

# Design and Analysis of Recommender System for Business Data

Prabhat Kumar\*, Dr. KiranJyoti\*\*

\*(Computer Science and Engineering, Guru Nanak Dev Engineering College, Ludhiana)

\*\* (Information Technology, Guru Nanak Dev Engineering, Ludhiana)

\*\*\*\*\*

## Abstract:

Recommender frameworks have changed the way individuals discover items, data, and even other individuals. They study examples to recognize what somebody will incline toward from among an accumulation of items he has never experienced. The innovation behind recommender frameworks has advanced in the course of recent years into a rich gathering of devices that empower the expert or scientist to create successful recommenders.

Collaborative filtering, a standout amongst the most generally utilized approach as a part of recommender framework, predicts a client's appraising towards an item by accumulating appraisals given by clients having comparative inclination to that client. In existing methodologies, client comparability is regularly processed all in all arrangement of items. In any case, on the grounds that the quantity of item is frequently huge, as is the assorted qualities among items, clients who have comparative inclination in one class of items may have very surprising judgment on items of another kind. Keeping in mind the end goal to manage this issue, we propose a strategy for grouping items, so that inside a cluster, closeness between clients does not change altogether. After that, when anticipating rating of a client towards a items, we just total appraisals of clients who have high likeness degree with that client inside the cluster to which that item has a place. Investigations assessing our methodology are completed on the genuine dataset taken from motion pictures suggestion arrangement of MovieLens site. Preparatory results recommend that our methodology can enhance expectation precision contrasted with existing methodologies.

**Keywords —Recommender System, Clustering.**

\*\*\*\*\*

## I. INTRODUCTION

With the development in item options, picking a suitable item inside of an imperative period nowadays is exceptionally troublesome, to beat this downside, recommender frameworks are being utilized. [5]A recommender framework in a layman term is a subclass of data separating framework that can help the client to discover an item in the pursuit space as indicated by their taste. A normal recommender framework assemble its database of client taste by at first gathering huge measure of client tastes and afterward tries to relate the enlisted taste with another arrangement of clients. Recommender frameworks these days are turning out to be to a great extent well known and have wide zone of utilization. Recommender framework in it part is only a rating framework that can foresee client tastes taking into account their history of

procurement, a recommender framework contrasts a client taste and different clients in the database, with some calculation, it gives the client a few suggestions, with some level of certainty. There are three types of recommender system:

1. Collaborative based recommender system
2. Content based recommender system
3. Hybrid recommender system

### A. Collaborative based recommender system

[6]In a collaborative framework, two or more client rate an arrangement of items, if some client demonstrates some regular taste, then it can be anticipated that another item enjoyed by a client will likewise be preferred by another client of same taste. Collective suggestion is in view of the connection, i.e. in the event that two client has comparable taste in the past then they will have same taste later on, for a synergistic recommender

framework to work, immense measure of client information, for example, rating and input is taken, then their taste are processed next time when the client visits the site, the item taking into account their taste are just introduced. Database for a synergistic recommender framework can be fabricated either explicitly or implicitly. In an explicit database development, client has gotten some information about arrangement of items in type of appraisals or content and in an implied development client exercises are checked and subsequent to checking for a few times their advantage are enlisted, for instance if a man P introspects an item I, then it can be concluded that he or she is occupied with item I, and next time recommender framework can securely prescribe item I. There are two types of collaborative recommender system:

1. User-User Based - [2]In a client based shared recommender framework, conclusions from different clients are enlisted and in light of their feeling and determination, their taste database is developed. Subsequent to building taste database, an item is prescribed to another client with comparative taste, for instance if two clients has comparable taste in motion pictures, then it is likely if a film M, being preferred by the first client will be enjoyed by the second and the other way around..

2. Item-Item Based - [9]It is in light of the guideline, if a client purchases an item I1, then it is likely he will purchase item I2. Case in point if a man purchases bread, then he is liable to purchase margarine too. Item-Item database is manufactured by watching different client, and their acquired made, and finding concealed examples in their buy.

#### **B. Content Based Recommender System**

[7]Not at all like community is oriented recommender framework, in substance based recommender framework, it critical to know the substance of the item being suggested. In content based recommender framework, client inclinations are recorded and substance of items are likewise being compressed, and in the wake of coordinating client inclinations with item content, the items are being prescribed to them. Case in point - if a client U1 purchases red shirt and its taste is like client U2, then red shirt will be prescribed to the client U2, if

and if client U2 has demonstrated any enthusiasm for red shading in their past buy. Essentially substance based recommender framework is all that much like Item-Item based collective framework, with the consideration that substance of the item and client is contemplated before foreseeing. Community based recommender framework are the most well-known type of recommender framework, principally in light of the fact that it is anything but difficult to create, takes almost no time to anticipate an outcome or item, and it devours less asset than whatever other recommender framework systems, however real disadvantage of synergistic based recommender framework is that it can't confirm or concentrate substance inside sound or feature document, besides the shared channel are just relevant presently, and can't even parse content inside a report, to defeat this defect, substance based recommender frameworks are utilized. The substance based recommender framework can parse any sound, feature, content or double document and can separate data inside those documents, however significant downside of substance based recommender framework is that it doesn't have any client's communication, all the choice made by substance based recommender framework are absolutely be the substance. Subsequently it is regularly called item-item cooperation.

#### **C. Hybrid Recommender System**

[3]Hybrid recommender framework is a hybrid between User-User based synergistic framework and content based recommender framework, where client taste are enrolled and items like client taste are prescribed yet contents of client inclinations and items data is considered. Both collaborative and content based recommender framework has their own particular defects, to conquer those blemishes, hybrid recommender frameworks are utilized, hybrid recommender framework has the property that it perform both client item coordinated effort and item-item cooperation, it is a blend of community oriented based recommender framework and content based recommender framework. Hybrid Recommender framework does take client's cooperation as appraisals furthermore has the ability to recover contents from non-literary document.

## II. LITERATURE REVIEW

With the expansion in electronic and consumer goods, the number of choice per head has grown exponentially and they also discussed about types of collaborative filtering, according to them there are two types of collaborative filtering systems and they are memory based and model based. The memory based collaborative filtering systems employ algorithms such as user-user based collaborative filtering algorithm whereas the model based collaborative filtering employs item-item based and singular value decomposition based filtering techniques. The quality of today's recommender system is much better than pure content based and demographics based recommender system, but still has sparsity and cold start problems, resulting in poor recommendations, to overcome these problems they proposed a boosted RDF scheme which is a hybrid recommendation system that can handle cold start and sparsity problems. For their experiment they took twenty percent of ratings from the user and used the rest as a test set, then applied their algorithm together with Pearson correlation as a similarity measure, and Naïve classification as a classifier to generate recommendation. Their algorithm took  $O(N)^2$  time and the strength of their recommendation was 0.725.

Liangxign et al (2010)[8] proposed a system for hybrid product recommender system for apparel retailing customers, in their system, every customer was given a RFID circuit embedded card, and on punching those card in magnetic swipe readers, would identify the customer and their history, and based upon their history their hybrid recommender system would recommend them with the list of apparel they might be interested in. Their system consisted of three different databases, one for on-sale product, next for membership customer historical purchase and last for membership customer information. These data were then processed with similarity algorithm to compute item-item and user-user collaboration, then initial recommendation is obtained for the customer then top N recommendation is presented to the customer.

The data set for their research were obtained from UNIQLO, SDEER, GOELIA, YISHION, MISSFOX and NIBBUNS. From their result they concluded that most of their customers were between 20 to 30 years old.

Mittal et al (2010)[6] proposed a recommender system based on recommender system framework using K-means clustering. In their work, at first user has to enter genre of the movies and then their opinion are partitioned into dataset based on their genre selection, by partitioning the resultant size of their database reduces considerably, further, K-means clustering technique is applied on those data set to figure out clusters, and then some movies from these clusters are presented to the user and user has to select movies from these clusters, their recommender system is item-item based and modelled using slope one algorithm. To model their system, they collected Movie Lens dataset with seventy thousand ratings, five hundred users and three thousand eight hundred and eighty three movies titles of different types. The system also takes into account the demographics of the user while providing the results. They also concluded that using space partitioning and K-means clustering algorithm substantially improves both time and prediction accuracy.

Devi et al (2010)[4] according to them, Community oriented Recommender system helps the web customers to appreciate the right protest in the midst of electronic securing. The collaborative recommender structure recognizes the comparable purchasers in perspective of the getting or rating behavior to the dynamic buyer and after that implies the thing in light of the relative buyers. Community oriented recommender structure is by and large used as a component of lion's offer of the present online recommender system, for occurrence, orkut, google, amazon, walmart and numerous others. Rather than it unmistakable quality, is perseveres for the reason that of sparsity, chilly start and versatility recommender structure. Vast investigation is going ahead to beat these issues. In their paper they mentioned that, Probabilistic neural system (PNN) is used to focus the trust between buyers in perspective of positioning structure. Using the figured accept, sparse score group is smoothed, by utilizing predicting the rating

estimations of the nonrated things in the score cross section. Using this smoothened rating system, the trust is figured for online element customers. The discovered trust is used to endorse thing. Appraisals are coordinated making utilization of dataset, for outline, movielens. In light of the execution estimations, it is affirmed that the proposed framework performs higher than the benchmark and some present structures.

Adomavicius et al (2010)[1] displayed a diagram of the field of recommender frameworks and portray the present era of suggestion techniques that are generally ordered into the accompanying three principle classifications: substance based, community, and cross breed proposal methodologies. The paper additionally depicts different limits of current suggestion systems and examines conceivable expansions that can enhance proposal capacities and make recommender frameworks appropriate to a much more extensive scope of utilizations. These expansions incorporate, among others, change of comprehension of clients and things, consolidation of the context oriented data into the proposal procedure, support for multi-criteria appraisals, and procurement of more adaptable and less nosy sorts of suggestions.

### **III. METHODOLOGY**

**Collect Data** - Data Gathering is the procedure of get-together and measuring data on variables of interest, in a set up methodical design that empowers one to answer expressed examination inquiries, test speculations, and assess results. The data collection segment of examination is normal to all fields of study including physical and sociologies, humanities, business, and so forth. While systems shift by order, the accentuation on guaranteeing exact and fair gathering continues as before. The objective for all information gathering is to catch quality confirmation that then means rich information investigation and permits the building of a persuading and solid response to inquiries that have been postured.

The principal thing that will be done is assembling or making dataset. Furthermore, the dataset should be sufficiently expansive so that the

assessment between two recommenders can be accomplished.

For this reason we have picked two datasets. Among the two, the first one contains 100,000 evaluations from 1000 clients on 1700 motion pictures and another contains 1 million evaluations from 6000 clients on 4000 films from grouplens.com.

**Cleaning the data-** Data cleaning is the procedure of recognizing and remedying degenerated or wrong records from a record. Utilized chiefly as a part of databases, the term alludes to distinguishing fragmented, off base, wrong, unessential, and so forth parts of the data and after that supplanting, adjusting, or erasing this coarse data.

Subsequent to purifying, a data set will be predictable with other comparable data sets in the framework. The irregularities identified or evacuated may have been initially brought on by client passage lapses, by defilement in transmission or capacity, or by diverse data word reference meanings of comparative substances in distinctive stores.

Data purging varies from data acceptance in that approval perpetually implies data is rejected from the framework at passage and is performed at section time, as opposed to on clusters of data.

In this research each dataset consists of various files. The first one contains 5 .dat files along with various other test case files. The other one contains 3 .dat files.

**Analyzing Data** - Dissecting data includes looking at it in ways that uncover the connections, examples, patterns, and so on that can be found inside of it. That may mean subjecting it to factual operations that can let you know not just what sorts of connections appear to exist among variables, additionally to what level you can believe the answers you're getting. It may mean contrasting your data with that from different gatherings (a control or correlation bunch, statewide figures, and so on.), to help make a few determinations from the data. The point, regarding your assessment, is to get a precise evaluation keeping in mind the end goal to better comprehend your work and its impacts on those you're concerned with, or so as to better comprehend the general circumstance.

Examining the data is a vital piece of any recommender framework. One can't manufacture a genuine Recommender System if does not know the data of the data. Since there are no standard routines for cleaning the data henceforth it is not indicated.

The most ideal approach to investigating the data is by making a graphical photo of the data. This can be accomplished by plotting them in N-Dimensional space.

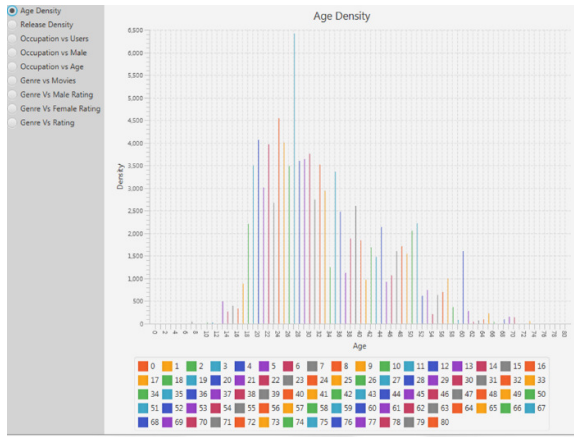


Figure 1: Density vs Age

**Selecting a Distance Function** - In this context Distance function is a function that depicts how similar or dissimilar two objects under consideration are. There are various different Distance functions measures but for this research, Weighted Manhattan distance has been chosen and is given as:

$$S(x, y) = \sum_{i=1}^d F_i - A * \lambda(g(x)), A \in \mathbb{N}$$

$$F_i = |x_i - y_i| * W(n)$$

$$W(x) = \frac{1}{w_i}, w_i \in \mathbb{R}$$

$$g(x) = \sum_{i=1}^d \frac{m(x)}{n(x)}$$

$$m(x) = |x_i - y_i|$$

$$n(x) = |x_i + y_i|$$

**Clustering** - The basic idea is to make sure if there are two nearby points then they end up in the same cluster, the agglomerative clustering is bottom up, each cluster contains one data point,  $c_i = \{P_i\}$ .

Algorithm:

- 1) For all data points  $p_i$ , put that points into an individual cluster  $c_i$ .
- 2) Find two cluster  $c_i$  and  $c_j$  such that their distance is less than every other cluster points, i.e.  $\forall_{i,j}, \min(\text{Dist}\{c_i, c_j\})$ . The distance function used is Weighted Manhattan Distance.
- 3) Merge the cluster  $c_i$  and  $c_j$  into a new cluster  $c_k$ .
- 4) Remove  $c_i$  and  $c_j$  from the list, and add  $c_k$ .
- 5) Stopping Condition: If number of cluster  $n(c) \neq$  required cluster then go to step 2 else stop

**Selection of Neighbors** - Selection of Neighbors is necessary because in real life scene the size of table is enormous and computing all the distance function between every pair of object is not feasible

- 1) Top N neighbors – Selecting Top N-Rows
- 2) Pick nearest neighbors from the same group, User group for user-user based and Item group for item-item based as formed by Hierarchical cluster
- 3) Randomly – Picking the objects from the matrix randomly.

In our case we selected N neighbours because it gave good performance with acceptable accuracy

#### IV. RESULT

To test the accuracy of the system, three important measures are taken, namely:

Precision - Percent of Selected Item that are correct, and is given as:

$$P = \frac{\text{Number of correctly identified selection}}{\text{All Selections}}$$

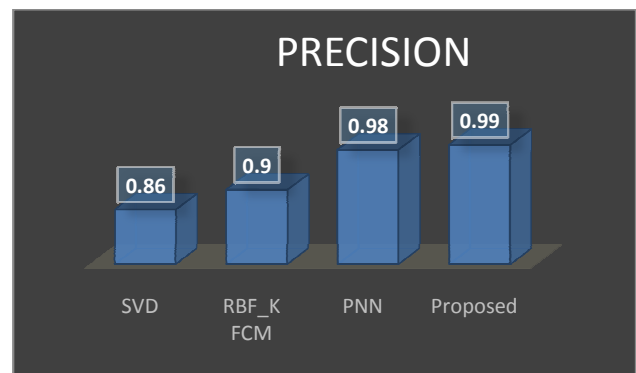


Figure 2: Precision[4]

The precision of the proposed system is 99%, to calculate precision 500 experiments were performed.

Recall - Percent of correct items that are selected

$$R = \frac{\text{Number of selection that are correct}}{\text{Total Correct Data Points}}$$

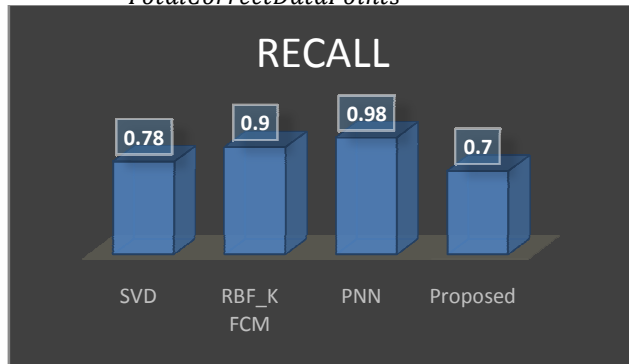


Figure 3: Recall[4]

The recall of the proposed system is 70 percent, the recall is somewhat lower than other systems, but recall is not a necessary factor in recommender system having large rating datasets. And there is a tradeoff between recall and precision; the proposed system chose precision over recall to yield better results.

F-measure - It is a harmonic mean between Recall and precisions and is given as:

$$F - \text{Measure} = \frac{1}{(\beta)^{\frac{1}{p}} + (1-\beta)^{\frac{1}{R}}}, \text{ where } \beta \text{ is a control}$$

parameter

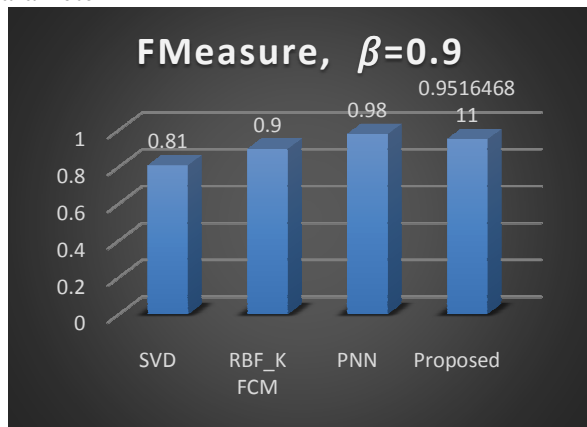


Figure 4: F-Measure[4]

The F-measure of the proposed system is approximate 95 percent, the lower F-measure score of the proposed system is due to lower recall score, but since precision score of the proposed recommender system is very high, the

recommendations generated by the recommender system is very accurate.

The precision, recall and F-Measure of the proposed algorithm are also shown in tabular format below:

Table 1: Comparison between Proposed and Pre-existed

	Precision	Recall	F-Measure
SVD	0.86	0.78	0.81
RBF_K FCM	0.9	0.9	0.9
PNN	0.98	0.98	0.98
<b>Proposed</b>	<b>0.99</b>	<b>0.7</b>	<b>0.95165</b>

## V. CONCLUSION AND FUTURE SCOPE

Compelling and important assessment of recommender frameworks is challenging. Till date, there has been no distributed endeavour to orchestrate what will be thought about the assessment of recommender frameworks, nor to deliberately comprehend the ramifications of assessing recommender frameworks for diverse assignments and distinctive connections. In this article, we have endeavoured to review the components that have been considered in assessments and presented new variables that we accept ought to be considered in assessment. Moreover, we have presented experimental results on exactness measurements that give some introductory knowledge into how comes about because of distinctive assessment measurements may fluctuate. Our trust is that this article will build the attention to potential inclinations in reported assessments, expand the differences of assessment measurements analysed where it is fundamental, and energize the advancement of more institutionalized routines for assessment.

As observed from the result the new improved Recommender system based on weighted Manhattan Distance and Agglomerative Hierarchical cluster gives better score and running time.

User Sensitivity to Algorithm Accuracy. It is known fact that client fulfilment is diminished when a critical level of slip is brought into a recommender framework. The level of blunder presented in that study, on the other hand, was

ordinarily bigger than the contrasts between the best calculations. Key inquiries meriting consideration include: (a) for diverse measurements, what is the level of progress required before clients notification or client conduct changes? (b) To which measurements are clients generally delicate? (c) How does client affectability to exactness rely on upon different elements, for example, the interface? (d) How do elements, for example, scope and luck influence client fulfilment? On the off chance that these inquiries are replied, it might be conceivable to manufacture a prescient model of client fulfilment that would allow broader logged off assessment.

The recommender system makes use of very large matrices whose space complexity is  $O(mn)^k$ .

Another problem has to deal with too many ratings, for most collaborative filtering systems, having to deal with too few ratings is a far more serious problem. This problem occurs when the amount of items become very large reducing the number of items users have rated to a tiny percentage. In such a situation it is likely that two people have few rated items in common making the correlation coefficient less reliable. This is called the sparsity problem. Several solutions have been proposed to overcome this problem like Dimensionality Reduction, Implicit ratings.

#### **ACKNOWLEDGMENT**

I would like to express my special appreciation and thanks to my advisor Professor Dr.KIRAN JYOTI, you have been a tremendous mentor for me. I would like to thank you for encouraging my

research and for allowing me to grow as a research scientist. Your advice on both research as well as on my career have been priceless. I would also like to thank my committee members, Professor RANINDERKAUR, Professor KAMALJEETKAUR, and Professor AKSHAYGIRDHAR for brilliant comments and suggestions, thanks to you.

#### **REFERENCES**

- [1] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," vol. 17, ed. 2005, pp. 734-749.
- [2] S. Berkovsky, T. Kuflik, and F. Ricci, "Mediation of user models for enhanced personalization in recommender systems," *User Modelling and User-Adapted Interaction*, vol. 18, pp. 245-286, 2008.
- [3] R. Burke, "Hybrid web recommender systems," ed. 2007, pp. 377-408.
- [4] M. K. K. Devi, R. T. Samy, S. V. Kumar, and P. Venkatesh, "Probabilistic neural network approach to alleviate sparsity and cold start problems in collaborative recommender systems," *2010 IEEE International Conference on Computational Intelligence and Computing Research, ICCIC 2010*, pp. 286-289, 2010.
- [5] Y. M. Li, C. T. Wu, and C. Y. Lai, "A social recommender mechanism for e-commerce: Combining similarity, trust, and relationship," *Decision Support Systems*, vol. 55, pp. 740-752, 2013.
- [6] N. Mittal, R. Nayak, M. C. Govil, and K. C. Jain, "Recommender System Framework Using Clustering and Collaborative Filtering," *2010 3rd International Conference on Emerging Trends in Engineering and Technology*, pp. 555-558, 2010.
- [7] M. J. Pazzani and D. Billsus, "Content-Based Recommendation Systems," *The Adaptive Web*, vol. 4321, pp. 325-341, 2007.
- [8] L. Yu and A. Dong, "Hybrid product recommender system for apparel retailing customers," *Proceedings - 2010 WASE International Conference on Information Engineering, ICIE 2010*, vol. 1, pp. 356-360, 2010.
- [9] X. Zhu, J. Huang, and Y. Qi, "An apparel recommender system based on data mining," *Proceedings - 2010 International Conference on Web Information Systems and Mining, WISM 2010*, vol. 1, pp. 48-52, 2010.