2nd International Conference on "Innovative Trends in Science, Engineering and Management"
YMCA, Connaught Place, New Delhi
05 November 2016, www.conferenceworld.in
ICITSEM-16
ISBN : 978-93-86171-10-8

GENETIC ALGORITHM BASED COLLABORATIVE FILTERING MODEL FOR PERSONALIZED RECOMMENDER SYSTEM

P.Prabhu[1], S.Selvbharathi[2]

[1]Assistant Professor, Directorate of Distance Education,
Alagappa University, Karaikudi, Tamilnadu, INDIA
[2]Assistant Professor (P.T), Department of ECE,
Alagappa Chettiar College of Engineering and Technology, Karaikudi.

ABSTRACT

Recommender systems are one of the business intelligence information filtering systems which provide intelligence by suggesting suitable items according to the interest given by the active user. These systems suffer from challenges like information overload, scalability, sparsity, cold-start problem, accuracy and high dimensionality and still needs improvements. Genetic algorithm (GA) is method for solving optimization problems based on evolutionary ideas for finding a best recommendation out of a large population of attributes. This paper introduces novel genetic algorithm based model for collaborative filtering personalized recommender system for effective retrieval of information. Several experiments have been carried out to evaluate the performance of the model with different factors and evaluation measures using Movielens and Jester real-world datasets. The experimental results show that the proposed model outperforms the conventional models in terms of accuracy.

Keywords: *Business intelligence, Clustering, Genetic Algorithm, Information Filtering, Personalization.*

I. INTRODUCTION

Due to the growth of internet, web users, there is a increase in large volume of web data. Such a huge amount of data cannot be processed by the expert in a short period to make any business decisions. in less time. Therefore, data mining has become important to the e-business world and active users especially in e-commerce. An important step in the data mining is data pre-processing, where the relevant features are selected using various techniques. In e-commerce feature selection is an important step in selecting relevant items to the active users. Web content mining is extraction of knowledge from web. Recommender systems are business intelligence systems which provide suggestions to the end users in online.

Collaborative filtering is one of the web content mining techniques which provide business intelligence to the active users and business. The users knowledge about the data such as profile and preferences is used search relevant items as recommendations. GA is a stochastic general search method, capable of effectively exploring large search spaces, which is usually required in case of attribute selection. Further, unlike many search algorithms, which perform a local, greedy search, GAs performs a global search [1].

This work use GA for selecting the relevant features and genetic k-means clustering is used to find the nearest neighbours and users navigation patterns.

II. RELATED WORKS

Research in any field requires a highly structured review and study of related literature. A critical analysis of related literature will provide information on what has previously been done in the relevant area this will lead to new approaches and investigation.

Amit Verma, Harpreet Kaur Virk[2] proposed hybrid method which combines content, collaborative filtering method, additional content features and demographic information to generate recommendations. Genetic algorithm and k-nn method is used to construct the RS system. The algorithm is evaluated using recall, precision and f1 measure.

Christakou and Stafylopatis [3] proposed a hybrid content-based collaborative filtering recommender system. The content-based recommender is implemented using three neural networks per user, each of them corresponding to one of the following features: “kinds”, “stars”, and “synopsis”. The authors trained the ANN using the Resilient Back propagation neural network method.

Shiva Nadi et al. [4] discussed about the merits of content and collaborative filtering based techniques for generating recommendations. They proposed a fuzzy based recommender system by using collaborative behaviour of ants (FARS). This system functions in two phases namely modeling and recommendation. In the first phase, modeling is constructed in offline. In the second phase the recommendations are generated online based on the results from the first phase.

This literature review identifies that, there is a research gap in the filed of generating more accurate recommendations from the multidimensional datasets.

III. METHODOLOGY (GABM)

The proposed Genetic Algorithm Based Model (GABM) works in three sub phases’ namely pre-processing, modeling and intelligence. The figure 1 shows the architecture of GABM model.

ISBN : 978-93-86171-10-8

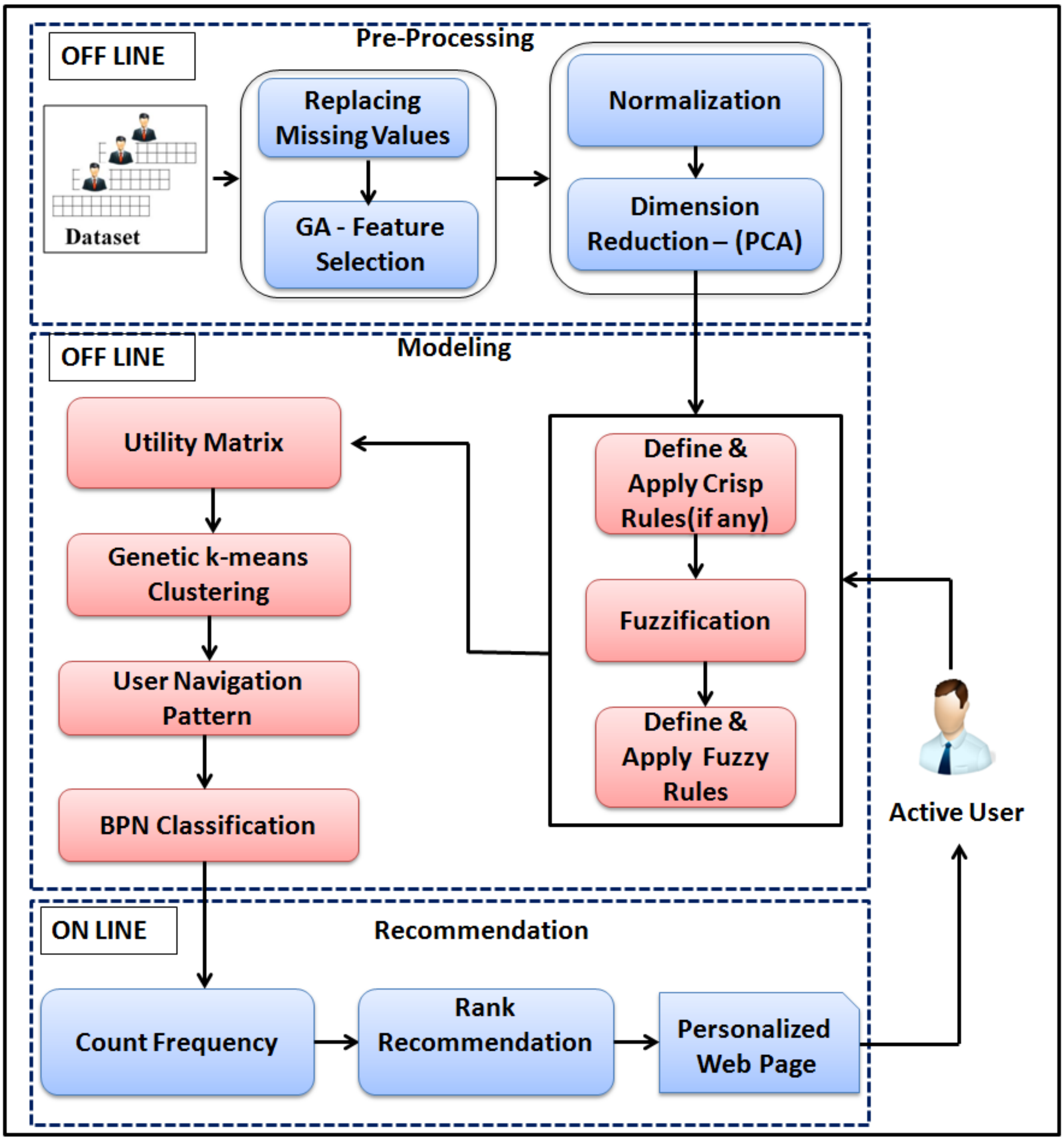

Fig 1. GABM Architecture

3.1 Pre-processing

The real-world datasets may contain inconsistent, irrelevant and missing values. These raw datasets lead to the loss of accuracy. Pre-processing is the process of removing and/or replacing inconsistent, irrelevant and missing values from the raw dataset. In this step missing values are replaced using neural networks. The relevant features are identified using genetic algorithm. Normalization is applied on datasets to optimize the recommendation performance. Principal Component Analysis (PCA) is applied to obtain the reduced representation of the dataset.

3.1.1 Feature selection using genetic algorithm

Genetic algorithms (GAs), a form of inductive learning strategy, are adaptive search techniques which have demonstrated substantial improvement over a variety of random and local search methods [5]. The feature selection is performed using genetic algorithm. The steps for feature selection are as follows;

Step 1: Set the number of desired features.
Step 2: Set the fitness function.
Step 3: Construction of the first population.
Step 4: ***Repeat***
Selection of parents from population
Apply Crossover
Apply Mutation
Until best feature is good enough

Algorithm 1. Genetic Feature Selection

3.1.2 Normalization

In order to optimize the recommender's performance, it is important to normalize the user-item matrix to computing the similarity matrix. Hence in this work the user-item matrix is normalized between 0 and 1 using min-max normalization technique.

3.1.3 Dimension reduction

The real world dataset consists of large volume of multidimensional sparse data. By reducing the dimension of dataset from high to low leads to reduction in sparsity. Dimensionality reduction techniques such as Principal Component Analysis (PCA), Singular Value Decomposition (SVD) and Decision Tree induction (DTI) help overcome scalability and sparsity problem by transforming the original high-dimensional space into a lower-dimensionality. Hence in this PCA is used to reduce the dataset from high to low dimension.

3.2 Construction of Model

In the second phase model is constructed offline. During this model construction phase the knowledge based fuzzy and crisp rules are defined and applied on utility matrix for information filtering. The Genetic k-means clustering method is applied to cluster filtered utility matrix to form neighbourhood of existing users. The active users Best Matching Cluster (BMC) is predicted using Back Propagation Neural Network (BPN).

3.3 Business Intelligence

The final phase in GABM is obtaining business intelligence by generating recommendations from the items of Best Matching Cluster (BMC) of active user. The frequency count of items from the identified BMC of users is calculated and ranked. Then, this model returns *top-N* items as recommendation that have not yet been purchased by the active user.

3.4 GABM Algorithm

This algorithm shows the pseudo code of the GABM model. The time complexity of this algorithm is O(n+K)+O(W^3)+ O(m^2+m) where K is total number of neurons. The space complexity of the algorithm is O(1).

Input: Training Dataset D and Test dataset TD;
The number of clusters *k*.
N = Potential number of recommendation.

Output: Recommendation List $\{I_1, I_2, \ldots, I_n\}$ of *Top-N* items.

// Phase I: Pre-processing

Replace missing values using NN.(if any)
Select relevant features using GA Feature Selection.
Perform Normalization.
Do Dimension Reduction using PCA.

// Phase II: Model Construction

Define KB crisp rules (if any) for user preferences and/or profiles.
//Formulate the fuzzy rules (FR).

- Define the linguistic variables and terms (initialization).
- Construct the membership functions (initialization).
- Construct the rule base (initialization).

Convert crisp input data to fuzzy values using membership functions.
// Fuzzification of attributes/user preferences.
$compute\ \mu_s(a_j)$ where j = 1 to m.
// Fuzzification of samples/user profile features.
$compute\ \mu_s(p_j)$ where j = 1 to m .
Define KB fuzzy rules for user preferences and/or profiles
Apply rules and generate the resultant dataset.
If New User **// Cold-Start Problem**
Register and Login.
Go to Step 2 of Phase III.
Else
Clustering of utility matrix using GA-kmeans
For each Active user in TD do // Predicting BMC
Find the Best Matching Cluster (BMC) using RBPN.
End

ISBN : 978-93-86171-10-8

TMAE ← Evaluate Matching Cluster

End If

//Phase III: Recommendation

For each Active user in TD **do**

Identify items from Best Matching Cluster (BMC) of users.

Calculate the frequency count and rank the items.

Select and recommend *top-N* items.

End

Algorithm 2. GABM

IV. RESULTS AND DISCUSSION

Movielens and Jester dataset is used for experimental evaluation [13][14]. Relevant features are selected using GA before finding nearest neighbours using GA-k-means clustering. This algorithm is implemented using MATLAB. Various factors with range of dimensions, clusters, *top-N* value, similarity and test/train size is considered to evaluate the performance of this model. Two metrics widely used in the information retrieval (IR) community namely recall and precision [6] are used to evaluate the model.

$$\text{Recall} = \frac{\text{size of hit set}}{\text{size of test set}} = \frac{|\text{test} \cap \text{top} - \text{N}|}{|\text{test}|}$$

$$\text{Precision} = \frac{\text{size of hit set}}{\text{size of top} - \text{N set}} = \frac{|\text{test} \cap \text{top} - \text{N}|}{\text{N}}$$

The standard F1 metric that gives equal weight to recall and precision and are computed as follows:

$$\text{F1} = \frac{2 * \text{Recall} * \text{Precision}}{(\text{Recall} + \text{Precision})}$$

The results of GABM model is compared with conventional k-Nearest Neighbour (k-NNBM) Approach[7][8][9][10] using MovieLens and Ant Based Recommender System (ARS) [11], k-nearest neighbour based Mean Squared Distance (MSD-CMB)[12] Jester datasets with commonly compared top-N value as 10.

The Table 1 shows the comparison of Mean F1 measure between KBFNM model and conventional models using MovieLens dataset for *top-N*=10.

TABLE 1. Mean F1 of GABM with conventional methods using MovieLens.

Algorithm	k-NNBM	k-NNBM(w)	GAC	k-NNBM(PCC)	GABM
F1 Measure	**0.44**	**0.53**	**0.60**	**0.66**	**0.78**

When compared with k-NNBM, k-NNBM (w), GAC and k-NNBM (PCC) algorithms 12% increase in F1-measure is obtained using GABM model. Therefore the performance of the GABM is more significant than the conventional methods. The Table 2 shows the comparison of mean F1 measure between GABM model and conventional models using Jester dataset for *top-N*=10.

TABLE 2. Mean F1 of GABM with conventional methods using Jester.

Algorithm	ARS	MSD-CMB	GABM
F1 Measure	0.49	0.68	0.89

The results listed in the Table 2 shows that, when compared with ARS and MSD_CMB algorithms GABM model outperformed using Jester dataset. Therefore the performance of the GABM is more significant than the ARS and MSD_CMB methods since it gives more recommendation accuracy measured in terms of F1 measure. This research work is useful to reduce the search time of online users, improve the loyalty of the business and design and develop efficient RS models for the Information Technology (IT) professionals.

V. CONCLUSION

One of the most important applications of the recommender systems is in the filed of e-commerce. However, since the repository of e-commerce items is very massive and these items have several features, when applying the existing recommendation algorithms, there are some problems such as sparsity, reliability, cold-start and scalability. The proposed method discovers and optimizes latent features by GA and generates recommendation using collaborative filtering. The experiment results show that the proposed approach performs better than conventional approaches in the terms of accuracy measures measurements and also can alleviate sparsity and scalability problem. Optimization techniques such as Particle swarm Optimization (PSO) and Ant Colony Optimization (ACO) can be considered as a future research work.

REFERENCES

[1] Priyanka khare, Kavita Burse, Feature Selection Using Genetic Algorithm and Classification using Weka for Ovarian Cancer, International Journal of Computer Science and Information Technologies, Vol. 7 (1) , 2016, 194-196,pp. 194-196.

[2] Amit Verma, Harpreet Kaur Virk, A Hybrid Recommender System using Genetic Algorithm and kNN Approach, *International Journal of Computer Science and Technology Vol. 6, Iss ue 3, July - Sept 2015,pp.131-134.*

[3] C. Christakou and A. Stafylopatis, A hybrid movie recommender system based on neural networks. In ISDA ’05: Proceedings of the 5th International Conference on Intelligent Systems Design and Applications, pp. 500–505, 2005.

[4] Shiva Nadi, Mohammad. H. Saraee, Mohammad Davarpanah Jazi3and Ayoub Bagheri, FARS: Fuzzy Ant based Recommender System for Web Users, IJCSI International Journal of Computer Science Issues, ISSN (Online): Vol. 8, Issue 1,pp.1694-0814, , January 2011.

[5] De Jong, K. “Learning with Genetic Algorithms: An overview,” *Machine Learning Vol. 3, Kluwer Academic publishers, 1988.*

[6] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl, Analysis of Recommendation Algorithms for Ecommerce,ACM,EC’00,October 17-20,2000.

[7] Deepa Anand, Enhancing Accuracy of Recommender Systems Through Various Approaches to Local and Global Similarity Measures, *PhD Thesis, Jawaharlal Nehru University, July 2011*.

[8] Hema Banati and Shikha Mehta, A Multi-Perspective Evaluation of MA and GA for Collaborative Filtering Recommender System, *International journal of computer science & information Technology (IJCSIT) Vol.2, No.5, pp. 103-122,October 2010.*

[9] Panagiotis Adamopoulos and Alexander Tuzhilin, Recommendation Opportunities: Improving Item Prediction Using Weighted Percentile Methods in Collaborative Filtering Systems, *ACM 978-1-4503-2409-0/13/10, RecSys’13, October 12–16, 2013.*

[10] Wang, W., Zhang, G. and Lu, J., Collaborative Filtering with Entropy-Driven User Similarity in Recommender Systems. *Int. J. Intell. Syst., 30: 854–870. doi:10.1002/int.21735,2015.*

[11] P.Bedi, R.Sharma, H.Kaur, Recommender system Based on collaborative behaviour of Ants, *Journal of Artificial Intelligence*, *ISSN 1994-5450,Asian Network for Scientific Information, 2 (2):pp.40-55,2009.*

[12] Symeonidis, P. Alexandros Nanopoulos, Apostolos N. Papadopoulos, Yannis Manolopoulos, Collaborative recommender systems: Combining effectiveness, *Expert Systems with Applications, doi:10.1016/j.eswa.05.013*, 2007.

[13] The MovieLens dataset. Available: http://www.movielens.umn.edu.

[14] The Jester data set. Available: http://goldberg.berkeley.edu/jester-data/.