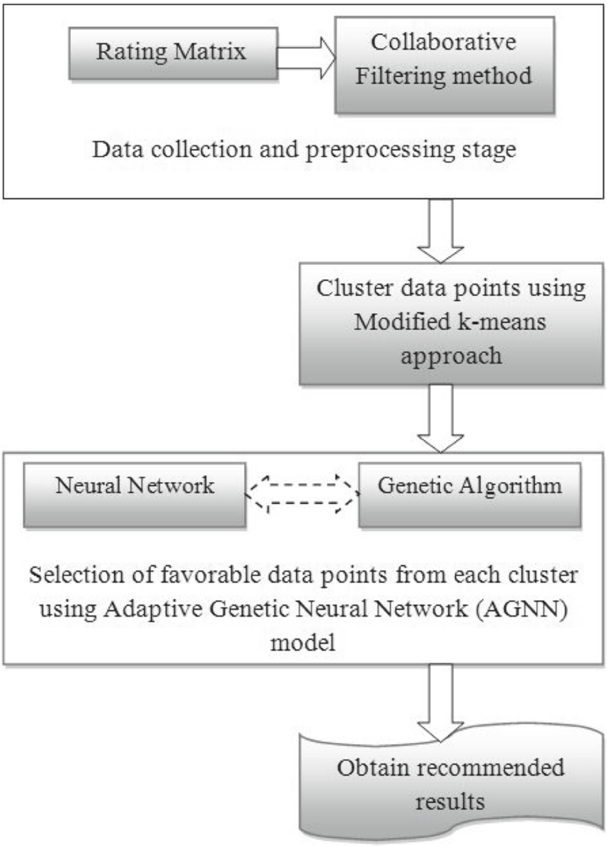

Fig. 1 Flow diagram of Proposed RS

of users, then taking such movies into further phases of RS will affect its performance. If the data points are clustered into groups then movies with dissimilar characteristics may fall into some cluster which in turn affects RS accuracy. To enhance RS performance, these dissimilar movies are considered as outliers and removed for further processing [49].

For effective preprocessing, a movie-based CF similarity measure is used. It finds similarity among movies which is the ratio between the product and addition of ratings for all users with co-rated movies. Let L be the set of online users $\{ou_1, ou_2, ou_3, ..., ou_x\}$ and M be the set of online movies $\{om_1, om_2, om_3, ..., om_y\}$. Let p be any arbitrary user in a set L who has rated both online movies om_s and om_t from the set of M movies ($om_s, om_t \in M$). Then, the similarity measure is defined as,

$$sim(om_s, om_t) = \frac{\sum\limits_{p \in (om_s, om_t)} r_{x,om_s} \times r_{p,om_t}}{\sum\limits_{p \in (om_s, om_t)} r_{p,om_s} + r_{p,om_t}} \tag{1}$$

where r_{p,om_s} and r_{p,om_t} represents the rating of online user p on online movies s and t respectively. $p \in (om_s, om_t)$ represents the user p is found to be common for rated movies om_s and om_t.

$$om_s = \frac{\sum\limits_{om_y \in M} sim(om_s, om_t)}{|M|_{om_y \in M}} \tag{2}$$

Where $|M|_{om_y \in M}$ defines the count of om_y in M. The average similarity value for each online movie is calculated as in (2). Finally, the movies are ordered based on their average similarity value and those with less average similarity value (based on outlier cutoff value) are removed from the dataset. Figure 2 shows the accuracy of the proposed RS for clustering techniques with size 5 regarding to change in the outlier cutoff value. From Fig. 2, it is clearly observed that accuracy of the proposed RS for MovieLens dataset increases till it reaches outlier cutoff value 0.3 and starts reducing afterwards. Outlier cutoff value for Netflix dataset is 0.35 which is fixed experimentally based on a sample dataset from standard benchmark datasets. The remaining preprocessed dataset is taken to process the RS model further.

3.2 Modified k-means clustering approach

An easy way to solve clustering problems is the k-means partitioning algorithm. It splits a set of data points into a predefined number of clusters with the end goal of finding the closeness between data points within the clusters which is more than that of data points among the clusters. That is, the sum of a squared distance between data points within the cluster is minimum. The similarity between data points is calculated with a common used distance measure called Euclidean Distance (ED). The computed squared ED from each data point of every cluster allocates data point to the nearest cluster.

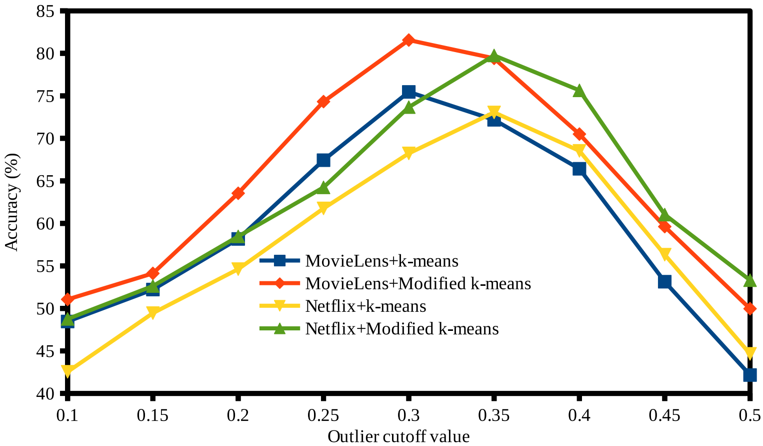

Fig. 2 Accuracy for proposed AGNN method of clustering techniques by varying outlier cutoff value for MovieLens and Netflix Datasets

Minimization of an objective function for a k-means approach is given in (3).

$$Objective\_function_{k-means} = \sum_{j=1}^{n}\sum_{i=1}^{l}(\parallel t_i - c_j \parallel)^2 \tag{3}$$

where t_i represents the i^{th} data points, c_j represents the j^{th} cluster centroid, $\parallel t_i - c_j \parallel$ denotes the ED between t_i and c_j, n denotes number of clusters and l denotes the number of data points.

The existing clustering k-means method groups users together where a similarity among them within the group is maximum. Generally, each data point has a probability of having a place in each cluster, as opposed to totally having a place in only one cluster as in a conventional k-means approach. Each cluster method run individually with a various group sets to discover the best clustering method that ensures the most noteworthy recommendation accuracy. Also, k-means suffers from a large error rate and more iterations to get well-framed clusters which thus reduces RS performance considerably. To offset this, the paper proposes a modified k-means clustering approach which places data points in more than one cluster. As some user's taste is similar to more than one cluster group, this improves recommendation accuracy (User's information is not lost in the possible cluster group) and reduces cluster error rate. As some users falls in to more than one cluster, the convergence happens quicker and also each cluster have more similar user that in turn reduces the error rate. The small reduction in cluster error rate value affects the RS performance.

Data points are assigned to a cluster based on the proposed modified k-means approach and stated as follows,

$$\begin{aligned} &Objective\_function_{Modified\ k-means} \\ &= \begin{cases} t_i\ to\ C_j & If\ \sum_{j=1}^{n}\sum_{i=1}^{l}(\parallel t_i - c_j \parallel)^2 < T_j \\ min(\sum_{j=1}^{n}\sum_{i=1}^{l}(\parallel t_i - c_j \parallel)^2) & Otherwise \end{cases} \end{aligned} \tag{4}$$

$$T_{ij} = \sum_{j=1}^{n}\sum_{i=1}^{l}\frac{1}{count(ED_{ij})}sum(ED_{ij}) \tag{5}$$

Threshold value T_{ij} is calculated by taking the average of calculated distance value from i^{th} data point to j^{th} cluster. The T value is calculated as given in (5), because the new centroid is calculated based on the average value of data points in the corresponding cluster. The centroid is updated based on clustered data points as in (6).

$$c_j = (\frac{1}{|s_j|})\sum_{t_i \in s_j} t_i \tag{6}$$

where $|s_j|$ is total data points in the j^{th} cluster. This process is repeated until it reaches the stopping condition, i.e., no data point was reassigned in the cluster. Algorithm 1 describes the steps for the proposed modified k-means clustering approach.

Algorithm 1 Modified k-means clustering algorithm

Input: Data points, Number of cluster;
Output: Clustered data points;

```
1:  begin
2:      Select random n data points as a cluster centroids;
3:      for each data point i do
4:          Calculate the ED between each data point and cluster centroids as given in (3);
5:          Allocate data point i to the cluster whose centroid is closest;
6:      for j = 1 to n cluster do
7:          Find the threshold as given in (5);
8:          Calculate the new cluster centroids using (6);
9:          Do Step 3 to 4;
10:         if ED_i ≤ T_j then
11:             Allocate data point i to the j^th cluster as given in (4);
12:         else
13:             Do Step 5;
14:     Repeat Step 6 to 13 until no data point was reassigned or it reaches MaxIteration;
15: end
```

The modified k-means assures that each data point is grouped at least in any one cluster. Finally, data points are clustered with minimum error value using the proposed modified k-means approach.

3.3 AGNN

3.3.1 Artificial Neural Network (ANN)

The main objective of RS is to suggest movies to targeted online users based on the most favorable user obtained from each cluster generated from Section 3.2. To obtain the most favorable user from every cluster, ANN is used in this paper. Initially, data points of clusters are given as ANN inputs and nodes are given random weights. Finally, node weights are updated to reduce error value from the earlier iteration (the difference between predicted and desired/actual output). In this paper, a multilayer feed forward backpropagation ANN is utilized with inputs, hidden and output layers. Distribution units of layers connected with a certain weight value in a range of -1 to 1.

Backpropagation algorithm is used to train the ANN, as described below.

Step 1: Generate arbitrary weights within an interval [-1, 1] and assign it to the hidden layer neurons and output layer neurons. Maintain a unity value weight for all input layer neurons.

Step 2: For every output neuron r, error rate is calculated as given in (7).

$$\delta_{op_r} = p_{op_r}(1 - p_{op_r})(t_{op_r} - p_{op_r}) \tag{7}$$

where δ_{op_r} represents backpropagation error, p_{op_r} denotes predicted output and t_{op_r} represents desired/actual output.

Step 3: For every hidden neuron q, error rate is calculated as given in (8).

$$\delta_{op_q} = p_{op_q}(1 - p_{op_q}) \sum_r (w_{qr} \times \delta_{op_r}) \tag{8}$$

where w_{qr} represents the link weight from hidden layer q to output layer r and p_{op_q} represents the output of hidden layer q. The output of hidden and activation layers are obtained by applying a sigmoid activation function. The formula of sigmoid activation function for hidden and output layers is given in Equation 9 and (10) respectively.

$$p_{op_q} = \frac{1}{(1 + e^{-(\sum w_{pq} \times ip_p)})} \tag{9}$$

$$p_{op_r} = \frac{1}{(1 + e^{-(\sum w_{qr} \times ip_q)})} \tag{10}$$

Step 4: Adjust weights of all output neurons as,

$$w_{rq} = w_{rq} + \triangle w_{rq} + (\gamma \times \triangle(t-1)) \tag{11}$$

where, γ is a momentum term, $\triangle(t-1))$ is previous iteration's change in weight and $\triangle w_{rq}$ is the change in weight which is determined as,

$$\triangle\, w_{rq} = \lambda \times \delta_{op_{rq}} \times ip_{rq} \tag{12}$$

where w_{rq} and ip_{rq} are weight and input values of neuron r coming from neuron q respectively and λ is the learning rate which varies from 0.2 to 0.5.

Step 5: Adjust weights of all hidden neurons as,

$$w_{qp} = w_{qp} + \triangle w_{qp} + (\gamma \times \triangle(t-1)) \tag{13}$$

where, $\triangle w_{qp}$ is the change in weight which can be determined as,

$$\triangle\, w_{qp} = \lambda \times \delta_{op_{qp}} \times ip_{qp} \tag{14}$$

where w_{qp} and ip_{qp} are weight and input values of neuron q coming from neuron p respectively.

Step 6: Repeat process from Step 2, until backpropagation error is minimized to the least value. Practically, the criterion to be satisfied is $\delta_{op_r} < 0.1$.

3.3.2 Weight optimization using genetic algorithm

The ANN backpropagation method reduces error rate gradually and uses many iterations to obtain the expected lower error rate. To overcome this, the paper proposes a novel AGNN method which uses, GA optimization to optimize ANN weights in each iteration.

In GA optimization, random weights are generated between -1 to 1 and considered as chromosomes in the initial iteration. Till optimal solutions are achieved the chromosomes are updated once in each iteration. The best chromosomes are generated by crossover and mutation operators. That is, relocation of individuals among dissimilar chromosomes chased by the application of genetic operators results in new individuals. This process is repeated for a specific number of iterations and the best chromosomes are identified to evaluate network error. The steps involved in GA are,

Step 1: Generation of chromosome

Generation of the chromosome for optimization is difficult in the early stages of GA. Here, N numbers of random chromosomes are produced using value encoding method in solution space. The produced chromosomes are considered early chromosome, which is signified as,

$$X_i = [x_0^{(i)}, x_1^{(i)}, x_2^{(i)}, ..., x_{N_L-1}^{(i)}]; \quad 0 \leq i \leq N_p - 1, \quad 0 \leq j \leq N_L - 1 \tag{15}$$

The objective function and the random population of GA are defined so that the entire solution space is covered and converges with the global optimal solution. where, x_j^i is the j^{th} gene of chromosome i, N_P is the population pool and N_L is the length of the chromosome.

Step 2: Fitness function

Fitness function is a kind of objective function, which is the top target parameter for an optimized value. The proposed objective function used to evaluate the chromosome is coined in (16).

$$F_i = \sum_{i=1}^{n_f} \frac{d_i}{n_f} \tag{16}$$

where d_i represents the weighted sum of input and n_f represents the total inputs. The fitness value of every chromosome is evaluated at this point. Chromosomes are selected if they reach the expected decreases in error rates or else they are subjected to perform GA operations such as crossover and mutation to attain the anticipated result.

Step 3: Selection

During selection, the generated chromosomes and new chromosomes are positioned in a selection pool based on their fitness values. Chromosomes with good fitness occupy top positions in the selection pool. The first chromosomes at the top are chosen for the next generation among other chromosomes. Here selection is based on fitness and execution time for each task. So utilizing the above AGNN, recommendations are effectively provided by choosing the most favorable data points in each cluster.

Step 4: Crossover operation

To obtain the latest chromosome called offspring, crossover operation is done between two selected parent chromosomes. The genes are chosen from both parent chromosomes and the latest child chromosome is produced based on crossover rate (C_R). Fitness function is applied to the new chromosome to check its worthiness. The formula to calculate crossover rate is defined in (17).

$$C_R = \frac{Number\ of\ genes\ crossovered}{Chromosome\ length} \tag{17}$$

The executed crossover operation used in this paper has certain advantages that the architectures maintain are authorized, as the supergenes continue unbroken through the crossover operation which means that no liability can go out of the system or is replaced. By managing a one-point crossover, the crossover operation joins two subsets of duties with their relevant architectural structures at this point.

Step 5: Mutation operation

Mutation operation used in GA adds or subtracts a small value on the chromosome's selected points. Mutation operation is based on Mutation Rate (M_R). Based on specified (M_R) genes are mutated subjectively at a certain point. The formula to calculate (M_R) is defined as,

$$M_R = \frac{M_P}{N_L} \tag{18}$$

where M_P is the mutation point.

3.4 Proposed RS model

To provide effective movie recommendation to a targeted online user, the dataset is preprocessed by a collaborative filtering based similarity measure. This measure removes movies which have different characteristics from other movies in the dataset. The processed reduced dataset is given to the proposed modified k-means clustering approach. Well framed clusters are formed by giving users a chance so that a movie has a probability of falling into more than one cluster group. This removes the drawback of traditional k-means and does not lose user details to ensure efficient recommendations is addition to lowering clustering error rate.

The AGNN model is used to choose a most favorable user from each cluster. The backpropagation ANN structure is framed with input, hidden and output layers. The number of neurons in the input layer is equal to the number of distinct movies in the dataset after preprocessing. The number of hidden and output layer is one. The number of neurons in the output layer is one and in the hidden layer is the mean value of neurons in the input and output layers.

In the AGNN's first iteration, for each cluster a normal ANN is implemented by assigning random weights to the links. GA is proposed to reduce ANN error rate and the number of iterations to back propagate the error. That is, node weights in the ANN model are optimized using GA. Initial random populations are taken as 75 and each chromosome is initialized with random weights linearly. The chromosome's length is the number of connected links in the ANN from the input to the output layer and ANN is fully connected in the network. The weight's random population is evaluated with the proposed GA fitness function as in (16). If the minimum error criterion defined for the GA is achieved then the process stops. Otherwise, mutation and crossover operations are executed to form an intermediate population. To do this, the best chromosomes from GA operations are chosen using an elitism mechanism. This process repeats for maximum iterations or till the expected network error is reached.

Then an optimal weight is selected and given to the ANN which finds the difference between the actual/desired result and the predicted network result. ANN iteration stops if the expected error is reached and the user with the minimum error rates is considered favorable and hence recommendations are provided to the targeted online users. Otherwise, ANN backpropagation is resorted to by calling the GA weight optimization for maximum iterations or if the network reaches the desired error value. Each time a new GA iteration starts by computing fitness functions of both GA and ANN. The global best user (Gbest) is the user who has trained with optimal weight value obtained by GA and ANN with reduced error rate. The advantages of the AGNN network is that, it reduces ANN iterations by optimizing node weights using GA.

Algorithm 2 Proposed RS model

Input: l, λ, p, q and r.
Output: Recommendable item from the obtained most favourable data points.
Let $[r]_{X\times Y}$ be the rating matrix

1: **begin**
2: Collect $[r]_{X\times Y}$ dataset;
3: Calculate $sim(om_s, om_t)$ using (1);
4: Remove $om_s \leq$ Outlier cutoff value and Obtain reduced dataset; //The cutoff value is fixed experimentally.
5: Call Algorithm 1.
6: **for** each n **do**
7: Build feed forward backpropagation ANN with p, q and r neurons;
8: Assign random weights to the ANN network and calculate the error rate δ_r;
9: **if** $Minimum\ error\ rate$ or $Stopping\ Criterion$ **then**
10: Select the user with minimum error rate as a Gbest user and provide recommendations to the other users
11: **else**
12: Optimize the node weights of ANN
13: **for** $t = 1$ to $MaxIteration$ or $Termination\ criterion$ **do**
14: Generate initial populations of weights;
15: Calculate F_i ;
16: **if** minimum error criterion **then**
17: Select best chromosome ;
18: **else**
19: Perform crossover;
20: Perform mutation;
21: Repeat Step 15 to 17;
22: Feed forward ANN with optimized weight by computing op_q and op_r of the network.
23: Calculate δ_r.
24: **if** Minimum criterion **then**
25: Find $Gbest$ user solutions and provide recommendations to other users;
26: **else**
27: Repeat from Step 13;
28: **end**

3.5 Performance measures

Precision, Recall, F-measure, and Accuracy are the measures used in this paper to evaluate the performance of the proposed RS. These metrics are evaluated based on entries in 2×2 contingency matrix [9].

3.5.1 Precision

This is a measure of exactness, which evaluates the ratio between the number of correct recommendations and the total recommendations presented to the user. The formula for precision is given in (19).

$$Precision = \frac{TP}{TP + FP} \tag{19}$$

3.5.2 Recall

It is a measure of completeness, which evaluates the ratio between the number of correct recommendations and total relevant recommendations presented to the user. The definition for a recall is given in (20).

$$Recall = \frac{TP}{TP + FN} \tag{20}$$

3.5.3 F-measure

Precision and recall are simplified into one single metric with equal weight called F-measure. It compares the algorithm across the dataset easily. It is defined in (21).

$$F - measure = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} \tag{21}$$

3.5.4 Accuracy

Accuracy is used to evaluate the overall performance of the RS. It evaluates the ratio between a total number of correct recommendations and total items considered for recommendations. Equation 22 shows the formula for calculating accuracy.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{22}$$

4 Experimental results and analysis

4.1 Experimental system design

The performance of the proposed RS model is evaluated by conducting an experimental analysis on standard benchmark datasets namely MovieLens [37] and Netflix [40]. Table 1 describes the details of datasets considered for the analysis.

The MovieLens dataset is of size 100K and includes 100,000 ratings of 1000 users on 1700 movies. Rating values are in a rating scale range of 1 to 5. The size of the Netflix Prize dataset is 1M comprising of 1,000,000 ratings from 480,189 users on 17,770 movies. The

Table 1 Details of Dataset

Dataset	Description	No. of users	No. of movies	No. of ratings	Rating range
MovieLens	Movie	1000	1700	1×10^5	1 to 5
Netflix	Movie	480,189	17,770	100×10^6	1 to 5

number of distinct movies are considered as a feature for further experimentation. First, the benchmark dataset is preprocessed as in Section 3.1 to remove features(movies) with different characteristics. Then the proposed modified k-means clustering algorithm is applied to the processed dataset to obtain well-framed clusters. Finally, the most favorable users in each cluster are identified using the proposed AGNN model and then movies are recommended to targeted online users. To show the effectiveness of the proposed RS model, 500 users are chosen randomly from the MovieLens and Netflix datasets and the results are compared with other state-of-the-art RS model with a 5-fold cross validation method. The experimental analysis of the proposed RS model is based on the steps provided in Algorithm 2.

4.2 Performance analysis of the proposed clustering approach

First, the dataset is collected and preprocessed to eliminate movies which have less correlation with other movies. Then, the processed dataset is grouped using the proposed modified k-means clustering approach. Clustering users plays an important role in RS to improve performance with reduced error rate. To prove the importance of using the proposed clustering approach which ensures less error rate, the analysis is done on standard relational dataset namely Iris, Class, and Diabetes [50]. The standard relational dataset Iris has 150 instances with 4 attributes of three classes. The Glass dataset consists of seven classes of 214 instances with 10 attributes. The Diabetes dataset contains 768 instances of two classes with 8 attributes. To show the superiority of the proposed clustering approach, it is compared to current k-means approach.

Figure 3 represents the error value comparison of k-means and the proposed modified k-means clustering approaches by setting its parameters as in Table 2. The experiment is conducted 10 times and average error values calculated. Error values are more in initial iterations, are reduced in later iterations and finally stabilize to a specific error value. Error values are normalized on a 0 to 1 scale.

Figure 3a, b and c represent the normalized error values of k-means and modified k-means clustering approaches for Iris, Glass and Diabetes datasets respectively. The primary x-axis in Fig. 3a, b and c represents the number of iterations and the y-axis represents the normalized error value. From Fig. 3a, b and c, it is clearly understood that, the error value of the proposed modified k-means is reduced drastically compared to k-means clustering. Hence, the modified k-means clustering method is considered the best clustering method and used by RS to group similar users in groups with reduced error. So, the modified k-means provides a well-framed cluster as the objective of RS is improving performance by reducing error rate. As in Algorithm 1, modified k-means clustering approach is applied to the preprocessed benchmark datasets and well-framed clusters obtained.

4.3 Performance analysis of AGNN for the movieLens dataset

Most favorable users from each cluster got from Section 4.1 are obtained using the proposed AGNN method. Suggestions are provided to targeted online users based on the selected favorable users from each cluster. To validate the performance of the proposed AGNN method, its performance is compared to the existing ANN and Fuzzy models. Parameters for implementing AGNN, ANN and Fuzzy based RS models are provided in Table 2.

The effectiveness of the proposed RS model is shown by comparing its results in all combinations of existing k-means clustering approaches, along with that of the ANN and Fuzzy models. The results shown in Tables and Figures. are used to clarify and enhance understanding of the relationship between parameters such as number of clusters, number

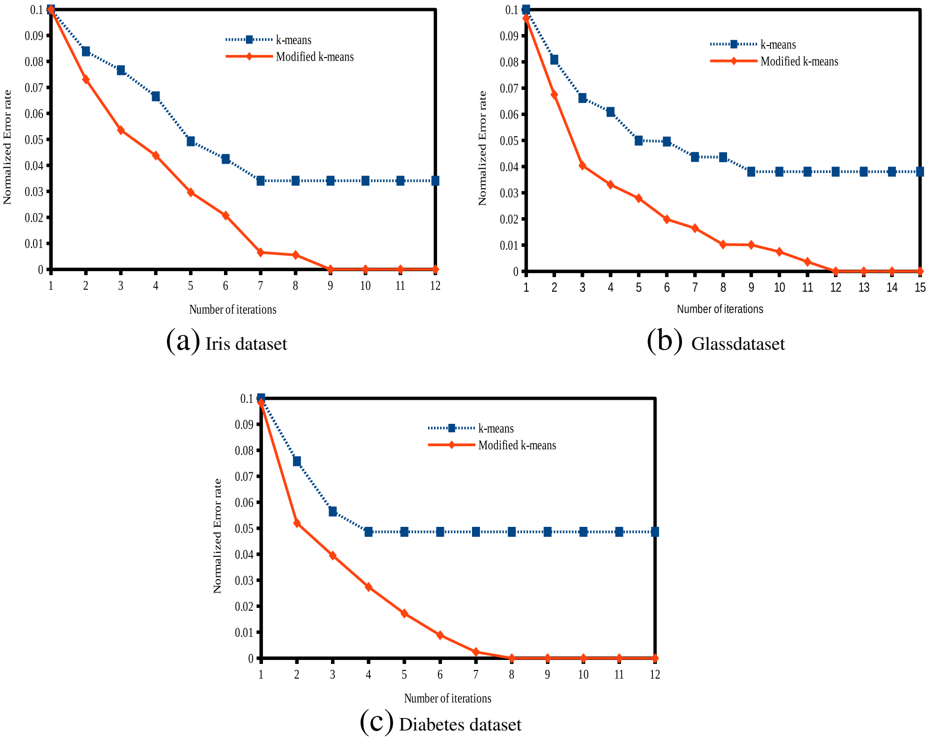

Fig. 3 Normalized Error value for Clustering Techniques

Table 2 Control Parameters considered for RS model

Clustering Approach	Algorithms		
Modified k-means	Fuzzy	ANN	AGNN
Number of cluster: n	Maximum Iteration: $MaxIteration$: 100	Maximum Iteration: $MaxIteration$: 100	Maximum Iteration: $MaxIteration$: 100
Threshold value: T	Population size: 100	Population size: 100	Momentum term γ: 0.9
Maximum Iteration: $MaxIteration$: 100	Membership function: Triangular	Momentum term γ: 0.9	Learning rate λ: 0.45
		Learning rate λ: 0.45	Mutation: Add/Subtract value in selected position Mutation rate: 0.8
			Crossover: One-point Crossover rate: 0.78
			Selection: Elitism
			Encoding: Value encoding

Table 3 Recommendation for different Cluster Size on Proposed Method: MovieLens Dataset

Number of clusters	Number of users	Correctly Recommended	Wrongly recommended
2	100	82	18
3	100	84	16
4	100	85	15
5	100	87	13

of iterations and number of populations which analyzes the performance of the proposed approach. The recommendations for different cluster sizes for the MovieLens dataset are given Table 3.

From Table 3, it is clearly understood that the number of correct recommendations increase when cluster size increases. Table 4 shows the recommendation results of the proposed RS model in terms of precision, recall and F-measure calculated from the results in Table 3. A high precision value indicates fewer users getting wrong recommendations and a higher recall value indicates improved recommendation for more users. Using these expressions, F-measure is calculated for different cluster values under 100% of population size.

Figure 4 shows the overall F-measure value of the proposed RS model (modified k-means clustering and AGNN model) with combinations of existing k-means clustering approach, ANN and Fuzzy models on the MovieLens dataset. From Fig. 4, it is seen that the proposed modified k-means with the AGNN combination of RS performs better when compared to other combination methods. It is also inferred that the proposed modified k-means clustering approach provides a better performance for all AGNN, ANN, and Fuzzy models. Due to the random nature of the GA, RS performance cannot be judged by a single run. AGNN weight values are optimized for varying iterations or until it reaches the termination criterion.

The overall F-measure value of the proposed AGNN on the MovieLens dataset is compared to existing ANN and Fuzzy models under the proposed modified k-means clustering approach and the results are shown in Fig. 5. F-measure for the proposed system is 0.93 for cluster size 5, 0.859 and 0.813 for existing ANN and Fuzzy based RS models respectively. It is inferred from Fig. 5, that the F-measure performance of RS increases with increased cluster sizes. From Fig. 5, it is clear that the overall performance of F-measure for the proposed system is better than that of the existing ANN and Fuzzy models.

Accuracy is a performance metric that measures the overall performance of the RS model. Accuracy results for the proposed RS model (modified k-means clustering and AGNN models) with combinations of existing k-means clustering approach, ANN and

Table 4 Overall F-measure value on different Cluster Size on proposed RS model: MovieLens Dataset

Number of clusters	Precision	Recall	F-measure
2	0.82	0.99	0.897016574585635
3	0.84	0.99	0.908852459016393
4	0.85	1	0.918918918918919
5	0.87	1	0.93048128342246

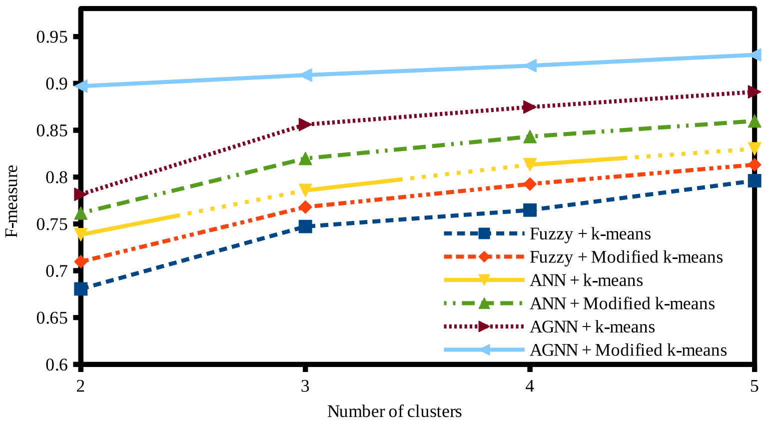

Fig. 4 F-measure for Proposed and Existing method of Clustering, AGNN, ANN and Fuzzy based RS model: MovieLens Dataset

Fuzzy models on various cluster sizes in the MovieLens dataset are shown in Fig. 6 which reveals that the combination of the proposed modified k-means with AGNN, ANN, and Fuzzy models perform better than k-means combinations. Also, it is concluded that the proposed RS model performs better than other models considered for the comparison.

Figure 7 shows overall accuracy measure for the modified k-means clustering with existing ANN, Fuzzy and proposed AGNN models on the MovieLens dataset. The proposed method achieves an accuracy value 81.56% whereas ANN and Fuzzy achieve 73.74% and 64.72% respectively. Figure 7 shows that the overall accuracy is improved for the proposed method than a combination of existing methods. It is also observed from Fig. 7, that the performance of RS increased when the cluster size increased.

Figures 8 and 9 show the convergence value of F-measure and Accuracy for the proposed and existing RS models using the modified k-means approach on the MovieLens dataset over many iterations respectively. F-measure and accuracy are measured for k-value as 5. From Figs. 8 and 9, it is clearly seen that the proposed AGNN model converges at 14^{th}

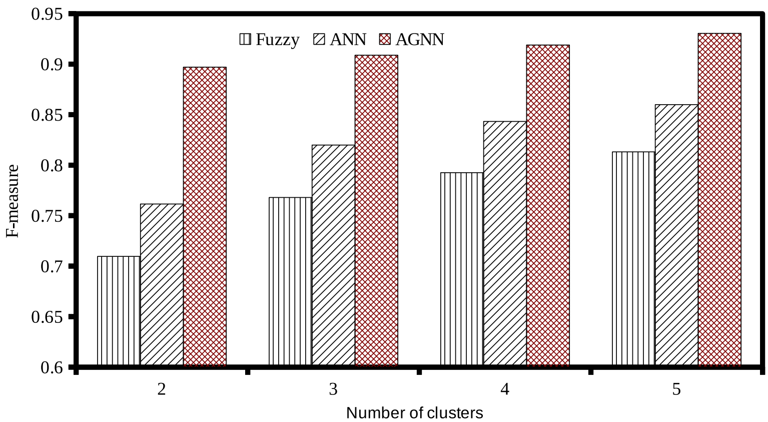

Fig. 5 F-measure values of RS models on different Cluster size: MovieLens Dataset

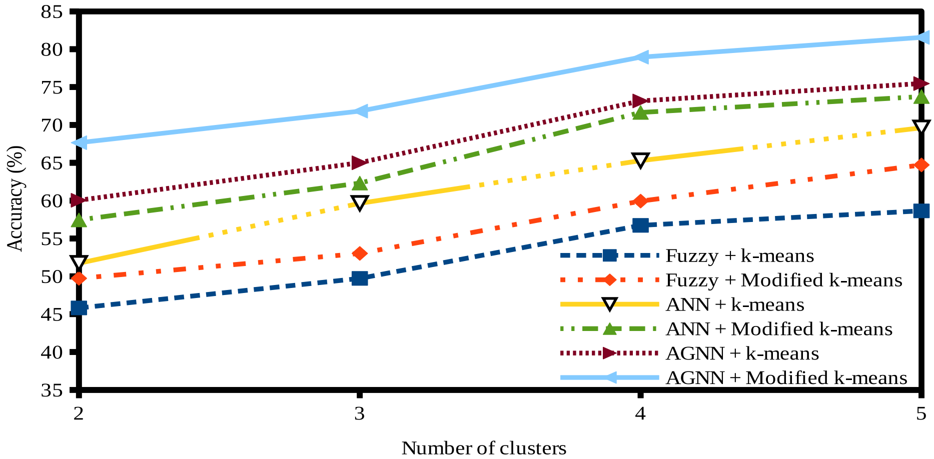

Fig. 6 Accuracy for Proposed and Existing method of Clustering, AGNN, ANN and Fuzzy Algorithms: MovieLens Dataset

iteration and ANN model converges at 30^{th} iteration. Also, F-measure and Accuracy value for AGNN are better than ANN.

The number of iterations for implementing ANN and AGNN based RS models on the MovieLens dataset using modified k-means approach by varying cluster size is given in Fig. 10. Figure 10 shows, that the number of iterations needed to obtain converged results for both AGNN and ANN methods are more for reduced cluster size values. That is, the number of iterations is reduced by increasing cluster size. Also, AGNN needs lesser iterations compared to the ANN based RS model.

Another important parameter that affects the RS model's performance is learning rate. The experimentation was done on the MovieLens dataset by varying the value for learning rate from 0 to 1. Figure 11 shows the convergence time for AGNN and ANN models by varying the learning rate. It is clearly understood from Fig. 11 that convergence time is less

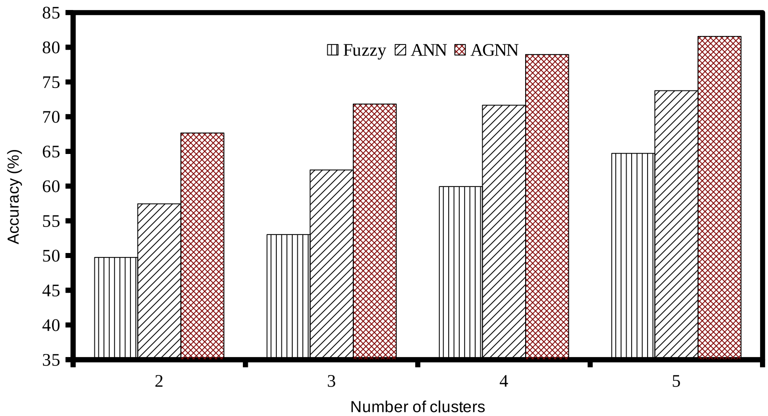

Fig. 7 Accuracy of AGNN, ANN and Fuzzy RS model using modified k-means Clustering Approach: MovieLens Dataset

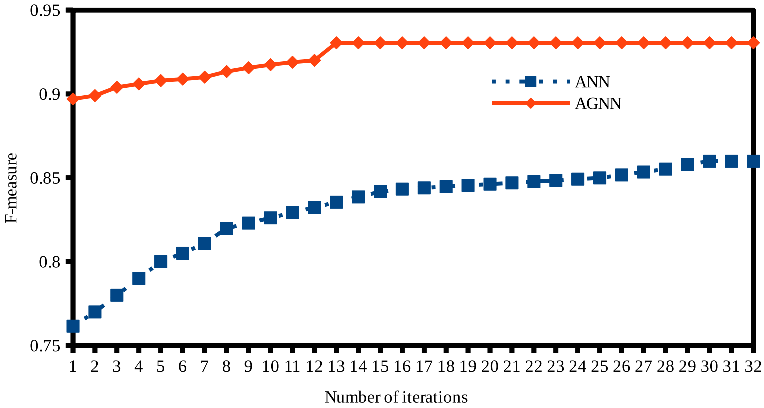

Fig. 8 F-measure of ANN and AGNN RS model using Modified k-means Approach based on number of iterations: MovieLens Dataset

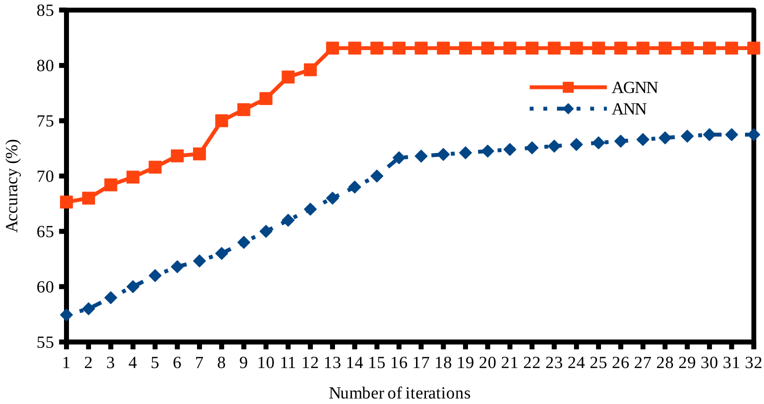

Fig. 9 Accuracy of ANN and AGNN using Modified k-means Approach based on number of iterations: MovieLens Dataset

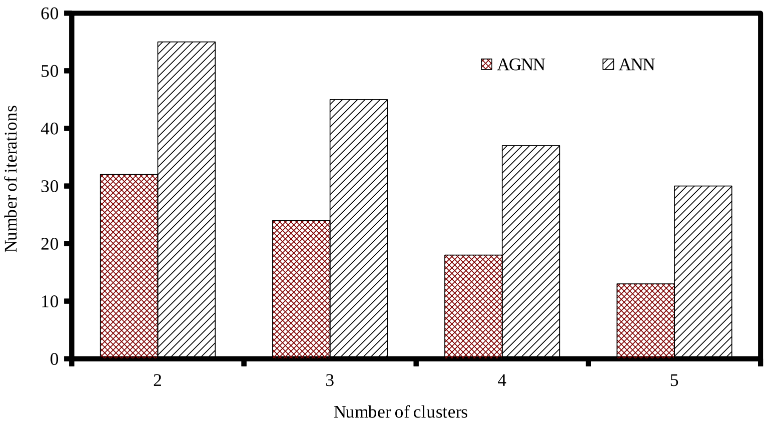

Fig. 10 Number of iterations for ANN and AGNN RS methods: MovieLens Dataset

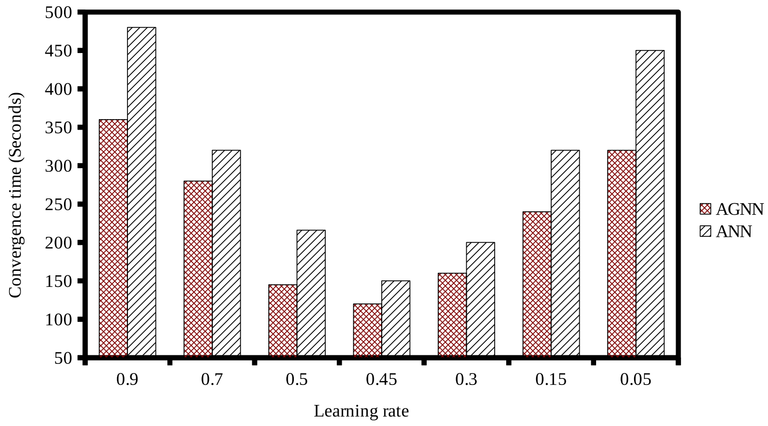

Fig. 11 Convergence time vs Learning rate of RS model: MovieLens Dataset

for AGNN and ANN based RS models which have a learning rate value 0.45, for the dataset considered for the experiment. It is also clearly shown from Fig. 11 that convergence time for the proposed AGNN is less compared to the existing ANN based RS model.

4.4 Performance analysis of AGNN for the Netflix dataset

Similar to that of the MovieLens dataset the performance of the proposed RS method that was evaluated using the Netflix dataset and its performance is compared with existing ANN and Fuzzy based RS models.

Table 5 shows the recommendation results on different cluster sizes using the proposed RS model for the Netflix dataset. From Table 5, it is clearly understood that the number of correct recommendations increase when cluster size increases. Precision, recall and F-measure values are calculated for the proposed RS model from the results in Table 5 and are shown in Table 6. A high value of precision represents fewer users getting the wrong recommendation and a higher recall value indicates improved recommendation for more users.

Overall F-measure value for all combinations of existing and proposed clustering techniques with ANN, Fuzzy and AGNN based RS models on the Netflix Dataset are shown Fig. 12. It is observed from the Fig. 12, that the proposed RS model (Modified k-means+AGNN) performs better compared to other combinations. Also, it is inferred, that the proposed modified k-means clustering approach gives better performance for all AGNN,

Table 5 Recommendation for different Cluster Size on Proposed Method: Netflix Dataset

Number of clusters	Number of users	Correctly Recommended	Wrongly recommended
2	100	76	24
3	100	79	21
4	100	81	19
5	100	83	17

Table 6 Overall F-measure value on different Cluster Size on proposed RS model:Netflix Dataset

Number of clusters	Precision	Recall	F-measure
2	0.76	0.96	0.848372093023256
3	0.79	0.97	0.870795454545454
4	0.81	0.98	0.886927374301676
5	0.83	0.98	0.89878453038674

ANN, and Fuzzy models for the Netflix dataset. GA weight values are optimized for maximum iterations or till it reaches termination. The AGNN model is back propagated till it reaches the stopping criterion.

Figure 13 shows the overall F-measure value using the proposed modified k-means clustering approach on the proposed AGNN, existing ANN and Fuzzy based RS models in the Netflix Dataset. F-measure value under cluster size 5 on the proposed RS model is 0.89 and 0.82 and 0.79 for existing ANN and Fuzzy based RS models respectively. It is inferred from Fig. 13, that the F-measure value of RS model increases with an increase in cluster size. It is also seen from the Fig. 13, that the overall performance of F-measure value on the proposed system is better compared to the existing ANN and Fuzzy based RS models.

Figure 14 shows accuracy results of the proposed RS model (modified k-means clustering and AGNN model) with combinations of existing k-means clustering approach, ANN and Fuzzy models on various cluster sizes for the Netflix dataset. It is seen from Fig. 14, that combinations of the proposed modified k-means with AGNN, ANN, and Fuzzy based RS models performs better than other k-means combinations and also that the proposed RS model performs better than other models considered for the comparison.

Overall accuracy value for the modified k-means clustering with existing ANN, Fuzzy and proposed AGNN based RS models on the Netflix Dataset are shown in Fig. 15 where accuracy values are 79.75%, 70.04, and 59.16% respectively. It is seen from the Fig. 15, that the overall accuracy of the proposed RS model is better compared to other combinations. It is also inferred from the same Fig. 15, that the performance of the RS model increases when cluster size increases.

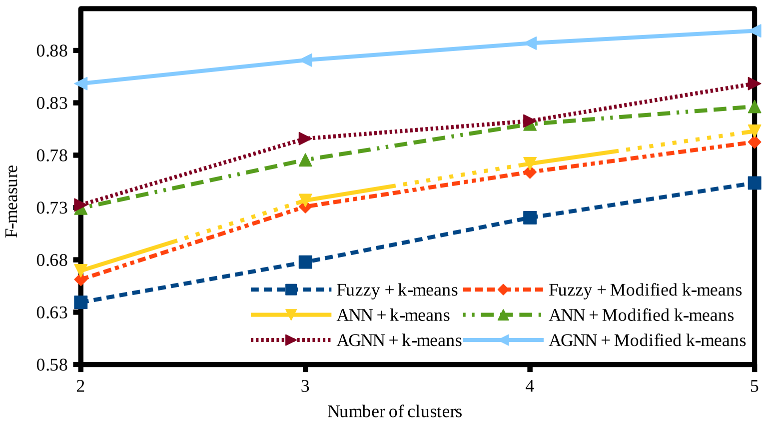

Fig. 12 F-measure for Proposed and Existing method of Clustering, AGNN, ANN and Fuzzy based RS model: Netflix Dataset

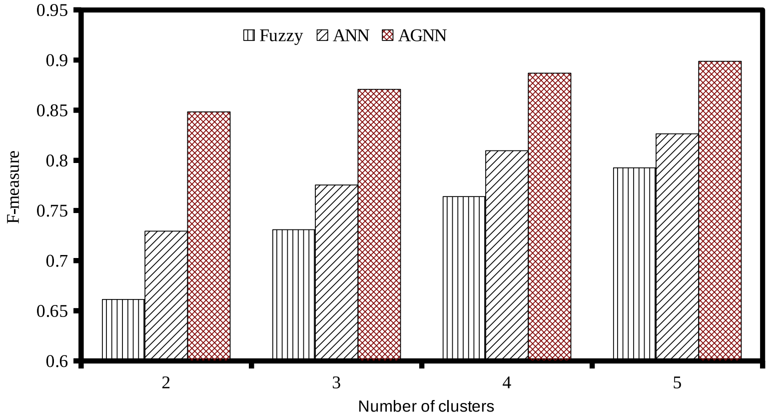

Fig. 13 F-measure values of RS models on different Cluster size: Netflix Dataset

Convergence value of F-measure and Accuracy for the proposed and existing RS models using modified k-means approach over varying iterations on the Netflix dataset are shown in Figs. 16 and 17 respectively. It is clearly observed from the Figures, that the proposed AGNN model converges at 16^{th} iteration while the ANN model converges at 36^{th} iteration on the Netflix dataset for k-value 5. Also, F-measure and Accuracy values for AGNN are better compared to ANN.

Figure 18 shows the iterations taken to obtain the converged results on ANN and AGNN based RS models with modified k-means clustering by varying cluster size from 1 to 5. It can be concluded from Fig. 18, that iterations for obtaining the converged result are more for less value of cluster size on both AGNN and ANN based RS models. Similar to the

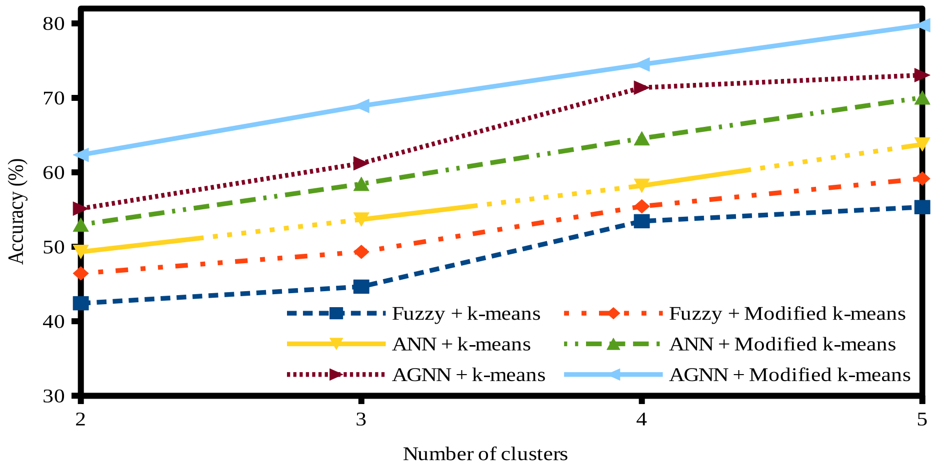

Fig. 14 Accuracy for Proposed and Existing method of Clustering, AGNN, ANN and Fuzzy Algorithms: Netflix Dataset

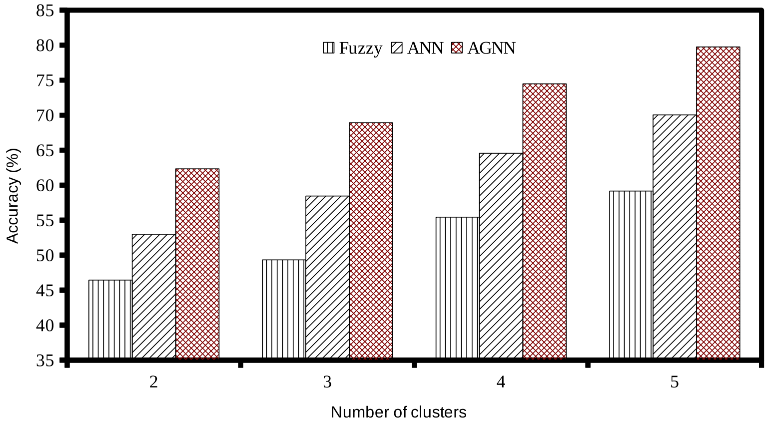

Fig. 15 Accuracy of AGNN, ANN and Fuzzy RS model using modified k-means Clustering Approach: Netflix Dataset

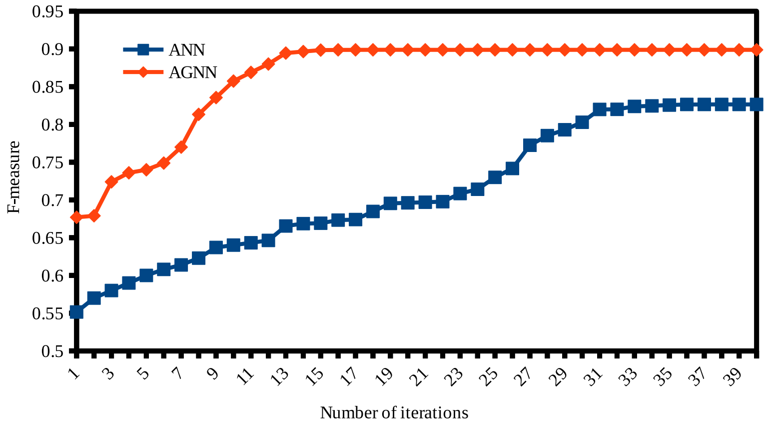

Fig. 16 F-measure of ANN and AGNN RS model using Modified k-means Approach based on number of iterations: Netflix Dataset

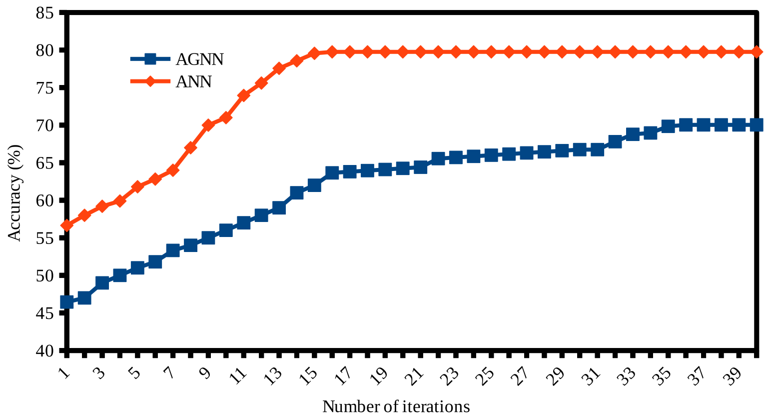

Fig. 17 Accuracy of ANN and AGNN using Modified k-means Approach based on number of iterations: Netflix Dataset

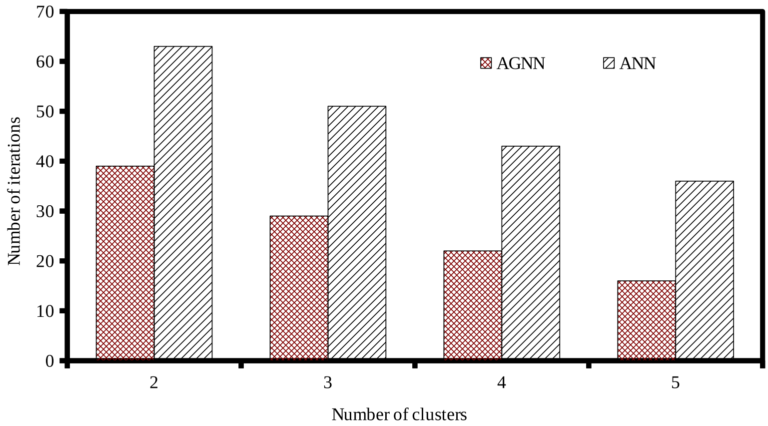

Fig. 18 Number of iterations for ANN and AGNN RS methods: Netflix Dataset

MovieLens Dataset the number of iterations is reduced by increasing cluster size. Also, AGNN takes lesser iterations compared to ANN based RS models.

The performance of the RS model is affected by another parameter called learning rate and an experiment was conducted on the Netflix dataset by varying the learning rate value from 0 to 1. The convergence time of AGNN and ANN based RS models regarding learning rate is shown in Fig. 19 which reveals that convergence time for both RS models is less with a learning rate of 0.45 on the Netflix dataset. It is also seen that convergence time for the proposed RS model is less compared to the existing ANN based RS model.

From the experimental results obtained for the MovieLens and Netflix datasets, it can be concluded that the proposed RS model obtains better recommendation results over the other models used in the comparison.

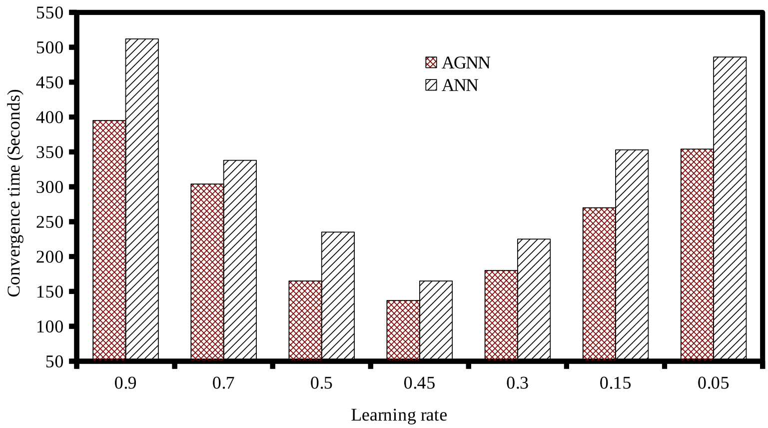

Fig. 19 Convergence time vs Learning rate of RS model: Netflix Dataset

5 Conclusion

This paper proposed a new Recommender System (RS) that originated with the Collaborative Filtering (CF) approach. Data points are clustered with minimal error rate using the proposed modified k-means clustering approach. The most favorable data points in the cluster are identified using the proposed Adaptive Genetic Neural Network (AGNN) model that provides effective recommendations to targeted online users. Using the benefits of AGNN with a modified k-means approach, the proposed RS reduces recommendation error rate drastically and ensures accurate recommendations. The performance of the proposed RS is evaluated using the standard benchmark MovieLens and Netflix datasets and the obtained results show its improved efficiency compared to other Artificial Neural Network(ANN) and Fuzzy based RS models.

References

1. Ahmed ME, Botvich D (2013) Multi-agent based middleware for protecting privacy in iptv content recommender services. Multimedia Tools Appl 64(2):249–275
2. Al-Shamri MYH, Bharadwaj KK (2008) Fuzzy-genetic approach to recommender systems based on a novel hybrid user model. Expert Syst Appl 35(3):1386–1399
3. Albanese M, Chianese A, d'Acierno A, Moscato V, Picariello A (2010) A multimedia recommender integrating object features and user behavior. Multimedia Tools Appl 50(3):563–585
4. Alhijawi B, Kilani Y (2016) Using genetic algorithms for measuring the similarity values between users in collaborative filtering recommender systems. In: 15th international conference on computer and information science (ICIS) IEEE/ACIS, pp 1–6, IEEE
5. Amatriain X, Pujol JM (2015) Data mining methods for recommender systems. In: Recommender Systems Handbook, pp 227–262. Springer
6. Amini M, Nasiri M, Afzali M (2014) Proposing a new hybrid approach in movie recommender system. Int J Comput Sci Inf Secur 12(8):40
7. Anand D (2012) Feature extraction for collaborative filtering: a genetic programming approach. Int J Comput Sci Issues 9(1):348–354. https://doi.org/10.17485/ijst/2016/v9i17/89936
8. Ar Y, Bostanci E (2016) A genetic algorithm solution to the collaborative filtering problem. Expert Syst Appl 61:122–128
9. Berry MJ, Linoff GS (2009) Data mining techniques. Wiley, New York
10. Bhaskaran S, Santhi B (2017) An efficient personalized trust based hybrid recommendation (tbhr) strategy for e-learning system in cloud computing. Clust Comput, pp 1–13
11. Bilge A, Polat H (2013) A scalable privacy-preserving recommendation scheme via bisecting k-means clustering. Inf Process Manag 49(4):912–927
12. 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
13. Braida F, Mello CE, Pasinato MB, Zimbrão G (2015) Transforming collaborative filtering into supervised learning. Expert Syst Appl 42(10):4733–4742
14. Christakou C, Vrettos S, Stafylopatis A (2007) A hybrid movie recommender system based on neural networks. Int J Artif Intell Tools 16(05):771–792
15. Cochocki A, Unbehauen R (1993) Neural networks for optimization and signal processing. Wiley, New York
16. da Silva EQ, Camilo-Junior CG, Pascoal LML, Rosa TC (2016) An evolutionary approach for combining results of recommender systems techniques based on collaborative filtering. Expert Syst Appl 53:204–218
17. Dooms S, Pessemier TD, Martens L (2015) Offline optimization for user-specific hybrid recommender systems. Multimedia Tools Appl 74(9):3053–3076
18. Dooms S, Pessemier TD, Martens L (2015) Online optimization for user-specific hybrid recommender systems. Multimedia Tools Appl 74(24):11297–11329

19. 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
20. Gupta A, Shivhare H, clustering SS (2015) Genetic algorithm based weighted similarity measure. In: 2015 international conference on computer, communication and control (ic4) recommender system using fuzzy c-means, pp 1–8, IEEE, p 2015
21. Gupta A, Tripathy BK (2014) A Generic hybrid recommender system based on neural networks. In: Advance Computing Conference (IACC) IEEE International, pp 1248–1252, IEEE, p 2014
22. Han J-W, Jo J-C, Ji H-S, Lim H-S (2016) A collaborative recommender system for learning courses considering the relevance of a learner's learning skills. Clust Comput 19(4):2273–2284
23. Hsu SH, Wen M-H, Lin H-C, Lee C-C, Lee C-H (2007) Aimed-a personalized tv recommendation system. In: European conference on interactive television, pp 166–174. Springer
24. Jia YB, Ding QQ, Liu DL, Zhang JF, Zhang YL (2014) Collaborative filtering recommendation technology based on genetic algorithm. In: applied mechanics and materials, vol 599, pp 1446–1452. Trans Tech Publ
25. 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
26. Katarya R, Verma OP (2016) A collaborative recommender system enhanced with particle swarm optimization technique. Multimedia Tools Appl 75(15):9225–9239
27. 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, pp V6–414. IEEE
28. Kim K-J, Ahn H (2008) A recommender system using GA k-means clustering in an online shopping market. Expert Syst Appl 34(2):1200–1209
29. Lee M, Choi P, Woo Y (2002) A hybrid recommender system combining collaborative filtering with neural network. In: International conference on adaptive hypermedia and adaptive web-based systems, pp 531–534. Springer
30. Li X, Wang Z (2017) A new recommendation algorithm combined with spectral clustering and transfer learning. Clust Comput 1–17. https://doi.org/10.1007/s10586-017-1161-4
31. Linden G, Smith B, York J (2003) Amazon. com recommendations: item-to-item collaborative filtering. IEEE Internet Comput 7(1):76–80
32. Liu X, Fu H (2014) Pso-based support vector machine with cuckoo search technique for clinical disease diagnoses. The Scientific World Journal, 2014
33. Lops P, De Gemmis M, Semeraro G (2011) Content-based recommender systems: state of the art and trends. In: Recommender systems handbook, pp 73–105. Springer
34. Lu J, Wu D, Mao M, Wang W, Zhang G (2015) Recommender system application developments: a survey. Decis Support Syst 74:12–32
35. Madadipouya K, Sivananthan C (2017) A literature review on recommender systems algorithms, techniques and evaluations. BRAIN Broad Research in Artificial Intelligence and Neuroscience 8(2):109–124
36. McCulloch WS, Pitts W (1943) A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys 5(4):115–133
37. Movielens 100k dataset. https://grouplens.org/datasets/movielens/100k/
38. Mukhopadhyay A, Maulik U, Bandyopadhyay S (2015) A survey of multiobjective evolutionary clustering. ACM Comput Surv (CSUR) 47(4):61
39. 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
40. Netflix dataset. http://www.netflixprize.com
41. Özkan C, Erbek FS (2003) The comparison of activation functions for multispectral landsat tm image classification. Photogramm Eng Remote Sens 69(11):1225–1234
42. Paradarami TK, Bastian ND, Wightman JL (2017) A hybrid recommender system using artificial neural networks. Expert Syst Appl 83:300–313
43. 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
44. Raja NSM, Vishnupriya R (2016) Kapur's entropy and cuckoo search algorithm assisted segmentation and analysis of rgb images. Indian Journal of Science and Technology 9(17)
45. Rashid SKLAM, Karypis G, Riedl J (2006) Clustknn: a highly scalable hybrid model-& memory-based cf algorithm. In: Proceeding of webKDD
46. Rosli AN, You T, Ha I, Chung K-Y, Jo G-S (2015) Alleviating the cold-start problem by incorporating movies facebook pages. Clust Comput 18(1):187–197

47. Salehi M, Pourzaferani M, Razavi SA (2013) Hybrid attribute-based recommender system for learning material using genetic algorithm and a multidimensional information model. Egypt Inform J 14(1):67–78
48. 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
49. Selvi C, Sivasankar E (2017) A novel optimization algorithm for recommender system using modified fuzzy c-means clustering approach. Soft Comput 1–16. https://doi.org/10.1007/s00500-017-2899-6
50. Standard relational dataset. http://storm.cis.fordham.edu/~gweiss/data-mining/datasets.html
51. Tsai C-F, Hung C (2012) Cluster ensembles in collaborative filtering recommendation. Appl Soft Comput 12(4):1417–1425
52. Tu JV (1996) Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. J Clin Epidemiol 49(11):1225–1231
53. 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
54. Zanardi V (2011) Addressing the cold start problem in tag-based recommender systems. Phd thesis UCL (University College London)
55. Valentina Zanardi, Licia Capra (2011) A scalable tag-based recommender system for new users of the social web. In: Database and expert systems applications, pp 542–557. Springer
56. 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
57. Zimmerman J, Kauapati K, Buczak AL, Schaffer D, Gutta S, Martino J (2004) Tv personalization system. In: Personalized digital television, pp 27–51. Springer

C. Selvi is a Ph.D. Research scholar in the Department of Computer Science and Engineering, National Institute of technology, Tiruchirappalli, Tamilnadu. Her current research interests include Data Mining, Recommender Systems, Sentiment Analysis, Credit Risk Analysis, Algorithms & Data structures and Machine Learning.

E. Sivasankar is an Assistant Professor in the Department of Computer Science and Engineering, National Institute of Technology, Tiruchirappalli, Tamilnadu. His current research focuses on Data Warehousing & Data Mining, Web Technology, Database Management Systems, Machine Learning and Big data Analytics.