A Firefly Approach to Collaborative Filtering based Recommender Systems through Fuzzy Features

Krishan Kant Yadav, Venkatadri Marriboyina, Sanjiv Sharma

***Abstract:** Recommender system (RS) is most important methods which offer the recommendation to the online user with ease to make his right decisions on items or services. The User-based Collaborative Filtering (CF) technique is one of mainly important method amongst various recommender systems.Collaborative Filtering (CF) approaches are either model-based/ memory-based. While the previous is more precise, it's not flexible in compare of model-based approach. Here we proposed a hybrid fuzzy-firefly method to RS, which maintain the precision of memory considered as CF & scalability of model considered as CF. Utilizing the hybrid characteristics, new user model (UM) has been created, which assisted in reaching vital reduction in system difficulty, sparse & create the grip of neighbour transitivity association. UM is working to discovery group of compatible clients in which a memory-based hunt is performed. Experimental results on Movie Lens dataset shows that proposed method not only improves recommendation accuracy significantly but also increases quality of prediction and recommendation performance.*

***Index Terms:** Collaborative filtering; fuzzy logic; Firefly algorithm; Recommender systems.*

I. INTRODUCTION

Due to the massive utilization of internet, there are large amount of data and web documents that are almost impossible to manage by the use of traditional tools. The reason of these huge & diversity of web contents, there is an immediate necessity of a forceful automated web personalization device which could help us to understand massive amount of data in beneficial information & knowledge. Personalized recommender system (RS) is an efficient device to get information on the burdened issue [1]. The most fruitful instance of Web personalization devices is the web RS which reduces information burden & guides the clients in personalized approach to stimulating things inside extremely huge storage of conceivable choices [2]. RS are utilized in many areas like movies, news, books, music, search queries and products in general. Since mid of 90s, many researches has been done on recommender method &consequently significant growthis achieved in this field. Recommender systems provide personalized recommendations according to user needs by eliminating irrelevant items and proposing the most interesting items according to user preferences. Therefore, RSs save the required time for searching. Recommender systems have four techniques that suggest user i.e. content based filtering (CBF), demographic filtering (DMF), collaborative filtering (CF) & hybrid filtering systems [3]. DMF labelled the user considered on user's personal attributes & creates approval on the basis of demographic sections, while CBF advises previously identical items to the users liked. CF is broadly utilized for RS filtering. It offers recommend through examining the rating details of items / the clients. Because of its ease, productivity & capability to generate precise & personalized recommendations, CF is considered to be a major technique in RS. In our work we are incorporating fuzzy logic in recommender system. Fuzzy logic was presented via LotfiA Zadeh, where fuzzy sets are determined through the membership function (MF) values lies between 0 & 1 [8]. Although many researchers have introduced fuzzy logic in different directions but some of them are still unlisted. Nasraoui et al. [9] the fuzzy approximate analysis is used to create common system for approval method, whereas the relational fuzzy subtractive clustering method is used by the Suryavanshi [10] used. Shahabi et al. [5] introduced Yoda RS, which gradually classifies the active user depends on specific samples of clients & it gives soft commendations for it. User profiles have many features which is terms as fuzzy. But it is tough to fuzzify the profile at the item level, since it will need prohibitively huge space & lengthy processing period. Al-Shamri et al. [7] made a fuzzified hybrid model that fuzzy distance measurement has been presented to calculate the similarities between the user profiles. Another incorporation that yields in our work is learning the optimal weights on many features [11]. In general, each user gives multiple preferences on numerous features; some clients provide extra prominence on specific characteristics, whereas others clients show no interest in certain features. Many attempts have been made in the previous to include numerous evolutionary methods [12] in RS to learn optimal weights on it for several characteristics. Al-Shamri et al. [7] a hybrid fuzzy-genetic RS was developed through the deployment of genetic algorithm (GA) of progress proper weights speciality of each users. Likewise, Ujjin et al. [13] working a genetic algorithm to know the user's personal favourites & additional they enhanced it by Particle swarm optimization (PSO), to know that favourites & outcome was equated to those found in the GA recommender system. In the end, they summarized that the PSO works very fast compared to GA. In the performance, we evolve the fuzzy-firefly CF (FF-CF) through employing the firefly method to discover optimal individual preferences for multiple features like age, rating, gender etc. Afterward, finding the appropriate weights for multiple characteristics, we calculate adequate likenesses among clients & produce suitable recommendations for clients.

Revised Manuscript Received on May 10 ,2019

Krishan Kant Yadav, Research Scholar, Department of Computer Science & Engineering, Amity University, Gwalior, Madhya Pradesh, India.
Dr. Venkatadri Marriboyina, Professor, Department of Computer Science & Engineering, Amity University, Gwalior, Madhya Pradesh, India.
Dr. Sanjiv Sharma,, Professor, Department of CS & IT, Madhav Institute of Technology & Science, Gwalior, Madhya Pradesh, India.

Remaining part of this research has been prearranged as follows: Current studies related to collaborative filtering have been observed in section 2 & provides fundamental overview to firefly algorithm. In the Section 3, it defines our implemented fuzzy-firefly CF & give comprehensive explanation of how organization services firefly for CF. In the Section 4, set of the data & the outcomes of the appraisal. In the end, Section 5 offers closing remarks & advises some upcoming research direction.

II. BACKGROUND

A. Collaborating Filtering RS

The CF methods could be further labeled into memory-based &model based technologies [4]. In model-based technology, a model offline with data used for online recommendations [5] &memory based technology is used to recommends whole data [6]. Memory considered as method are extra precise, except it agonize from issue of the scalability, while in the form of scalability the model considered as methods are further scalable but less precise. The model was also evolved to maintain scalability of model considered as CF & correctness of the model based CF [7].

Normally, in the CF recommenders, it take group of the users U = {u_1, u_2,.....,u_m} rating group of objects S = {s_1, s_2,....., s_n}, like movies, CDs/ books. Spaces S & U are huge & could be extremely huge in few applications. To every user u_i, i = 1, 2,....,m takes rated the subset of objects S_i. Exactly, user u_c rating for object s_j, j = 1, 2, . . . ,n is represented via$r_{c,j}$. All achieved ratings are together in m × n user object, matrix represented via R.

In CF, following steps are required for recommendation:

Phase1. Data gathering

Phase2. Formation of user profile

Phase3. Generation of neighbourhood set

Phase4. Prediction and recommendation

Phase1: Data collection

Datasets are collected based on explicit and implicit ratings of the user. For our experiment we have composed 3-kinds of data from the clients, demographical data by the registering method, clear ratings for subsection of the offered objects, & underlying data from the online performance of the users.

Phase2: Formation of user profile

The profile of the client is only a gathering of private information of user. User could be modelled considered as user's information that is stored in the user profile [14]. Information of the profiling could be removed as of demographical data (example, gender, age of the clients, profession, etc.), clients favourites about the features of the objects (example, film name, director, category, release year, main lead thespians, etc.), & user ratings over the skilled objects (example, formerly viewed films) [15].

Phase3: Generation of neighbourhood set

Once the user models are installed, the system could similar an active user for the existing database for neighbourhood group. The item-based similarity matrix is used for forecast that creates representation of things likenesses through salvaging overall objects rated via an active clients viauser things matrix, then find out, alike salvaged objects are to the goal object, after then it chooses the k alike objects & their analogous likenesses are also identified. Forecasts are done through considering weighted average of online clients rating on alike objects k. Varieties of likeness measures are used to calculate the similarities among the item or clients. Measures of two most recommended likenesses are considered as correlation & cosine. The Pearson correlation coefficient (PCC) is utilized to compute that range, where two variables lay linear to one another & it is described as

$$sim(u,v) = \frac{\sum_{i\epsilon C}(r_{u,i} - \bar{r_u})(r_{v,i} - \bar{r_v})}{\sqrt{\sum_{i\epsilon C}(r_{u,i} - \bar{r_u})^2}\sqrt{\sum_{i\epsilon C}(r_{u,i} - \bar{r_v})^2}}, \quad (1)$$

where *C* is group of objects ranked via both clients *u* &*v*. Since the equation (1), receives data of general objects for both clients, this isn't appropriate if other cited attributes are also involved in the prototype. Therefore, alternate method to calculate likeness is the improved Euclidean distance function (2), that considers many attributes

$$d(U,V) = \frac{1}{x}\sum_{i=1}^{x}\sqrt{\sum_{j=1}^{m}\left(u_{i,j} - v_{i,j}\right)^2} \quad (2)$$

It shows$u_{i,j}$is j^{th} attribute in general object C_i, *m* is no. of attributes, &*x* = |*C*| is cardinality of *C*.

Phase4: Prediction and recommendation

Here, when all the neighbors are found, say k, prediction value is computed by applying numerous methods to join the neighbor's ratings on the hidden object for active clients. Next, forecasting how an active client would prefer definite objects, whichare not yet rated yet via active clients, the top-N object group identifies group of objects ordered with highly forecast value & is suggested.Predicted rating, $pre_{u,i}$, of object *i*of client's *u* is calculated through subsequent structure

$$pre_{u,i} = \bar{r_u} + k\sum_{u'\epsilon C} d(u,u') \times \left(r_{u',i} - \bar{r'}_u\right), \quad (3)$$

Where*C* symbolize group of neighbours that rated the object. Multiplier *k* is a normalizing factor & it is generally chosen

$$k = 1/\sum_{u'\epsilon C} |\, d(u,u')|$$

useru'. $have$ the average rating is$\bar{r'}_u$

B. Firefly Algorithm

The Firefly method is consisting as swarm intelligence algorithm, an evolutionary model inspired by social performance & nature [16]. It is used to resolve optimization difficulties. Fireflies produce short and rhythmic lights that their light patterns are different from each other. Population of algorithm is in fact fireflies, each of which has some lighting or fitness characteristics. In this procedure, fireflies are equated with each other & firefly, which is a fewer attractive moves toward the more attractive firefly. For easiness, the following 3- rules, we can make these flashing features ideal.

• All the fireflies are unisex so that according to their gender a firefly is concerned to different fireflies;

• Attraction is comparative to brightness, therefore for several 2- flashing fireflies, not as much of bright will shift in direction of a bright. Attraction is comparative to brightness &both of themreduce when their distance rises. Whether it is not brighter as a specific firefly, then it transfers arbitrarily;

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

• The illumination/ light force of firefly is influenced / resolute through objective purpose for landscape, which is adapted.

Attractiveness is a relative parameter and from the view of other fireflies is measured and it depends on distance of fireflies from each other. Attractiveness changes by distance according to the following equation:

$$\beta = \beta_0 . e^{-\gamma . r^2} \quad (4)$$

Where $\beta_0$is the maximum of attractiveness in interval [0,1], $\gamma$is absorption factor in [0, ∞), &thedistance between two fireflies is r. there is distance $r_{i,j}$is Cartesian distance among several of 2-fireflies *i*&*j* at x_i&x_j, correspondingly. It is obtained from the following equation:

$$r_{i,j} = \sqrt{\frac{\sum_{k=1}^{d} (x_{i,k} - x_{j,k})^2}{d_{max}}} \quad (5)$$

Here $x_{i,k}$is *kth* constituent of spatial coordinate $x_i$of *ith*firefly. Speed of firefly *i*, which attracted to alternative extraattractive (brighter) firefly *j*, is calculated as follows:

$$x_i = x_i + \beta_0 . e^{-\gamma . r_{i,j}^2} (x_j - x_i) + \alpha . e \quad (6)$$

Where e=delta*rand

Delta= $0.05 \times (Var_{max} - Var_{\min})$

$$d_{max} = (Var_{max} - Var_{\min}) \times \sqrt{21}$$

Standard Firefly algorithm pseudo code is given below [17]

```
Begin
    1. Objective purpose f(x), = (x1,x2,…x4)^T
    2. Make preliminary population of fireflies xi (i=1,2,…,n)
    3. Light intensity Ii at xi is define through f(xi)
    4. Illustrate light absorption coefficient γ

While (t > Max Generation )
    for i=1:n ; all n fireflies
   for j=1:i ; all n fireflies
      if ( Ij> Ii), move firefly i towards j in d-dimension
           end if
         Attractiveness varies via distance r via exp [− γ, r]
    Evaluate latest solution & modernize light intensity
   end for j
    end for i
Rank fireflies & search present finest
end while
Post process output & visualization
End
```

III. Proposed Methodology

The implemented recommender system is considered as fuzzy logic & firefly algorithm. In this research, we create a group of some hybrid structures which joins the users' & items' characteristics. In our proposed work, we have 3 phases to recommend using firefly algorithm.

Phase1: User information profile formation.

Phase2: Generation of neighborhood set.

Phase3: Predictions and recommendations.

Phase1: User information profile formation

According to this set of films, we get a hybrid structures based over the user's ratings for the content descriptions of the style & a set of high-quality pictures. For preparing a genre interestingness measure (GIM) as a basis, the hybrid structures are used. When we take a suitable structure for GIM, user prototype could be created via the DMF client profile & GIMs. Hybrid structures are utilized as the base for the interest measurement of the genre, therefore the client is more involved in G_i if its rating is greater, so it is good, very good/excellent. Numeral of hybrid structures based on numeral of genres. GIM can b calculated as

$$GIM(a,j) = \frac{2 \times N \times RGR(a,j) \times MRGF(a,j)}{RGR(a,j) + MRGF(a,j)}, \quad (7)$$

Wherever MRGF has improved comparative genre occurrence of genre$G_j$to the client u_a ,which is articulated

$$MRGF(a,j) = \frac{\sum_{g\epsilon G_{j\subset C_i}} \delta_3(r_{a,g}) + 2\times\delta_4(r_{a,g}) + 3\times\delta_4(r_{a,g})}{3\times TF(a)}, \quad (8)$$

RGR (comparative genre rating), is a proportion of u_i's rating to the highly rated objects of G_j , that is calculated as its overall rating

$$RGR(a,j) = \frac{\sum_{g\epsilon G_{j\subset C_{i\geq 3}}} r_{a,g}}{TR(a)}, \quad (9)$$

N is normalization feature of the provided system. TF & TR are full occurrence & rating complete, after this

Phase2: Generation of neighborhood set

Our profile is dependent on the fuzzy sets, that is utilized to the subsequent formulation that is introduced through [7].Distance between 2-fuzzy sets / points are talk about in [18]. Equality between two users can be calculated

$$Gfd(U,V) = \sqrt{\sum_{f=1}^{21} w_f \times \left(Lfd(x_f, y_f)\right)^2}, \quad (10)$$

Where Lfd is local fuzzy distance and w_f is the weight for the f[th] feature. The weight in our proposed work is learned using firefly algorithm. Following are the steps for incorporating firefly algorithm.

- **Initializing population**

Here, we are using firefly algorithm to learn the 21 feature weight of the population. Each feature weight is denoted via 8-bit binary numbers &weight ranges among 0 to 255. In this paper, we have reserved 10 as a people size comprising 10 fireflies.

- **Fitness Function**

Fitness function is calculated as the avg gap between the real & estimated ratings overall films in set training for that set of weights.

$$fitness = \frac{1}{t_R} \sum_{j=0}^{t_R} |r_j - pre_j|, \quad (11)$$

where $t_R$is termed as cardinality of training group for the active client &pre_jdefined as an item predicted rates *j* of a client in the training set.

IV. Experimental Results

The experiment is performed using MATLAB R2018a simulation tool on Movie Lense Dataset. Based on Movie Lens dataset we only accommodate clients whoever rated minimum 60 films, 20 to create a user model, & 40 for tests. Only 497 users out of 943 users completed this condition & donated 84,596 rates from 100,000. This dataset is utilized on the basis to create 5-arbitrary separations into training & active users.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

For each arbitrary separation, 50 users were selected randomly as active clients, & 447 clients are remaining to consider as users to train – as for RS to the historical data. Such an arbitrary parting was envisioned to perform five-fold cross-validation, in which overall researches are frequent 5-times, once with each division. These divisions aredevoted to as split-1, split-2 to split-5. Group of clients training (447 clients) is utilized to discover a group of neighbours to the active clients, whereas a group of active users (50 users) is utilized to analysis presentation of methods. While analysis stage, every active user's ratings are separated randomly into two disjoint sets, training ratings (34%) and test ratings (66%). The training ratings are used to model the user and to supervise the learning process of FGRS, whereas the test ratings are treated as unseen ratings that the system would try to forecast. The 2estimation metrics are utilized to estimate efficiency of various RS.

- Mean absolute error (MAE).
- Total coverage of system.

For the active clients $u_i$is provided through the subsequent formulation, MAE:

$$MAE(i) = \frac{1}{t_i}\sum_{j=1}^{t_i}\left|pre_{i,j} - r_{i,j}\right|, \qquad (12)$$

The coverage is given by

$$coverage = \frac{\sum_{i=1}^{T_n} q_i}{\sum_{i=1}^{T_n} t_i} \qquad (13)$$

Where $q_i$is total numeral of predicted stuffs &$t_i$is cardinality of examination rates a group of users u_i.

- **Result Analysis**

To determine the performance of our implemented FF-CF approach, the results are compared with FPSO-CF on the basis of MAE and coverage. In Table 1, performances of these schemes for each fold for 50 users. The MAE for FF-CF is always minor than the respective values for FPSO-CF, whereas the coverage value is always superior for all the divisions.

Table 1: Total MAE & Coverage.

MAE			**COVERAGE**	
Folds	FPSO-CF	FF-CF	FPSO-CF	FF-CF
1	0.8033	0.7887		
2	0.7954	0.6557	0.9654	0.9706
3	0.7892	0.7391	0.9546	0.9867
4	0.8321	0.8469	0.9564	0.9937
5	0.7833	0.6794	0.9424	0.9635
			0.9579	0.9868

V. Conclusion

In this paper we have proposed fuzzy firefly collaborating filtering (FF-CF) approach for recommendation. In our work, we create a group of hybrid geographies which joins few clients & objects properties. To order to deal with the imprecise nature of user feature, it is prepared fuzzy sets to efficiently represent user features. We have provided optimal similarity in order to increase speed of finding nearest neighbours of active user and reduce its computation time. Experimental outcomes demonstrate superiority of the implemented method (FF-CF) in terms of MAE, coverage & accurate forecasts. Apart from this, comparison with existing approach shows that our proposed approach is significantly faster.

REFERENCES

1. Linden G, Smith B, York J , "Amazon. com recommendations:item-to-item collaborative filtering". IEEE Internet Comput 7(1):76–80, 2003.
2. Burke, R., "Hybrid recommender systems: Survey and experiments".November 2002, Volume 12, Issue4, pp 331–370, 12, 331–370, 2002.
3. D. Jannach, M. Zanker, A. Felfernig and G. Friedrich, "Recommender Systems AnIntroduction," Cambridge University Press, Printed in the United States of America, 2011.
4. Su X, Khoshgoftaar TM, "A survey of collaborative filtering techniques". Advances in Artificial Intelligence, 2009.
5. Shahabi, C., Banaei-Kashani, F., Chen, Y., & McLeod, D. Yoda:, "An accurate and scalable web-based recommendation systems". In Proceedings of the sixth international conference on cooperative, 2000.
6. Resnick, P., Iakovou, N., Sushak, M., Bergstrom, P., & Riedl, J. GroupLens: , "An open architecture for collaborative filtering of Netnews". In Proceedings of ACM conference on computer supported cooperative work (pp. 175–186). NC: Chapel Hill,1994.
7. MYH. Al-Shamri and KK. Bharadwaj, "Fuzzy-Genetic Approach to Recommender Systems Based on a Novel Hybrid User Model", ExpertSystems with Applications, vol. 35, pp. 1386–1399, 2008.
8. LA. Zadeh, Fuzzy Sets, Information and Control, vol. 8, pp. 338–353, 1965.
9. O. Nasraoui and C. Petenes, "An Intelligent Web Recommendation Engine Based on Fuzzy Approximate Reasoning", In Proc. of the IEEEInternational Conference on Fuzzy Systems, pp. 1116–1121, 2003.
10. B. Suryavanshi, N. Shiri and S. Mudur, "A Fuzzy Hybrid Collaborative Filtering Technique for Web Personalization", In Proc. of 3rd International Workshop on Intelligent Techniques for Web Personalization (ITWP 2005), 19th International Joint Conference on ArtificialIntelligence (IJCAI 2005), pp. 1–8, 2005
11. Tyagi Shweta and KK. Bharadwaj, "Enhancing Collaborative Filtering Recommendations by Utilizing Multi-Objective Particle SwarmOptimization Embedded Association Rule Mining", *Swarm and Evolutionary Computation*, vol. 13, pp. 1–12, 2013.
12. V. Kant and KK. Bharadwaj, "Incorporating Fuzzy Trust in Collaborative Filtering Based Recommender Systems", In *Proc. of the 2ndInternational Conference of SEMCCO*, Springer Berlin Heidelberg, pp. 433–440, 2011.
13. S. Ujjin and P. J. Bentley, Learning User Preferences Using Evolution, In *Proc. of the 4th Asia-Pacific Conference on Simulated Evolutionand Learning*, Singapore, 2002.
14. Froschl, C. (2005). User Modeling and User Profiling in Adaptive Elearning Systems. Master Thesis, Graz University of Technology, Graz, Austria.Gadi, T., Daoudi.
15. Massa, P., & Avesani, P. (2004). Trust-aware collaborative filtering forrecommender systems. CoopIS/DOA/ODBASE (1), 492–508.
16. Yang, X. S. (2008). *Nature-Inspired Metaheuristic Algorithms*. Luniver Press. ISBN 1-905986-10-6.
17. X. S Yang, "Firefly Algorithms for Multimodal Optimization," *Proc. On Stochastic Algorithms:Foundations and Applications*, Vol. 5792, pp. 169-178, 2009.
18. Klir, G., & Yuan, B. (1995). Fuzzy sets and fuzzy logic, theory andapplications. Prentice-Hall.

AUTHORS PROFILE

Krishan Kant Yadav, he is a research Scholar in Department of Computer Science & Engineering in Amity University,Gwalior. He have 10 Year of teaching experience. He did MCA, M.Tech. (CS) from Rajiv Gandhi Technical University, Bhopal. He has more than 10 research publications in various reputed national journals, conferences and National seminars. His area of Research is Network Security, Data mining and data warehousing.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Prof (Dr.) Venkatadri Marriboyina, presently working as Professor and Head Dept of Computer Science and Engineering, in Amity School of Engineering and Technology. Amity University Madhya Pradesh. He has more than 25 research publications in various international journals and conferences. Editor for Smart and Innovative Trends in Next Generation Computing, Vol-827 and vol-828, Communications in Computer and Information Science Series, Springer. (Scopus Indexed). Guest Editor for "Smart and Innovative Trends for NexGen Computing and Communications", in SCIE and Scopus indexed journals. His research interests in Cloud Computing and Cyber Security, Data Analytics and Artificial Intelligence.

Dr. Sanjiv Sharma, he works as an assistant professor in Department of Computer science Engineering and Information Technology in Madhav Institute of Technology and Science, Gwalior. He have 12 year of teaching and research experience. He has more than 70 research publications in various reputed international journals and conferences. His area of research is Network security, Data Mining and Social Network Analysis.