ISSN:0975-9646

G.P Karpagam et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (6) , 2015, 5114-5117

Proficient Enhancement of Accuracy for Movie Rating Recommendation System via Algorithmic Approaches

G.P Karpagam[1],S.Balaji[2]

Department of Computer Science and Engineering,
Akshaya college of engineering and technology,India.

Abstract-Online readers require tools to help them cope with the enormous of content available on the world-Wide Web. Selections are made by readers in traditional media with the help of assistance. Recommender system based on web data mining is very useful, more exact and provides worldwide services to the user. Recommender systems analyze patterns of user interest in items or products to provide recommendations for items that will suit a user's taste.This includes both implicit intervention in the form of editorial oversight and explicit aid in the form of recommendation services such as movie reviews and restaurant guides.Several opportunities are provided by the electronic medium to offer recommendation services, ones that adapt over time to trace their evolving interests.Both content-based and collaborative systems can provide such a examine, but individually they both face shortcomings. To improve the accuracy various techniques are used.Main proposal of the project is the Singular value decomposition and Naive bayes classification to increase the accuracy of movie rating recommendation system.

Index Terms - Recommender system, recommendation stability,iterative smoothing,Singular value decomposition and Naive bayes classification.

1 INTRODUCTION

Recommendation systems (RS) provide technique to select the relevant items from the vast data available in the web by predicting the "rating". Recommend useful and interesting items to users in order to increase both sellers profit and users satisfication. RS donate to the commercial success of many on-line ventures such as Amazon.com or Netflix. Often a RS attempts to predict the rating a user will give to items based on her past ratings and the ratings of the other users. The input used for the recommender systems includes explicitly provided feedback in the form of ratings or tags, as well as feedback that can be implicitly contingent by monitoring users behavior such as browsing, linking, or buying patterns. For example an online movie rental company Netflix, ask the user to rate movies using 5 star numeric scales. In the recommender systems literature, calculating performance of recommendation algorithms has been the key issue and recommendation accuracy and stability has been the main focus in evaluating metrices [1],[2].

Research in the recommender systems area helps to propose new technique to enhance the accuracy and stability of the recommendation algorithm. Accuracy typically compare the rating values estimated by the recommendation algorithm against the actual rating values and reflect the proximity of the system's prediction to users true ratings. Stability is designed to capture the level of internal consistency among predictions made by the recommendation algorithm.Stability is the important and essential property to avoid inconsisitent recommendation may create negative impact on the user's acceptance of future suggestion by the system. Different level of stability will be exhibited by the different recommendation algorithm with similar accuracy.Maximizing accuracy may not necessarily help in increasing stability. Recommender systems are a useful alternative to search algorithms since they help users discover items they might not have found by themselves. Interestingly enough, recommender systems are often implemented using search engines indexing non-traditional data.

2 WEB RECOMMENDATION METHOD

2.1 CONTENT BASED FILTERING

Content based filtering perform its work by using the profiles created for the user at the beginning. Recommends items based on a comparision between the content of the items and user profile. Set of descriptors or terms are used to represent the content of the each item.In other words,these algorithms try to recommend items that are similar to these that a user liked in the past.In particular ,various candidate items are compared with items formerly rated by the user and best matching items are recommended.Content based filtering has roots in information recovery and information filtering research.

2.2 COLLABORATIVE FILTERING

Collaborative filtering methods are based on gathering and analyzing a large amount of information on users' behaviours, activities, preferences and predicting users similarity to other users. A key advantage of the collaborative filtering approach is that it does not rely on system analyzable content and it is capable of accurately recommending complex items such as movies without requiring an 'understanding' of the item itself. The recommender system compares the collected data to similar and dissimilar data collected from others and calculates a list of recommended items for the user.Collaborative filtering, also referred to as social filtering, filters information by using the recommendations of other people. It is based on the idea that people who agreed in their evaluation of certain items in the past are likely to agree again in the future. A person who wants to see a movie for example, might ask for recommendations from friends. The recommendations of some friends who have similar interests are trusted more than recommendations from others. This information is used in the decision on which movie to see.

2.3 HYBRID RECOMMENDER SYSTEM

Combining collaborative filtering and content based filtering is the hybrid approach with more effectiveness.Particular approach can be implemented by making collaborative approach and content based approach separately and then combining them by adding content based capabilities to collaborative approach The term hybrid recommender system is used here to describe any recommender system that combines multiple recommendation techniques together to produce its output. There is no reason why several different techniques of the same type could not be hybridized, for example, two different content-based recommenders could work together, and a number of projects have investigated this type of hybrid: NewsDude, which uses both naive Bayes and kNN classifiers in its news recommendations.

3 CHALLENGES AND PROBLEMS OF RECOMMENDATION METHODS

3.1 COLD-START PROBLEM: When you create a profile in a few recommender systems have solved this problem with the survey. They are new in the system and when the item has not previously been declared a cold start it can be. Both of these problems can be solved with the hybrid approach.

3.2 BELIEVE: With a brief history of the people's voice in their rich history, which is as the voice of those that may not be relevant? The issue of trust arises to evaluate a particular client. The problem can be solved by the users for the distribution of preferences.

3.3 SCALABILITY: With the increase of number of users and items, and recommendations for the formation of information processing systems need more resources. This problem also filters and systems are solved by combining different types of physical improvement. Many parts of the computations in order to accelerate the assurance of online recommendations can be applied offline.

3.4 SPARSITY: Users and a large amount of items that online stores those users have rated only a few items are almost always there. Collaborative recommender systems using other methods to access their profiles, users typically create neighborhood. If a user has rated only a few items, it is very difficult to determine his taste and he / she may be wrong neighborhood.

3.5 PRIVACY: The most important issue privacy. To get the most accurate and recommendation systems, demographic data, and the data about the location of a particular user with the most amount of information possible about the user, should receive. Naturally, the information's reliability, security and privacy questions arise. Many online stores by using special algorithms and programs provide effective protection of users' privacy [3].All these challenges and problem should be overcome to provide better accuracy and stability.

4 INITIAL APPROACH : ITERATIVE SMOOTHING

High instability results from predictions that are inconsistent with each other. While bagging is expected to provide some stability benefits, it represents an indirect approach to improving stability, as discussed earlier (i.e., the bagging approach has not been explicitly designed with stability improvement in mind). In this section we propose an iterative smoothing approach, which is aimed more directly at stability improvement. This approach involves multiple iterations for repeatedly and collectively adjusting the rating predictions of a recommendation algorithm based on its other predictions and, thus, explicitly improves consistency of predicted ratings. The key idea of iterative smoothing is that the predictions computed during current iteration will be fed back into the data to predict other instances in subsequent iterations.

Fig. 1 provides an overview of the iterative smoothing approach and gives a high-level illustration of the overall process. Given rating space S and training set D where ratings are known, predictions on unknown ratings S\D are first estimated using some standard recommendation algorithm T.

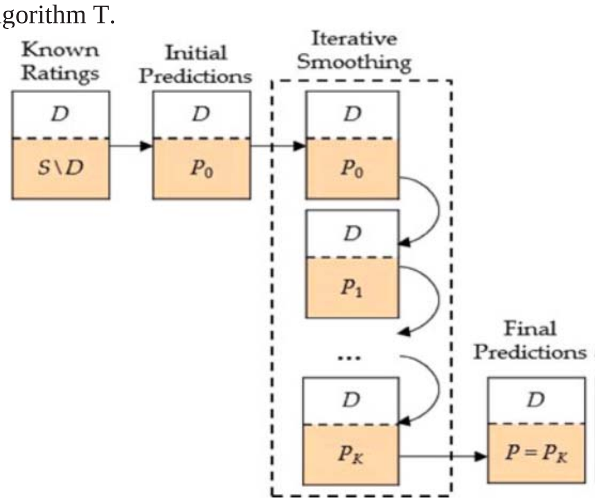


Figure 1: Illustration of the general iterative smoothing process.

These predictions are denoted as P0. Then, the main idea is to iteratively adjust estimations for each rating in S\D based on all other ratings in the rating space S (i.e., both known as well as predicted) in order to proactively improve consistency between different predicted ratings.

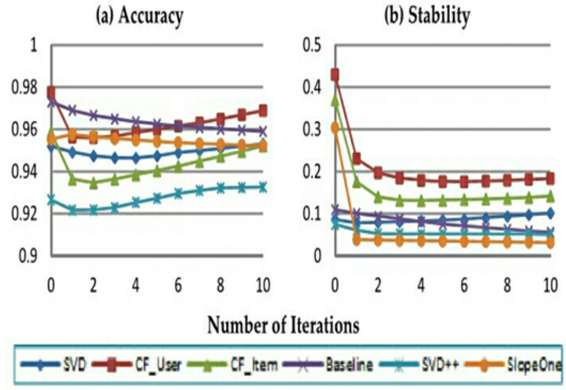


Figure2:Accuracy and stability of iterative smoothing

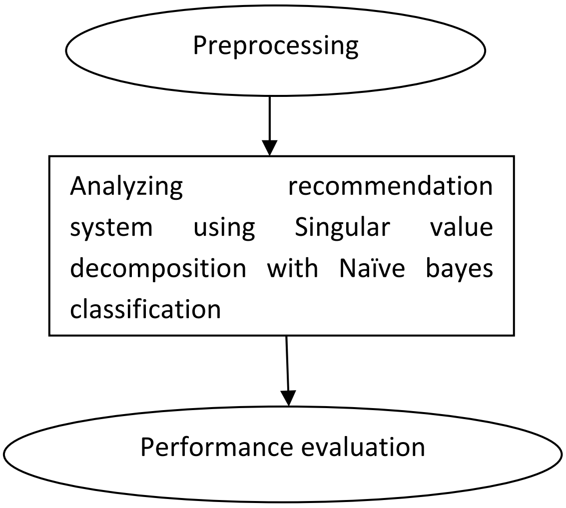
Fig. 2 suggests that the iterative smoothing models can “over-adjust” rating predictions after a number of iterations, in their attempt to maximize the performance on training data. Thus, when configuring iterative smoothing approaches, it is important to be aware of the best number of iterations for a given algorithm, in order to avoid performance deterioration.

5 PROPOSED WORK: SINGULAR VALUE DECOMPOSITION WITH NAIVE BAYES CLASSIFICATION

Fig 3 represents the module split up of the proposed system.Singular value decomposition(SVD) reduces the dimensionality of our dataset and captures the features to reduce the number of users and predict the ratings.SVD is the well known matrix factorization technique that factors an m by n matrix X into three matrices USV^T.

$$\underset{m \times n}{\overset{X}{\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & \\ \vdots & \vdots & \ddots & \\ x_{m1} & & & x_{mn} \end{pmatrix}}} = \underset{m \times r}{\overset{U}{\begin{pmatrix} u_{11} & \cdots & u_{1r} \\ \vdots & \ddots & \\ u_{m1} & & u_{mr} \end{pmatrix}}} \underset{r \times r}{\overset{S}{\begin{pmatrix} s_{11} & 0 & \cdots \\ 0 & \ddots & \\ \vdots & & s_{rr} \end{pmatrix}}} \underset{r \times n}{\overset{V^T}{\begin{pmatrix} v_{11} & \cdots & v_{1n} \\ \vdots & \ddots & \\ v_{r1} & & v_{rn} \end{pmatrix}}}$$

X represent the dataset as a matrix where the users are rows, movies are columns and the individual entries are ratings.The matrix S is a diagonal matrix containing the singular values of the matrix X.There are r singular value,where r is the rank of X.To accomplish simply keep the first k singular values in S,where k<r.This will give as the best rank–k approximation to X.In order to provide a baseline,fill all the empty cells with the average rating for that movie and then compute the singular value decomposition.It is applied on MovieLens 1m Subset to find the movies average rating.



Naive bayes classification to be comparable in performance with decision tree and selected neural networks classifiers.It exhibit high accuracy and speed when applied to large databases.

It is a classification technique assumption of independence among predictors. In simple terms,Naive Bayes classifier assumes that the presence of a particular feature in a class unrelated to the presence of any other feature.It is easy to build and particularly useful for very large datasets.Naive bayes is known to outperform even highly sophisticated methods.Bayes theorem provides a way of calculating posterior probability P(c/x) from P(c),P(x) and P(x/c).

P(c/x) =[P(x/c)P(c)]/P(x)

- P(c/x)is the posterior probability of class (c, target) given predictor (x, attributes).
- P(c) is the prior probability of class.
- P(x/c) is the likelihood which is the probability of predictor given class.
- P(x) is the prior probability of predictor.

It is easy and fast to predict class of test data set. It also perform well in multi class prediction. When assumption of independence holds, a Naive Bayes classifier performs better compare to other models like logistic regression and you need less training data.Naive bayes classification is used to find the correctly rated users,movies.

6 RESULT AND CONCLUSION

Singular value decomposition is used to find the average rating for the movie.Instead of looking for all ratings given by the users for particular movie one can easily analyze the movie average rating using SVD. Naive bayes classification is used to find the correctly rated movies and correctly rated users detail.This work provides several interesting directions for future research. Providing larger improvements for some algorithms and smaller improvements for some others lead to larger improvement in recommendation system.

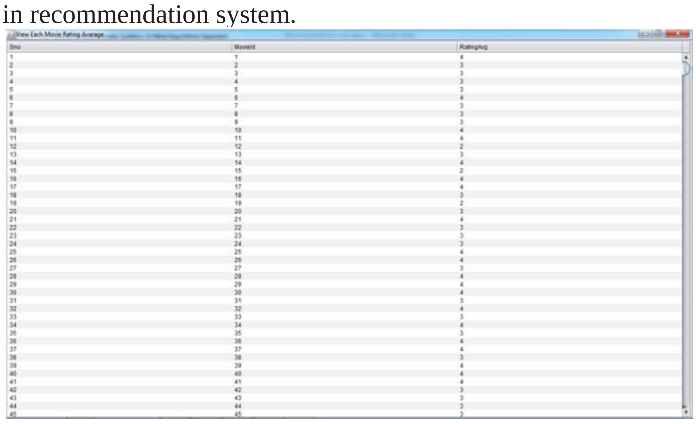


Figure 4: Each movie rating average

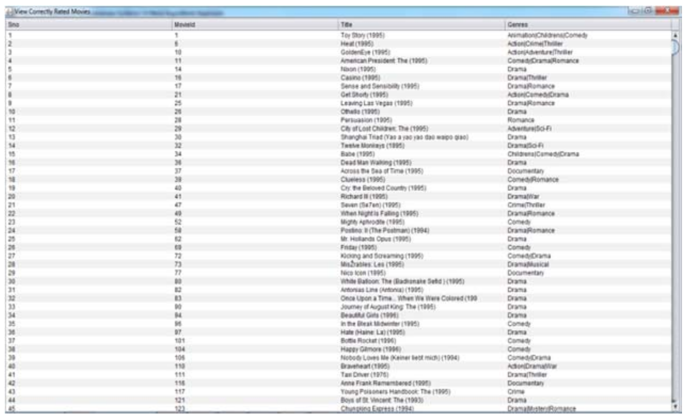
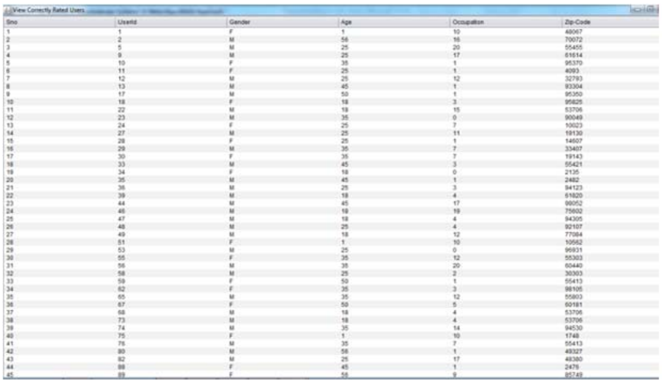


Figure 5: Correctly rated movies



View Correctly Rated Users

Sno	UserId	Gender	Age	Occupation	Zip-Code
1	1	F	1	10	48067
2	2	M	56	16	70072
3	5	M	25	20	55455
4	9	M	25	17	61614
5	10	F	35	1	95370
6	11	F	25	1	4093
7	12	M	25	12	32793
8	13	M	45	1	93304
9	17	M	50	1	95350
10	18	F	18	3	95825
11	22	M	18	15	53706
12	23	M	35	0	90049
13	24	F	25	7	10023
14	27	M	25	11	19130
15	28	F	25	1	14607
16	29	M	35	7	33407
17	30	F	35	7	19143
18	33	M	45	3	55421
19	34	F	18	0	2135
20	35	M	45	1	2482
21	36	M	25	3	94123
22	39	M	18	4	61820
23	44	M	45	17	98052
24	46	M	18	19	75602
25	47	M	18	4	94305
26	48	M	25	4	92107
27	49	M	18	12	77084
28	51	F	1	10	10562
29	53	M	25	0	96931
30	55	F	35	12	55303
31	56	M	35	20	60440
32	58	M	25	2	30303
33	59	F	50	1	55413
34	62	F	35	3	98105
35	65	M	35	12	55803
36	67	F	50	5	60181
37	68	M	18	4	53706
38	73	M	18	4	53706
39	74	M	35	14	94530
40	75	F	1	10	1748
41	76	M	35	7	55413
42	80	M	56	1	49327
43	82	M	25	17	48380
44	88	F	45	1	2476
45	89	F	56	9	85749

Figure 6: Correctly rated users

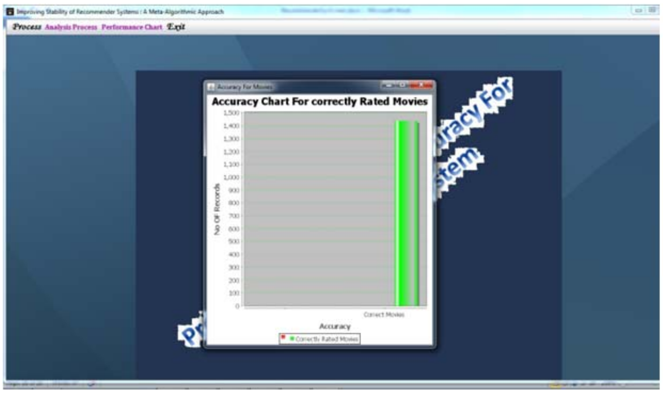


Figure 7: Accuracy chart for correctly rated movies

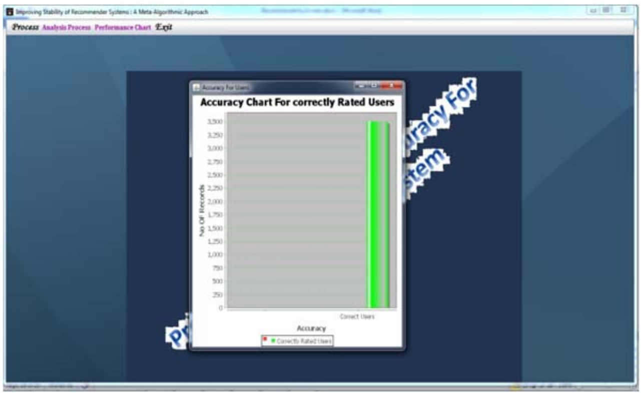


Figure 8: Accuracy chart for correctly rated users

Figure 9: Performance comparision for wrongly and correctly rated movies

Fig 4 illustrate the average rating of each movie.Fig 5 represents the correctly rated movies.Fig 6 represents the correctly rated users.So the user can view the profile of all those correctly rated users to know more about other movies.Another interesting direction would be to perform user behavior studies to investigate the value of stable recommendations (as opposed to unstable recommendations) on users' usage patterns and acceptance of recommender systems.In future K-Means clustering algorithm is decided to use to increase the accuracy and stability of the recommendation system.

REFERENCES

[1] J. L. Herlocker, J. A. Konstan, K. Terveen, and J. T. Riedl,"Evaluating collaborative filtering recommender systems," ACM Trans. Inform. Syst., vol. 22, no. 1, pp. 5–53, 2004.

[2] G. Shani and A. Gunawardana, "Evaluating recommendation systems,"in Recommender Systems Handbook, F. Ricci, et al. Eds. NewYork, NY, USA: Springer, 2011., pp. 257–294.

[3] R. Burke, "Hybrid Recommender System: Survey and Experiments,"*User Modeling and User-Adapted Interac- tion*, Vol. 12, No. 4,2002, pp. 331-370

[4] D. Dubois, E. Hüllermeier and H. Prade, "A Systematic Approach tothe Assessment of Fuzzy Association Rules," *Data Mining and Knowledge Discovery Journal*, Vol. 13, No. 2, 2006, pp. 167-192. doi:10.1007/s10618-005-0032-4

[5] Agrawal R, Imielinski T & Swami A N, (1993) "Mining association rules between sets of items in large databases", *ACM SIGMOD International Conference on Mgt. of Data*, Vol.22, Issue 2, pp.207-216.

[6] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommendation system: A survey of the state-of-the-art and possible extensions," IEEE Trans. Knowl. Data Eng., vol. 17, no. 6,pp. 734–749, Jun. 2005.

[7] G. Adomavicius and J. Zhang, "On the stability of recommendation algorithms," in Proc. ACM Conf. Recommender Syst., 2010,pp. 47–54.

AUTHORS

Balaji Subramani currently working as Assistant Professor in Akshaya College of Engineering and Technology. He has 4 years of Teaching Experience. He did his Master Degree M.Tech (IT) in K.S.Rangasamy College of Technology. He has participated in various International Conferences and workshops held at different places. His area of interest includes Word sense disambiguation, web mining, Information retrieval and social network analysis.

G.P Karpagam currently pursuing Master Degree(CSE) in Akshaya College of Engineering and Technology.Area of interest includes Image processing and Web mining.