Student Psychology Prediction and Recommendation System Using Rough Set Theory

Bhakti Ratnaparkhi[1*], Lokesh Katore[2], J. S. Umale[3] and Niharika Upadhyay[4]

[1*,2,3,4]Computers, Pune, India

Abstract— Big data analysis includes many theories and methods for prediction system. Statistical methods such as Person's correlation, Regression analysis and Rough Set Theory etc are being used for predicting facts. Also theory like collaboration filtering uses word's filtering to predict and provide recommendations. We have studied all these methods and selected most appropriate method for student's psychology prediction. In our proposed work we have used Rough sets to extract the rules for prediction of student's psychology. Rough Set is a comparatively recent method that has been effective in various fields such as medical, geological and other fields where intelligent decision making is required. Our experiments with rough sets in predicting student's psychology produced attractive results.

Keywords— Psychology; Prediction; RST

I. INTRODUCTION

Prediction systems are used to predict future scenarios. E.g. prediction system can be used to in medical domain to predict whether particular disease is present in patient. It has been observed that number of students facing psychological problem is increasing per year. Psychologically disturbed person cannot work efficiently. Same problem is observed in students because of which their educational performance gets affected. To control this situation a prediction system is needed which will accurately predict student's current psychological state.

Recommender system takes help of prediction system to provide recommendations to user. Top ranked predictions are provided as recommendations to user. In real life we take suggestions from our friends or elders because we have trust relation with them. Researchers working in this domain are trying to bring this trust relation in automated systems by increasing accuracy of prediction and recommendations provided from such systems. We need recommender system which will give remedial suggestions to make change in psychological state and improve performance. With pure intension of applying technical knowledge to solve social problem we have proposed our model of prediction & recommendation approach for student's psychology.

Various methods are available for prediction. By studding these methods we are applying a Rough Set Theory for student's psychology prediction. Rough set theory is useful for rule induction from incomplete data sets. Using this approach we can distinguish between three types of missing attribute values: lost values (the values that were recorded but currently are unavailable), attribute-concept values (these missing attribute values may be replaced by any attribute value limited to the same concept), and "do not

care" conditions (the original values were irrelevant). A concept (class) is a set of all objects classified (or diagnosed) the same way. Rough set reduction technique is applied to find all the reducts of the data, which contains the minimal subset of attributes that are associated with a class label for prediction. Finally, rough sets dependency rules are generated directly from all generated reducts.

II. LITERATURE SURVEY

A. Various Techniques

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.

Rough sets can be use to extract the rules for prediction of personality traits. Rough Set is a recent statistical method that has been used in various fields such as medical, geological and other fields for intelligent decision making. Author performed experiments with rough sets to predict personality traits [17].

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 after implementation were analyzed.

Yang Guo [4] used Bayes network for developing a system for people having Type-2 diabetes. Diabetes can be categorized as 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. Graph can be use to show dependencies among attributes. 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].

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

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].

B.   Rough Set Theory

Rough set theory is a new mathematical approach to imperfect knowledge. The problem of imperfect knowledge has been tackled for a long time by philosophers, logicians and mathematicians. Recently it became also a crucial issue for computer scientists, particularly in the area of artificial intelligence. There are many approaches to the problem of how to understand and manipulate imperfect knowledge. The most successful one is, no doubt, the fuzzy set theory. Rough set theory proposed by the author in [18] presents still another attempt to this problem. The theory has attracted attention of many researchers and practitioners all over the world, who contributed essentially to its development and applications. Rough set theory has an overlap with many other theories. However we will refrain to discuss these connections here. Despite of the above mentioned connections rough set theory may be considered as the independent discipline in its own rights. Rough set theory has found many interesting applications. The rough set approach seems to be of fundamental importance to AI and cognitive sciences, especially in the areas of machine learning, knowledge acquisition, and decision analysis, knowledge discovery from databases, expert systems, inductive reasoning and pattern recognition. The main advantage of rough set theory in data analysis is that it does not need any preliminary or additional information about data – like probability in statistics, or basic probability assignment in Dempster-Shafer theory, grade of membership or the value of possibility in fuzzy set theory.

In day to day life, we come across with the incomplete or imprecise information or knowledge to understand our surroundings, to learn new things, and to make plans for the future. Rough Set has been introduced by Pawlak in 1982 as a tool to deal with, uncertain knowledge. Its philosophy is based on the assumption that, in contrast to the classical set. This theory uses different approach to uncertainty. The main concept of this Theory is Lower and Upper Approximations. The main advantage of this theory is that it doesn't need any base about data like probability in statistics or grade of membership in fuzzy set theory.

The theory has found many interesting applications in medicine, pharmacology, business, banking, market research, engineering design, conflict analysis, image processing, decision analysis, and other fields. Advantage of this theory is that it allows analyzing both [19].

Due to the imprecision, which exists in real world data, there is sometimes conflicting classification of objects contained in a decision table. The conflicting classification occurs whenever two objects have matching descriptions, but are deemed to belong to different decision classes. In such cases, the decision table is said to contain inconsistencies. Often we wonder whether there are features in the information system, which are more important to the knowledge represented in the equivalence class structure than other features and whether there is a subset of features

which by itself can fully characterize the knowledge in the database. Such a feature set is called a reduct. Calculation of reducts of an information system is a key issue in RS theory and we use reducts of an information system in order to extract rule-like knowledge from an information system [20].

III. PROPOSED WORK

Student	P1	P2	P3	P4	P5	Prediction
1	20.1	25	73	70	65	Satisfied
2	60	65	62	63.3	66.6	Average
3	40	41	60.1	65	70	Satisfied
4	65	70.3	67.3	63	62	Satisfied
5	36.8	24	27.4	23.2	30	Bellow Average
6	45.78	40.4	63.2	67.4	61	Satisfied
7	25.4	38	73	71.4	69.9	Satisfied
8	70.8	71.1	65	40	20.4	Bellow Average
9	22	28.7	40.2	39.99	25	Bellow Average
10	78	75.2	87	84.3	85	Excellent

a.   Data Set

Table a shows part of data set which has 60 student's data. P1 to P5 are parameters based on which we are predicting student's psychological state. Students were given questioner and asked to give their responses for same. Values for attributes are assigned based on response given by students. For example we have given data of 10 students.

Student	P1	P2	P3	P4	P5	Prediction
1	A	A	C	C	C	Social Personality
2	C	C	C	C	C	Common Attitude
3	B	B	C	C	C	Social Personality
4	C	C	C	C	C	Social Personality
5	B	A	B	A	B	Confused State
6	B	B	C	C	C	Social Personality
7	B	B	C	C	C	Social Personality
8	C	C	C	B	A	Confused State
9	A	B	B	B	A	Confused State
10	D	D	D	D	D	Positive Attitude

b.   Processed Data

Table b is nothing but decision table which used for rough set theory. 1 to 10 are objects, P1 to P5 are attributes and last is decision attribute. It is difficult to operate on so many different numeric values so we then convert these numbers into range values. Range is given bellow. E.g. if number fall in between 1 to 20 then it will come under A category. 1 to 25- A, 26 to 50- B, 51 to 75- C and 76 to 100- D.

SR. No	P1	P2	P3	P4	P5	D
1	D	D	D	D	D	1
2	A	A	C	C	C	2
3	A	A	A	A	A	3
4	A	B	B	B	A	4
5	B	A	B	B	A	4
6	B	B	C	C	C	2
7	B	C	C	D	C	2
8	B	A	B	A	B	4
9	C	C	C	C	C	5
10	C	C	C	C	C	2
11	C	C	C	B	B	4
12	C	C	B	B	B	5
13	C	C	C	D	D	2
14	C	C	C	C	D	2

c.   Reduced Data Set

Table c shows 14 Equivalent classes among all data. Here 9, 10 is case of uncertainty as decision is different even if values of all attributes are same. So when value of all attributes is C decision can 2 or 5. Meaning of Decision 1 to 5 is given below:
1- Positive Attitude, 2- Social Personality, 3- Depression, 4- Confuse State, 5- Common Attitude.

IV. DECISION ALGORITHM

1. If   (P1=4) Then   (D=1)
2. If   (P2=2 and P5=3) Then   (D=2)
3. If   (P2=3 and P4=4) Then   (D=2)
4. If   (P3=3 and P5=4) Then   (D=2)
5. If   (P2=1 and P4=3) Then   (D=2)
6. If   (P3=1) Then   (D=3)
7. If   (P4=1 and P5=2) Then   (D=4)
8. If   (P4=2 and P5=1) Then   (D=4)
9. If   (P4=2 and P5=2) Then   (D=5)
10. If   (P1=3 and P5=3) Then   (D=2) or
    (D=5)

Above is list of all rules extracted from given data. ROSE2 software is used to generate these rules. Here we have considered A=1, B=2, C=3 & D=4 for attribute's values. E.g. If P1 is 4 i.e. D Then Decision attribute will be 1 i.e. Positive attitude as per 1st rule. We have seen case of uncertainty in previous section. 10th rule is Approximation rule which is for case of uncertainty. When P1 and P5 both are 3 i.e. C then decision can be either 2 or 5 i.e. prediction can be Social personality or Common attitude.

V. DECISION ALGORITHM

SR .No	P1	P2	P3	P4	P5	D	Accuracy
1	D	D	D	D	D	1	100
2	A	A	C	C	C	2	100
3	A	A	A	A	A	3	100
4	A	B	B	B	A	4	100
5	B	A	B	B	A	4	100
6	B	B	C	C	C	2	100
7	B	C	C	D	C	2	100
8	B	A	B	A	B	4	100
9	C	C	C	C	C	5	33.33
10	C	C	C	C	C	2	71.42
11	C	C	C	B	B	4	100
12	C	C	B	B	B	5	100
13	C	C	C	D	D	2	100
14	C	C	C	C	D	2	100

d. Results

Above table shows accuracy for all cases. Accuracy can be calculated using formula;
Accuracy= (Lower approximation)/ (Upper approximation)
As 9, 10 is case of uncertainty, accuracy should be calculated for only these 2 cases. For all other cases as there is no conflict in decision accuracy will be 100%. For case 9;

lower approximation (for D = 5) is 1 and upper approximation is 2+1 so accuracy will be (1/3) = 0.33 i.e. 33.33%. For case 10; lower approximation (for D=2) is 5 and upper approximation is 5+2 so accuracy will be (5/7) = 0.71 i.e.71.4%.

CONCLUSION

We studied applicability of Rough set theory to student's psychology prediction system. Basic concept of mathematics, the set, leads to antinomies, i.e., it is contradictory. In our proposed work applied RST for student psychology prediction and extracted rules using ROSE2 software. Accuracy was calculated for all cases.

REFERENCES

[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, DimitrisPlexousakis, IoannisRousidis 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, GuohuaBai, Yan Hu, "Using Bayes Network for Prediction Of Type-2 Diabetes", 2012, Ieee, 7th International Conference For Internet Technology And Secured Transactions (Icitst).
[5]AymanKhedr,"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 ComputEngInfTechnol 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] ArturasKaklauskas, EdmundasKazimierasZavadskas, VaidotasTrinkunas, Laura Tupenaite, Justas Cerkauskas, PauliusKazokaitis, "Recommender system to research students' study efficiency", Procedia - Social and Behavioral Sciences 51

(2012) 980 – 984.

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

[16] Bhakti Ratnaparkhi, Prof. Dr.J. S. Umale, "State of the art of Prediction and Recommender System", International Journal of Computer Applications (0975 – 8887) Volume 108 – No. 11, December 2014.

[17] Umang Gupta, Niladri Chatterjee, "Personality Traits Identification using Rough sets based Machine Learning", IEEE 2013 International Symposium on Computational and Business Intelligence.

[18] Niharika Upadhyay, Pragati Jain, "Applying Rough Set Theory In Feedback Analysis", International Journal of Integrative Sciences, Innovation and Technology.

[19] Hameed Al-Qaheri, Aboul Ella Hassanien, Ajith Abraham, "A Generic Scheme for Generating Prediction Rules Using Rough Set".

[20] Zdzisław Pawlak, "Rough Sets", Institute of Theoretical and Applied Informatics, Polish Academy of Sciences, ul. Bałtycka 5, 44 100 Gliwice, Poland.