© 2018 IJSRST | Volume 4 | Issue 2 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X
Themed Section: Science and Technology

Knowledge Based Deep Learning Collaborative Filtering (KBDLCF)

E. Sandeep Krupakar[1], Dr. A. Govardhan[2]
[1]Research Scholar, Department of Computer Science & Engineering, Rayalaseema University, Kurnool, India
[2]Professor & Rector, Department of Computer Science & Engineering, JNTUH, Hyderabad, Telangana, India

Abstract: Since from past few years huge volume of users are interacting with website for their day to day transactions are being going up. In this regards numbers of companies are encroaching towards internet to tell their products and services. This revolutions towards E-Commerce has changed, Conventional way of doing business this rapid expansion has resulted in new challenges to both companies as well as customers. Thus customers are overloaded with multiple choices for individual product which results in a confused and lost state. It has become a trivial for the webmasters to evaluate whether products and services are provided are catering the needs of customer or not. Recommender system has been playing a more vital and essential role in various information access systems to boost business and facilitate decision-making process.in this paper we proposed Knowledge Based Deep Learning Collaborative Filtering has been presented. The use of this approach joining the advantages of fuzzy logic and neural networks in order to improve CF recommender system that recommends items to the active user on the basis of fuzzy rules. The exe cution of neural networks for collaborative filtering recommendations was enlightened by an example real world dataset of movies and Reviews rated by user i.e Netflix. Results of a simulation study using various similarity measures and recommendation methods show that the KBDLCF model performs better than other conventional techniques. The experimental results reveal that forming ambiguity using fuzzy logic and neural networks mends the performance of personalized recommender systems.

Keywords: E-Commerce, fuzzy logic and neural networks, Knowledge Based Deep Learning Collaborative Filtering

I. INTRODUCTION

Now a day's several internet users using web to perform day to day transactions are increasing. In this regards numbers of companies are encroaching towards internet to tell their products and services. This revolutions towards E-Commerce has changed, Conventional way of doing business this rapid expansion has resulted in new challenges to both companies as well as customers. Thus customers are overloaded with multiple choices for individual product which results in a confused and lost state. It has become a trivial for the webmasters to evaluate whether products and services are provided are catering the needs of customer or not. Therefore it's essential to formulate novel marketing strategies such as one –to-one marketing and Customer relationship management. Effective solution for handling this issue is to provide personalized recommendation to individual users such as providing the customer with the type of product recommendation list he or she is interested in. A promising solution to overcome this issue is recommendation system, which provides and guides the customer with type of product he or she is interested in Buying/purchasing. So far a wide variety of recommendation system have been proposed and implemented.

Traditional recommendation system can be classified into THREE categories i.e 1. Content based 2. Collaborative filtering system. 3. Hybrid Recommender System

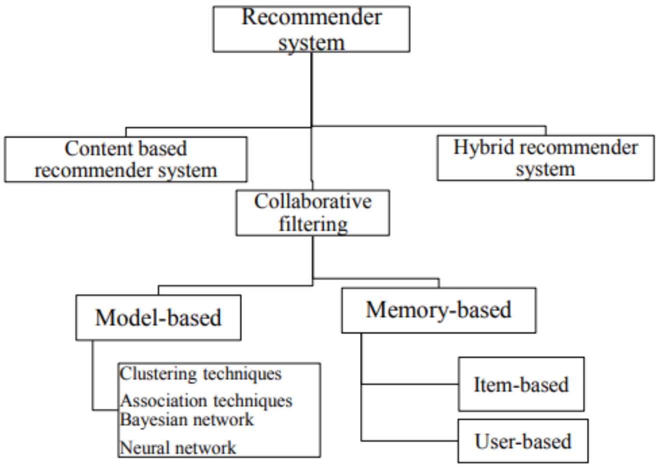

Figure 1. Anatomy of recommendation techniques

Content Based system

Takes into consideration product attributes to generate recommendations, Content based filtering methods are based on a description of the item and a profile of the user's preference. In a content-based recommender system, keywords are used to describe the items; beside, a user profile is built to indicate the type of item this user likes. In other words, these algorithms try to recommend items that are similar to those that a user liked in the past (or is examining in the present). In particular, various candidate items are compared with items previously rated by the user and the best-matching items are recommended. This approach has its roots in information retrieval and information filtering research.

Collaborative filtering system make use customer – product interaction and ignore other attributes to generate recommendations despite significant progress and widespread usage of the recommendations system suffers from following limitations.

1. Scalability : Consumer cum Products increases in numbers ,it slows down neighbor selection per second
2. Binary transaction (Clickstream) data: Most of the recommendation system make use of binary transaction data i.e. whether a specific item is purchased or not. However many times they are unable to exploit intrinsic characteristics of these data , that can be used to provide better recommendation
3. Retargeting : Providing the customer with same product that is already purchased

Hybrid Recommender System

Hybrid recommendation system is a method which combines content based with collaborative filtering techniques to gain better prediction or recommendation performance better prediction or recommendation performance. Hybrid approaches can be implemented in various ways: Implement collaborative and content-based methods individually and aggregate their predictions, Integrate some content-based characteristics into a collaborative approach, Comprise some collaborative characteristics into a content based approach, and Construct a general consolidative model that integrate both content based and collaborative characteristics. Cold start and the sparsity are common problems in recommender systems which are resolved by using hybrid methods.

Our proposed framework uses a Knowledge Based Deep Learning Collaborative Filtering (KBDLCF) that makes usage of data and integrates it with user, sales data to provide better recommendations to all the visitors of E-Commerce sites.

II. RELATED WORK

In this section the related works of fuzzy logic and artificial neural network based recommender systems have been discussed.Recommender systems are one of the information filtering and information retrieval methods used to provide suggestions to the items of active user on web. The first recommender system

G. Castellano et al.[1] presented a neuro-fuzzy strategy that combines soft computing techniques to develop web recommendation system that dynamically suggest interesting URLs for the current user. User access logs are analysed to identify user sessions as a preliminary step. The groups of user's common browser behaviour are discovered by applying a fuzzy clustering algorithm to the user sessions. Finally, a knowledge extraction process is carried out to derive associations between user profiles and relevant web pages to be suggested to users. In order to derive knowledge from session data and represent it in the comprehensible form of fuzzy rules a hybrid approach based on the combination of the fuzzy reasoning and connection paradigm has been proposed by this author.

G. Castellano et al. [2] proposed the use of a neuro-fuzzy strategy to develop a Web personalization framework for the dynamic suggestion of URLs retained interesting for the currently connected users. In particular, a hybrid strategy exploiting the combination of the fuzzy logic with the neural paradigm is proposed in order to discover useful knowledge from session data identified from the analysis of log files and represent it in a set of fuzzy rules expressed in an interpretable form.

De Campos et al. [3] proposed a model by combining Bayesian network for governing the relationships between the users and fuzzy set theory for presenting the vagueness in the description of user's ratings.

Soheila Ashkezari-T et al. [4] contribution using genre based information in a hybrid fuzzy-Bayesian network collaborative RS is suggested. The interest to the different genres is computed based on a hybrid user model. They calculated similarity of likeminded users according to the fuzzy distance and also Pearson correlation coefficient is involved in a Bayesian network. The author call the RS uses as Bayesian network RS (BNRS), use Pearson for similarity computations named it PBNRS and using fuzzy distance for similarity computation called it FBNRS.

Maral Kolahkaj et al. [5] presented the fuzzy-neural network for finding users movement pattern using fuzzy clustering technique and web usage mining. The author has presented a method to deal with limitations of conventional methods. Their system finds useful movement pattern of users and also presents user's future requests by using fuzz-neural network.

Collaborative Filtering

Goldberg et al. [6] proved that Collaborative Filtering CF is very successful in both research and practice. Most of the existing systems use CF as its base algorithm for providing recommendation. In 1994 the GroupLens system [67] has implemented a CF algorithm. The author noted that the recommendations are generated based on the users' similarities for the formation of the nearest neighbours, it is nowadays known as userbased collaborative filtering algorithm.

John S. Breese et al. [8] classified CF algorithms into two major types: memory based and model based. Memory based algorithms predictions are based on the similarity of the previously rated items by the users. Model based algorithms predicts based on the collection of ratings by learning the designed model. This model is used to make predictions of recommendations of a new active user.

Deepa Anand [7] proposed methods to enrich the set of user connections obtained using measures such as Pearson Correlation Coefficient(PCC) and Cosine Similarity(COS).Their experiments comparing various global similarity schemes show that different similarity schemes perform best under different circumstances.

III. System Study

3.1 Proposed Framework

Our Proposed Research framework illustrated in detail in figure 2

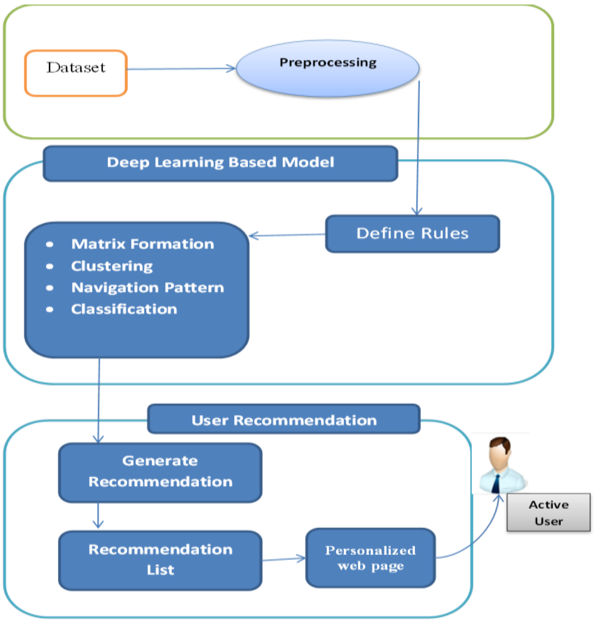

Figure 2. Proposed Knowledge Based Deep Learning Collaborative Filtering (KBDLCF)

3.2 Methodology

Knowledge Based Deep Learning Collaborative Filtering (KBDLCF) Model approach is separated into three sub phases. The First phase is pre-processing. In this phase data is pre-processed using various pre-processing techniques such as feature selection, normalization and data reduction. In the Second phase deep learning modeling is constructed. Active user's knowledge about the data such as profile and preferences are defined and applied in terms of fuzzy rules (crisp rules) on effectiveness matrix. Personalized Artificial Bee Colony algorithm (PABCA) is deploying to determine the web user pattern or to cluster similar users. Once the user navigation patterns are recognized using clustering the active user's best matching cluster is predicted using BPN classification to generate recommendations. Final phase is obtaining business intelligence i.e. recommendation generation process for the active user is performed online process.

3.2.1 Data Pre-processing

The essential pre-processing steps discussed in the earlier chapters such as feature selection i.e. Minimum Redundancy Maximum Relevance Feature Selection (mRMR) technique has been applied to select relevant features and samples. The feature selection algorithm is designed with different approaches broadly fall into three categories: Filter, wrapper and embedded models. The Filter model relies on the general characteristics of data and evaluates features without involving any learning algorithm. The wrapper model requires a predetermined learning algorithm and uses its performance as evaluation criterion to select features. Algorithms with embedded model Incorporate variable selection as a part of the training process, and feature relevance is obtained analytically from the objective of the learning model.

Missing values are replaced with values using neural networks. The back propagation feed forward neural network is used to replace the missing values. The user-item rating matrix is normalized between 0 and 1. The Principal Component Analysis technique is used to reduce the dataset from high to low.Good quality input data needs to be saved for better analysis. In this phase in consisted and duplicate data eliminated using following steps in figure 7.3 Field separation stage focuses on separates one attributes from another by making use of separator character such as space. In data cleaning stage we perform filter out outlier data, we check for URL suffixes. Log entries having file name suffixes such as gif,jpeg,tif,jpg are discarded , All records having failed http status code removed i.e status code greater than 200 amd less than 299 are eliminated in user differentiation phase we assign unique user ID to each IP address and registered users to separate one customer from other. Finally we construct session in session identification phase we group together session belonging to unique user.

Bringing all of the ideas together into one algorithm there is mRMR feature selection. The idea behind this algorithm is that you want to minimize the redundancy of features while maximizing the relevancy. To do this, we have Equations for calculating relevancy and Redundancy.

$$Relevancy = \frac{\sum_{i=1}^{n} c_i x_i}{\sum_{i=1}^{n} x_i} \qquad Redundancy = \frac{\sum_{i,j=1}^{n} a_{ij} x}{\left(\sum_{i=1}^{n} x_i\right)}$$

Session Provide us with complete set of activities done by the user in specific time period. Finally in data formatting stage we place the data in tabular form.

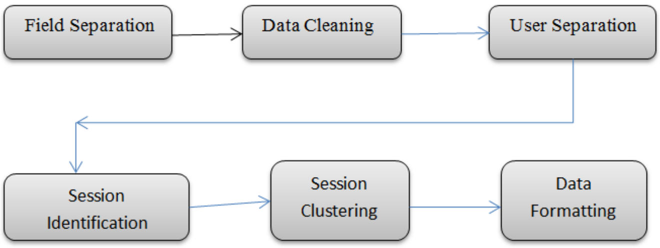

Figure 3 Steps for Pre-processing

Field separation: It focuses on separating individual fields by making use of separator character such as space.

Data cleaning: Data cleaning is a process of filtering out irrelevant and outliers' data [9]. It eliminates all irrelevant items by checking the suffix of the URL name. Therefore, all log entries with filename suffixes such as gif, jpeg, GIF, JPEG, and JPG are removed. All records of failed HTTP status code i.e. Status code less than 200 and greater than 299 are eliminated. For the present study, we consider only the GET and POST methods. Data cleaning reduces the total number of records and also log file size.

User differentiation: It is important to distinguish between different users for analysing different user access behaviour patterns. A different user ID will be assigned to different IP address. In case of same IP address referrer information and browser details will be used to distinguish among different web users.

Session identification: A session is defined as an ordered sequence of web pages visited by a user. A new session is constructed based on new IP address. Each new IP addresses implies (correspond) to a unique user. A maximum session time limit is considered to be 30 minutes.

Data formatting: Finally data will be formatted to appropriate tabular format for further analysis.

3.2.2 Deep learning Model Construction

In this second level, Knowledge Based Deep Learning Collaborative Filtering (KBDLCF) is made using Fuzzy Logic and Artificial Neural Networks. This stage has three steps, knowledge base step is used to filter of users and/or items using procedures, clustering of user-item rating matrix using **Minimum Redundancy Maximum Relevance Feature Selection** (mRMR) and predicting BMC of active user using Back Propagation Neural Network (BPN) classification.

Knowledge Base

In the first step, the active user's knowledge information such as profile and preferences are presented in the form of if-then rules and applied. Users or items in the dataset are filtered based on knowledge defined in terms of these crisp/fuzzy rules to identify the user's navigation pattern.

When there is ambiguity in user's preferences and profiles, fuzzy rules are defined on membership function. The profile and preferences are also considered as fuzzy variables. When the active user is new to recommender system (Cold-start problem) the proposed model generates recommendation based on knowledge filtering of users based on demographic information and preferences of active user. The users or items satisfying the crisp/fuzzy rules are filtered to reduce the dataset for mining business intelligence. Clustering of Utility Matrix using Minimum Redundancy Maximum Relevance Feature Selection (mRMR).

In this second stage, the user-item rating matrix is clustered to identify the group of similar users. Since the dataset contains multidimensional data with more number of features and user, neural network based mRMR clustering is used to group the users to find the users navigation patterns.

Identifying BAC using BPN

Finally, active user's Best Alike Cluster (BAC) is identified using back propagation neural network classification technique. Resilient back-propagation (RPROP) is a neural network training algorithm is used to identify matching cluster of active users. The steps are involved in KBDLCF classification.

3.2.3 KBDLCF Algorithm

Knowledge Based Deep Learning Collaborative Filtering (KBDLCF), algorithm is presented in this section. This algorithm demonstration the pseudo code of the KBDLCF 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 $\{I_1, I_1, I_1, I_1 ... I_n\}$ of *Top-N* items.
II Phase I: Pre-processing
Select relevant features using Feature Selection.
Replace missing values using NN.(if any)
Perform Normalization.
Do Dimension Reduction using PCA.
II Phase II: Deep learning Model Construction
KBDLCF Based crisp rules (if any) for user preferences and/or profiles. //Formulate the fuzzy rules (FR).

- Outline 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 $p_s(a_i)$ where j = 1 to m.
// Fuzzification of samples/user profile features.
Compute $\mu_s(p_j)$ where j = 1 to m .
Define KBDLCF 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 mRMR
//Predicting Best Alike Cluster (BAC) of active user.
For each Active user in TD do
Find the Best Alike Cluster (BAC) using resilient back-propagation (RPROP).
End
TMAE <— Evaluate Matching Cluster
End If
//Phase III: Recommendation
For each Active user in TD **do**
Identify items from Best Alike Cluster (BAC) of users. Calculate the frequency count and rank the items.
Select and recommend *top-N* items.
End

IV. Experimental Results

This section presents comprehensive investigational outcomes of KBDLCF model. First the quality of clustering is analysed and validated using Silhouette Index. Second, the results of prediction evaluation using MAE are presented. Third, recommendation quality evaluation with different parameters is presented. KBDLCF model has been experimentally simulated and evaluated with two real-world benchmark datasets namely Netflix and MovieLens 1M data set with different active users profile and preferences.

4.1Clustering Quality Analysis

The outcome of different numbers of clusters k=2 to N/2 on clustering prominence is examined. The finest numbers of are usually verified between 2 and N/2. The Table 1. Shows Mean Silhouette Index value for k=2 to N/2 number of clusters deliberate using **mRMR** and Fuzzy C Means (FCM) Clustering algorithm.

Table 1. Mean Silhouette Index of SOM and FCM clustering

Dataset	mRMR Clustering	FCM Clustering
MovieLens	0.40	0.30
Netflix	0.45	0.39

The investigated result indications that there is a substantial variance in mRMR Clustering and FCM is clustering. The mean Silhouette Index of Deep learning based mRMR Clustering is 10% better quality panel based FCM clustering using MovieLens dataset. The Mean Silhouette Index of deep learning based mRMR clustering is 8% better Quality panel based FCM clustering using Netflix dataset. Hence, the mRMR clustering is capable of handling large volume of dataset with better accuracy.

4.2 Recommendation Quality Analysis on MovieLens Dataset

The KBDLCF model is tested by compelling range of recommendations using MovieLens dataset from 10 to 50 in the step of 5 and 5 to 30 in the step of 5 using Netfix dataset for top-N=10.

Table 2. Mean Recall, Precision and Fl Measures of KBDLCF model using MovieLens dataset.

	ML-dataset-1			ML-dataset-2			ML-dataset-3		
Top-N	Recall	Precisio	Fl	Recall	Precisio	Fl	Recall	Precision	F
10	0.703	0.8600	0.712	0.742	0.7812	0.790	0.895	0.6022	0.756
15	0.821	0.7300	0.613	0.907	0.6265	0.643	0.948	0.5034	0.523
20	0.878	0.6415	0.645	0.945	0.5324	0.676	0.950	0.4121	0.516
25	0.932	0.5528	0.623	0.963	0.4546	0.556	0.967	0.3535	0.498
30	0.961	0.5277	0.576	0.979	0.4033	0.543	0.964	0.3038	0.432
35	0.976	0.4520	0.523	0.981	0.3624	0.489	0.979	0.2712	0.398
40	0.983	0.4221	0.530	0.981	0.3323	0.465	0.981	0.2215	0.334
45	0.982	0.3845	0.488	0.981	0.2834	0.411	0.981	0.1901	0.298
50	0.984	0.3332	0.471	0.981	0.2612	0.385	0.981	0.1812	0.277

From the table, it is pragmatic that the maximum mean Fl measure of KBDLCF model is 77 % and the resultant recall and precision are 75% and 79% correspondingly.

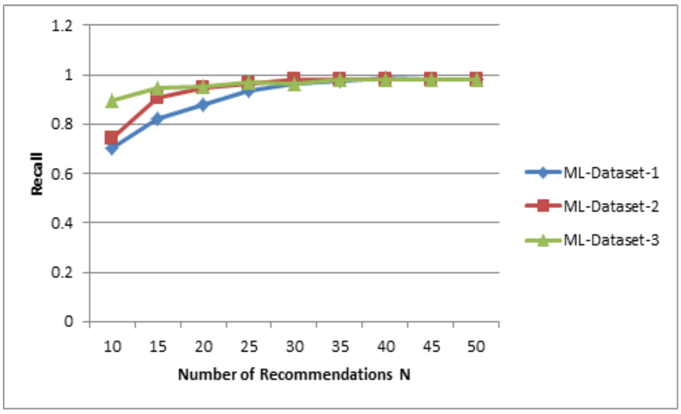

Figure 4.1 Mean Recall of top-N recommendations using Movielens dataset

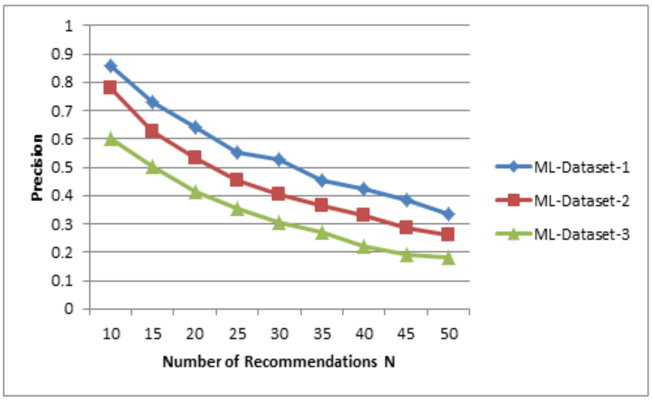

Figure 4.2 Mean Precision of top-N recommendations using Movielens dataset

Comparison of Precision

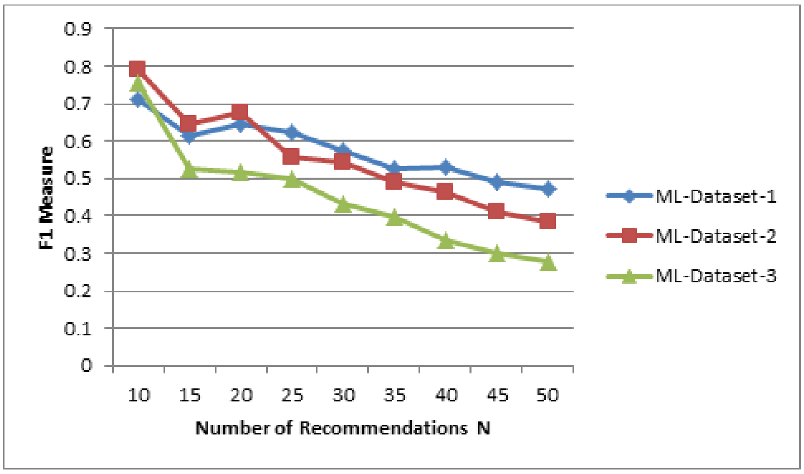

Figure 4.3 Mean F1Measure of top-N recommendations using Movielens dataset

4.3 Recommendation Quality Evaluation Netflix dataset

The experiment is repeated with range of N values from 5 to 30 in the step 5 and calculated the mean Recall, Precision and Fl-Measure using Jester dataset. The

Table 3. Shows the mean Recall, Precision and Fl Measures of KBDLCF model using Netflix dataset.

TOP-N	Netflix dataset - 1			Netflix dataset - 2			Netflix dataset - 3		
	Recall	Precision	Fl	Recall	Precision	Fl	Recall	Precision	Fl
5	0.3651	0.9765	0.5965	0.3748	0.9850	0.5326	0.3908	0.9859	0.5600
10	0.6814	0.9758	0.8124	0.7500	0.9788	0.8500	0.7750	0.9789	0.8798
15	0.9916	0.9701	0.9658	0.9723	0.9866	0.9770	0.9598	0.8781	0.9377
20	0.9865	0.7598	0.8546	0.9890	0.7291	0.8450	0.9594	0.6599	0.8245
25	0.9856	0.5987	0.7598	0.9859	0.5936	0.7440	0.9606	0.5247	0.6988
30	0.9645	0.5577	0.7785	0.9775	0.5843	0.7456	0.9604	0.4801	0.6388

From the table, it is observed that the maximum mean Fl measure of KBDLCF model is 87% and the corresponding mean recall and precision are 77% and 97% respectively for top-N=10.

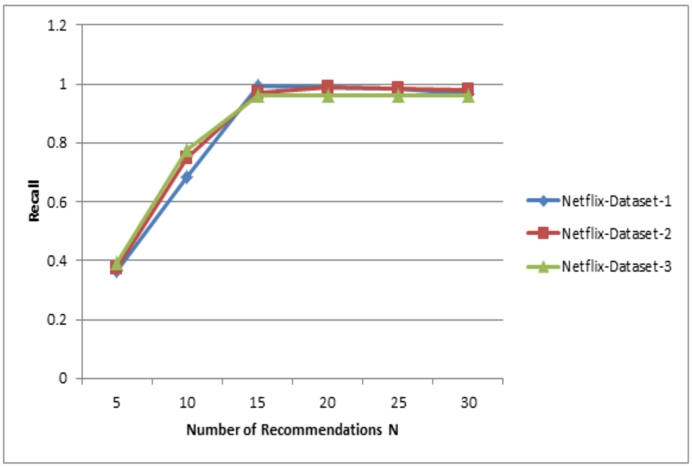

Figure 4.4 *Mean Recall of top-N recommendations using Netflix dataset*

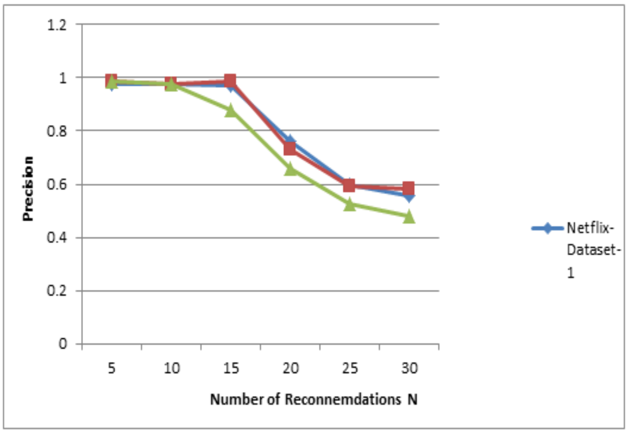

Figure 4.5. *Mean Precision of top-N recommendations using Netflix dataset*

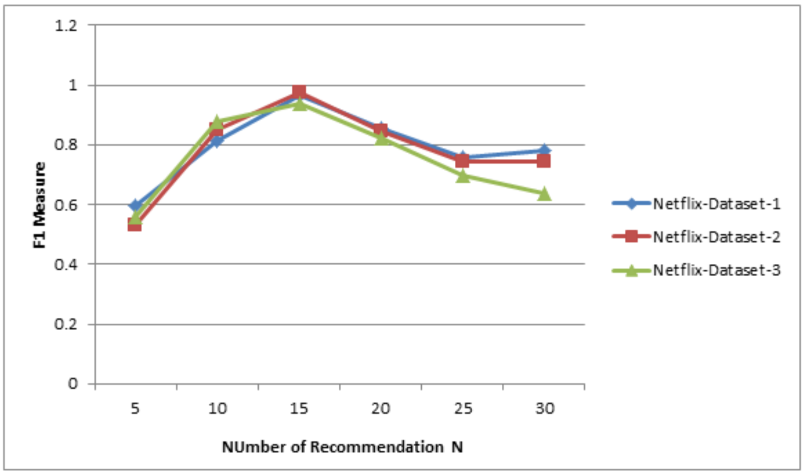

Figure 4.6 Mean F I-Measure of top-N recommendations using Netflix dataset

4.4 Comparison with Conventional Approaches

This division enlightens the comparison of KBDLCF model with conventional k-Nearest Neighbor (k-NNBM) Approach using MovieLens and Ant Based Recommender System (ARS), k-nearest neighbor based Mean Squared Distance *Netflix* datasets with commonly compared *top-N* value as 10.

The results listed in the table shows that, the accuracy of KBDLCF model is better than k-NNBM, k-NNBM (w),GAC and k-NNBM (PCC) algorithms.

Comparison using MovieLens dataset

The Table 4. Shows the comparison of Mean Fl measure between KBDLCF model and conventional models using MovieLens dataset for *top-N=10.*

Table 4 Mean Fl of KBDLCF with conventional methods using MovieLens.

Algorithm	k-NNBM	k-NNBM(w)	GAC	k-NNBM(PCC)	KBDLCF
Fl Measure	0.49	0.55	0.63	0.69	0.80

When compared with k-NNBM, k-NNBM (w), GAC and k-NNBM (PCC) algorithms minimum 11% increase in Fl-measure is obtained using KBDLCF model. Therefore the performance of the KBDLCF is more significant than the conventional methods.

V. CONCLUSION

In this paper Knowledge Based Deep Learning Collaborative Filtering has been presented. The use of this approach joining the advantages of fuzzy logic and neural networks in order to improve CF recommender system that recommends items to the active user on the basis of fuzzy rules. The execution of neural networks for collaborative filtering recommendations was enlightened by an example real world dataset of movies and jokes rated by users. Results of a simulation study using various similarity measures and recommendation methods show that the KBDLCF model performs better than other conventional techniques. The experimental results reveal that forming ambiguity using fuzzy logic and neural networks mends the performance of personalized recommender systems.

VI. REFERENCES

[1]. G Castellano, A.M. Fanelli, P. Plantamura, M.A. Torsello, A Neuro-Fuzzy Strategy for Web Personalization, Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence 2008.

[2]. G Castellano, A.M. Fanelli, M.A. Torsello, A neuro-fuzzy collaborative filtering Science,1992-6669 (Print) 1992-6677(Online) Global Information Publisher,Vol.1,No.1,28-39,2007.

[3]. De Campos LM, Fernandez-Luna JM, Huete JF. A collaborative recommender system based on probabilistic inference from fuzzy observations. Fuzzy Sets Systems 159(8):1554-1576, 2008.

[4]. Soheila Ashkezari-T, Mohammad-R, Akbarzadeh-T, Fuzzy-Bayesian Network Approach to Genre-based Recommender Systems, 978-1-4244-8126-2/10 ©IEEE 2010.

[5]. Maral Kolahkaj,Ali Harounabadi,Mehdi Sadegh Zade, A Recommender System for Web Mining using Neural Network and Fuzzy Algorithm, International Journal of Computer Applications (0975 – 8887) Volume 78 – No.8, September 2013.

[6]. D Goldberg, D. Nichols, B. M. Old, and D. Terry, Using collaborative filtering to weave an information tapestry, Communications of the ACM, vol. 35, no. 12, pp. 61-70, 1992.

[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]. John S. Breese, David Heckerman, and Carl Kadie, Empirical Analysis of Predictive Algorithms for Collaborative Filtering, TechReport, Microsoft Research, Morgan Kaufmann Publishers, 1998.

[9]. Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P. & Riedl, J. GroupLens: an open architecture for collaborative filtering on netnews in Proceedings of the conference computer supported collaborative work, pp. 175-18