State of the Art of Prediction and Recommender System

Bhakti Ratnaparkhi
M.E. Student, Department of Computer Engineering,
Pimpri Chinchwad College of Engineering
Pune University, Pune

J. S. Umale, Ph.D.
Professor, Department of Computer Engineering,
Pimpri Chinchwad College of Engineering
Pune University, Pune

ABSTRACT
Recommender system is the system which gives suggestions. It takes help of prediction system to give recommendations. Prediction system will do predictions about future actions. Recommender system provides top ranked predictions as recommendations. It is very essential to do correct prediction for giving best recommendation. In order to improve quality of recommender system researchers have been trying different approaches which we are see through this survey paper.

Keywords
Recommender system, Techniques, Student's Psychology

1. INTRODUCTION
Competition in every field is increasing day by day. To stand up firmly in such situation people are coming up with new ideas. Same is the condition in word of computation. Recommender systems are used to provide suggestions and indirectly to attract customers. Recommender system is system which predicts items or product in which user might be interested [1]. Example can be system which suggests movie names to user according to requirements and interest or system which suggests songs list as per user's liking.

In routine life people take suggestions from friends or from people those are close to them because of trust that they have on them as shown in Fig. 1. Suggestion will be useful to a person if it is accurate, and accuracy of such suggestion depends on how much other person who is making suggestion knows about receiver. Stronger the relationship accurate will be recommendation given.

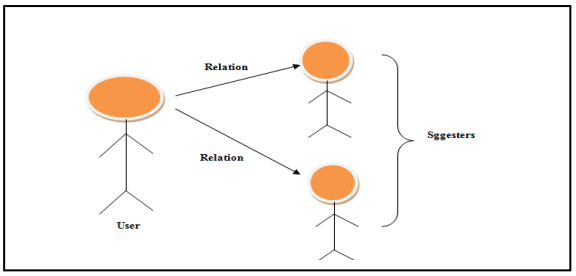

Fig 1: Recommendation in routine life

Experts from computation field are trying to bring same idea in the computer technology's word. Interest of users will be captured and it will be studied to recommend them products. Capturing interest can be done by either studying user's history or current behavior pattern. It can be compared with other users with similar interest to provide suggestions.

Recommendation can be of various types: Content based, Collaborative, Using data mining techniques or by using statics methods as shown below [1]. We will see in next section how various researchers have used these techniques to provide Recommendation Systems for different applications.

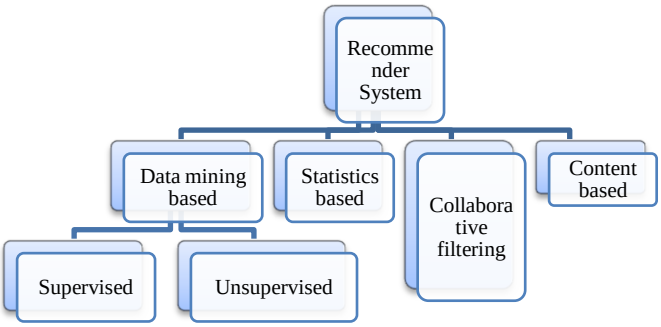

Fig 2: Types of Recommender System

Content based is a mechanism which tries to match user requirements to everything available and provide best suited once. In collaborative filtering users with same interest are searched and items preferred by them are suggested. Recommender systems use prediction techniques from data mining such as Clustering or Classification. All clustering techniques are unsupervised where class labels are not known previously and all classification techniques are supervised where class labels are known previously. Statistics based methods such as Chi-square test or Pearson's co-relation is also used in Recommender systems [2].

2. TECHNIQUES USED

2.1 Statistical Method
Manos Papagelis [2] studied many statistics based method and among them selected Pearson's co-relation method to find user with similar interest. Matrix was used to store information related to items and customers as well as group of items. Comparison was made among user based; item based, implicit rating and explicit rating. Here data set used was movie recommendation system, named MRS.

Sakchai Tangwannawit [15] used SVM technique. Data set used was collected from 304 vocational students of the academic year 2012 at Singburi Vocational College.

2.2 Data Mining based Methods
Shuai Zhang [3] developed a system for people with Dementia. Data of 40 patients was collected. Various data mining techniques such as feature selection, feature reduction also different classification algorithms like decision tree (DT), Naïve Bayes (NB) and k-nearest neighbor (kNN) were applied on it. Results obtained were compared.

Yang Guo [4] used Bayes network for developing a system for people having Type-2 diabetes. Diabetes is of two types: Type-1 and Type-2. Type-1 is found in children, which cannot be controlled without insulin. But Type-2 is found in elderly people and it can be controlled without insulin provided it is detected accurately on time. Author developed structure

showing dependency of attributes on one another. The data set used was from the UCI Machine Learning Repository. The original owner of this dataset is the National Institute of Diabetes and Digestive and Kidney Diseases.

Bayes network can be used when attributes are dependent on each other. Naïve Bayes can be used when attributes are totally independent [7].

Ayman Khedr [5] used FT-tree approach to diagnose Liver disease. The method was very helpful than traditional method used for diagnosis of Liver disease. Combination of FT and AT test was used. FT was to detect fibrosis is present or not and AT was to detect how active is the fibrosis. Newly invented method was less painful, it did not required hospitalization plus test was very easy to perform number of times which helps to see the progress of treatment. The dataset contained data on laboratory examinations, which were collected on Electricity Hospital in Egypt.

Samuel and Omisore [6] built a system which was combination of fuzzy system and neural network for prediction of Typhoid. Attributes used were fuzzy in nature to normalize values of fuzzy system author used neural network. Data was collected from hospital administrative.

Apriori algorithm can use to find frequently occurring patterns [8]. This generated patterns can use to recommend products or some important decision making process. In this way Recommender system can use data mining algorithms as Neural Network, Decision Tree, Bayes Network etc [9] [10] [11] [12].

2.3 Content based

Content based recommendation is used for recommending news, web sites and books as per user's preference [1] considering user's requirement.

2.4 Collaborative Filtering

Collaborative Filtering technique is used in social networking sites to suggest music, jokes [1]. Online shopping websites also use this technique to recommend items that user may want to buy. This technique has some limitations which will be discussed in next section of this document.

Now researchers are considering psychological factors as well such as Emotional Intelligence, Satisfaction, Psychological Effect and Personality-based to provide recommendation [1].

3. DISCUSSION AND COMPARISION

3.1 Comparison among Techniques used from Data Mining

Table 1. Comparison among data mining based methods

Criteria	Bayes Network	FT-tree	Neuro Fuzzy
Advantages	1. Gives more accurate results 2. Faster than Neural network	1. Simple 2. Fast	1. Can deal with vague attributes (FN) 2. Processing is similar to human brain (NN)
Disadvantages	Costly	Cannot deal with vague attributes	1. Complex 2. Costly 3. Time consuming
Data Set	The Pima Indians Diabetes Dataset (having lot of constrains)	From laboratory of hospital (Random sampling not used)	From hospital administration
Application	Diabetes Diagnosis	Liver disease Diagnosis	Typhoid Diagnosis

3.2 Difference between Naive and Bayes Network [4] [7]

- Naive Bayesian network assumes that effect one attribute is independent of effect of other attribute on result.
- Meaning of Naïve is strongly independent. Naïve Bayesian doesn't allow dependency among attributes.
- Whereas Bayesian network (Bayesian Believe network) allows dependency among attributes.
- Using Bayesian network attributes can be represented in the form of Directed Acyclic Graph (DAG).
- In real life examples attributes are dependent most of the time especially in medical domain. So Bayesian network is more useful in real life applications.
- Naive Bayesian is most accurate method of classification.

3.3 Problems with Collaborative Filtering

- **Cold Start**: There should be enough other users already in the system to find a match with similar likings.
- **Sparsity**: Due to large number of items to be recommended also large number of users matrix becomes sparse in nature which makes matching process difficult.
- **First Rater**: Cannot recommend an item that has not been previously rated e.g. new items.
- Recommendation cannot be provided to user with unique test using this technique.

3.4 Overall Comparison

Table 2. Comparative Analysis Different Mechanism And Applications

Title	Data Set Used	Data collection method (Standard/ Collection)	Mechanism	Application
1. Qualitative Analysis of User-based and Item-based Prediction Algorithms for Recommendation Systems movie recommendation system (MRS) Standard data set Pearson's Co-relation Movie Recommendation	movie recommendation system (MRS)	Standard data set	Pearson's Co-relation	Movie Recommendation
2. Comparing the Strengths and Difficulties Questionnaire (SDQ) and Behavior Consideration Assessment Using SVM Techniques	Singburi Vocational College students, 2012	Collected	SVM	Comparing the Strengths and Difficulties Questionnaire
3. A Predictive Model for Assistive Technology Adoption for People With Dementia	Data of 40 patients	Collected	feature selection, feature reduction also different classification algorithms like decision tree (DT), Naïve Bayes (NB) and k-nearest neighbor (kNN)	Assistive Technology Adoption for People With Dementia
4. Using Bayes Network for Prediction of Type-2 Diabetes	from the UCI Machine Learning Repository	Standard data set	Bayes Network	Diagnosis of Type-2 Diabetes
5. Business Intelligence framework to support Chronic Liver Disease Treatment	from Electricity Hospital in Egypt	Collected	FT-Tree	Diagnosis of Liver Disease
6. Hybrid Intelligent System for the Diagnosis of Typhoid Fever	from hospital administrative	Collected	Fuzzy logic and Neural Network	Diagnosis of Typhoid

4. CONCLUSION

We have studied approaches used in prediction and recommender system. Comparison was performed among them. Study says that among all techniques collaborative filtering has many challenges. Quality can be improved if data mining techniques combined with statistical methods like similarity measure. Pearson's co-relation is the best similarity measure. As a extension to this study in future we can develop a technique with greater accuracy by combining some of these approaches.

5. REFERENCE

[1] Maria Augusta S. N. Nunes, "Towards to Psychological-based Recommenders Systems: A survey on Recommender Systems", SCIENTIA PLENA VOL. 6, NUM. 8 2010.

[2] Manos Papagelis, Dimitris Plexousakis, Ioannis Rousidis and Elias Theoharopoulos,"Qualitative Analysis of User-based and Item-based Prediction Algorithms for Recommendation Systems".

[3] Shuai Zhang, Sally I. McClean, "A Predictive Model for Assistive Technology Adoption for People With Dementia", IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, VOL. 18, NO. 1, JANUARY 2014.

[4] Yang Guo, Guohua Bai, Yan Hu, "Using Bayes Network for Prediction of Type-2 Diabetes", 2012, IEEE, 7th International Conference for Internet Technology and Secured Transactions (ICITST).

[5] Ayman Khedr,"Business Intelligence framework to support Chronic Liver Disease Treatment", International Journal of Computers & Technology Volume 4 No. 2, March-April, 2013, ISSN 2277-3061.

[6] Samuel and Omisore, "Hybrid Intelligent System for the Diagnosis of Typhoid Fever", J Comput Eng Inf Technol 2013, 2:2, Journal of Computer Engineering & Information Technology.

[7] "Diagnosis of Heart Disease for Diabetic Patients using Naive Bayes Method", International Journal of Computer Applications (0975 – 8887) Volume 24– No.3, June 2011.

[8] "Finding Locally Frequent Diseases Using Modified Apriori Algorithm", International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 10, October 2013.

[9] "Importance of Artificial Neural Network in Medical Diagnosis disease like acute nephritis disease and heart disease", International Journal of Engineering Science and Innovative Technology (IJESIT) Volume 2, Issue 2, March 2013.

[10] "Lung cancer differential diagnosis based on the computer assisted radiology: The state of the art"

[11] "The Application of Machine Learning Technique for Malaria Diagnosis"

[12] "Performance Evaluation of Levenberg-Marquardt Technique in Error Reduction for Diabetes Condition Classification", International Conference on Computational Science, ICCS 2013.

[13] "An Investigation into the Feasibility of Detecting Microscopic Disease Using Machine Learning", Keynote Lecture of IEEE International Conference on Bioinformatics and Biomedicine November 2-4, 2007, Silocon Valley, California, USA.

[14] Arturas Kaklauskas, Edmundas Kazimieras Zavadskas, Vaidotas Trinkunas, Laura Tupenaite, Justas Cerkauskas, Paulius Kazokaitis, "Recommender system to research students' study efficiency", Procedia - Social and Behavioral Sciences 51 (2012) 980 – 984.

[15] Sakchai Tangwannawit and Montean Rattanasiriwongwut, "Comparing the Strengths and Difficulties Questionnaire (SDQ) and Behavior Consideration Assessment Using SVM Techniques", DOI: 10.7763/IPEDR. 2014. V70. 16.