SE International Journal of Computer Sciences and Engineering Open Access

Volume-3, Issue-5

Study & Analysis of Statistical Methods and Their Application to Career Prediction System

Lokesh S. Katore^{1*}, Bhakti S. Ratnaparkhi², Jayant S. Umale³ and Niharika Upadhyay⁴

^{1,2,3,4} Dept. of Computer Engineering, Savitribai Phule Pune university, India

www.ijcseonline.org

Received: Apr/22/2015Revised: May/02//2015Accepted: May/16/2015Published: May/30/ 2015Abstract—Now a day's prediction systems are key software systems in business and social contexts. We observe the change in
decision making and success obtained by organization and individuals due to accurate prediction systems. Various algorithms
and theories are used by researchers to formulate prediction systems. We feel the study of state of art methods for design of
prediction systems is highly necessary. This work studies the mathematical approaches used for prediction. We propose the
application of Rough Set Theory for design of career prediction system.

Keywords— RST(Rough Set Theory), Accuracy, Prediction System.

I. INTRODUCTION

In Today's world Career recommendation to the First Year Engineering students is a Herculean task.[15] From earlier days we look forward to recommend or predict something. Prediction is a Greek word, where pre means before and diction means future. Fortune teller used to see the palm and predict the future. Likewise Meteorologist used to analyze the scientific data and tell about the weather, climate, rainfall, snowfall, etc. [1].

The college student's career counseling Expert System is composed of the basic and intelligence career counseling, career counseling solutions management, auxiliary decision information management, and making, evaluation management.[2] The basic and intelligence career counseling agent: this is the core module of the system. Based on the experience and knowledge of career counseling, we can build a well-structured, functional, and relatively complete knowledge expert system.[2] It is able to response according to career dynamic provided by consultants from different angles. According to the existing rules system, it also can identify the interviewer quickly and accurately whether visitors are confused, and can propose solutions with measures targeted. In addition, the agent should be able to explain the logical reasoning process and its diagnosis to visitors in some way so that visitors are easy to understand.[16] Mental frustration by choosing the career. Some students do suicide after choosing such careers. Less growth in career phase. The students are not able to survive in their job.[13] After choosing the career they feel humiliation.

It is very much important to consider the interest, talent, projected growth or sustainability in a particular area,[3] before choosing a career.[14] It is commonly seen that, many of the students have their poor academic record just

because of choosing their career without considering their own capabilities and it will cause waste of time and the money, so it is important to choose the right career in the first place[20]. It is also observed that there is an impact of psychological parameters for choosing a right career option.[3]The psychometric parameters such as Team work, Learning ability, Hardworking nature, Discipline, Attitude, Intellectual property, reliability, decision making ability, self-respect, analytical skills, Interest, Hobbies etc.

II. LITERATURE SURVEY

Rough set theory can be regarded as a new mathematical tool for imperfect data analysis. [4] The theory has found applications in many domains, such as decision support, engineering, environment, banking, medicine and others. Rough set philosophy is founded on the assumption that with every object of the universe of discourse some information (data, knowledge) is associated [5]. Objects characterized by the same information are indiscernible (similar) in view of the available information about them. The in-discernibility relation generated in this way is the mathematical basis of rough set theory.[19] Any set of all indiscernible (similar) objects is called an elementary set, and forms a basic granule (atom) of knowledge about the universe. Any union of some elementary sets is referred to as a crisp [6] set – otherwise the set is rough (imprecise, vague). Each rough set has boundary line cases, i.e., objects which cannot be with certainty classified, by employing the available knowledge, as members of the set or its complement.[18] Obviously rough sets, in contrast to precise sets, cannot be information about their characterized in terms of elements.[7] With any rough set a pair of precise sets, called the lower and the upper approximation of the rough set, is associated. The lower approximation consists of all objects which surely belong to the set and the upper approximation contains all objects which possibly belong to the set.[17] The difference between the upper and the lower approximation constitutes the boundary region of the rough set. Approximations are fundamental concepts of rough set theory[12].

The main concepts related to Rough Set Theory are presented as the following.

A. SET:

The A set of objects that possess similar characteristics it is a fundamental part of mathematics. All the mathematical objects, such as relations, functions and numbers can be considered as a set. The various components of a set are known as elements, and relationship between an element and a set is called of a pertinence relation [11].

B. Rough Set:

Rough Set Theory [12] is similar to Fuzzy Set Theory, however the uncertain and imprecision in this approach is expressed by a boundary region of a set, and not by a partial membership as in Fuzzy Set Theory. Rough Set concept can be defined quite generally by means of interior and closure topological operations know approximations (Pawlak, 1982).

C. Information Sysytem:

Information system is a table consist of information for formulating the rules.[1] It consist of data table I=(U,A), where A is an attribute and U be the finite set of objects.

D. Indiscernibility relation:

The relation who gives different prediction for same number of attributes or objects. These are fundamentals of rough set theory. In psychological prediction it is obvious to have such relation as two people can give choose different career though having same characteristics.[1] Any subset B of A determines a binary relation on U which will be called indiscernibility relation.

E. Decision Table:

Decision table is having attribute D; which shows the decision of a particular person based on different characteristics.

F. Approximation Relation :

Two approximation relations are defined in the RST, they are lower approximation and upper approximation. Lower approximation means the set having particular decision Yes or No for different attributes. Upper approximation is the set which is the combination lower set as well as the attributes which does not show confirmed decision on the attributes.

G. Boundary region:

The difference between the upper and lower approximation is called boundary region. Boundary region sets the line of decision for a particular decision.

H. Accuracy:

Accuracy of a particular decision is defined by the ratio of Lowe approximation to the lower multiplied by hundred. We can find the accuracy of a particular decision. To find the accuracy of whole system, We need to do the average of all accuracy for decisions present in the Information Table.

I. Core:

Core is found out by intersection of two reduct classes. It can be the combination of two or more attributes

III. PROPOSED METHODOLOGY

A. RST Algorithm:

Following steps need to follow while generating the rules from the corpus of data. These are:

Step 1	:	Normalize the Dataset.									
Step 2	:	Find out Indiscernibility relation.									
Step 3	:	Find out Discernibility relation.									
Step 4	:	Define Approximation of set.									
Step 5	:	Find out Accuracy.									
Step 6	:	Reduction of Attribute and finding out									
equivalence class		-									

Step 7 : Defining the core of a set by intersecting two equivalence classes.

Step 8 : Rules are generated and decision is made based on rules.

B. Advantages of RST

- 1. It is easy to understand.
- 2. Identifies relations which cannot be found using statistical method.
- 3. Generates set of decision rules from data.
- 4. Easily find out hidden patterns in data.

IV. EXPERIMENTATION ENVIRONMENT

The proposed algorithm is implemented on Rose 2 tool. This tool is specially meant for statistical and mathematical calculation. Tool is run on Windows platform.



tudent No	Strenghts	Reliabilty	cision Mak		Discipline				eadership	Feam Worl	Verbal Co	Persuading	Career	Career 2
1	50	46	78	47	25	55	46	78	55	62	36	49	Engineer	Manager
2	30	48	45	25	63	23	72	45	23	29	28	29	Engineer	Accountant
3	48	24	18	48	66	62	12	18	62	41	49	43	Barrister	Engineer
4	88	54	72	37	85	48	26	72	48	61	37	35	Manager	Engineer
5	78	45	46	16	55	47	47	46	47	16	47	31	Engineer	Socila Worker
6	48	45	17	58	25	27	86	17	27	18	34	38	Accountai	Social Worker
7	39	78	47	72	63	26	24	47	26	19	78	54	Barrister	Engineer
8	22	15	13	18	78	33	36	13	33	63	19	24	Engineer	Manger
9	36	47	46	18	45	36	78	46	36	28	27	27	Engineer	Accountant
10	26	13	15	16	65	38	18	15	38	27	36	26	Barrister	Engineer
11	42	76	28	78	35	36	24	28	36	36	27	33	Engineer	Manager
12	52	84	16	18	37	34	36	16	34	33	16	29	Engineer	Journalist
13	10	16	73	74	41	31	75	73	31	38	85	84	Engineer	Manager
14	12	76	47	36	15	32	16	47	32	48	78	88	Proffesor	Barrister
15	15	79	57	18	19	36	75	57	36	46	16	29	Engineer	Socila Worker
16	19	19	69	46	53	38	27	69	38	69	34	33	Manager	Barrister
17	82	28	49	72	44	37	46	49	37	43	36	36	Engineer	Manager
18	25	31	18	12	25	39	74	18	39	19	47	37	Engineer	Journalist
19	25	78	46	26	33	34	36	46	34	65	19	49	Engineer	Manager
20	26	46	85	47	31	31	16	85	31	56	27	48	Manager	Barrister
21	30	12	17	86	16	32	27	17	32	53	13	42	Engineer	Manager
22	40	73	36	24	43	22	16	36	22	54	47	43	Engineer	Proffesor
23	45	48	28	36	35	29	42	28	29	28	79	16	Engineer	Journalist
24	54	56	49	78	32	30	75	49	30	29	16	15	Engineer	Social Worker
25	18	78	37	18	23	49	14	37	49	78	74	12	Engineer	Manager

Table 1: Sample database and Career

V. RESULTS AND DISCUSSIONS

We took the dataset of 90 students. Each student answered 10 questions on each attribute. Attributes are Strength, Reliability, Decision Making, Attitude, Discipline, Social Ability, Leadership, Team work, etc. All 12 attributes were considered while forming the dataset. Each answer to the question is measured in scale of 1 to 10. We considered two careers for each student, i.e. career choice 1 and career choice 2.

Table 1 shows the sample dataset. In this sample dataset all 12 attributes and its value for each students is specified. The value we got by doing the summation of each answers to the question for a particular attribute.

In table 2, we normalized the dataset by removing the equivalence class and found 40 entries were same. So our dataset reduced to 50 entries. In the interval scale of 1-25 as A, 26-50 as B, 51-75 as C, 76-100 as D, we normalized the data.

From table 2, for each career; lower and upper approximations are:

Lower Approximation for Professor in Career choice 1 :{ 42}

Upper Approximation for Professor in Career choice 1: {42, 49, 50}.

Lower approximation for Engineer in Career choice 1: {3-14, 15-17, 20, 22-24, 26-30, 33-35, 37, 39-41, 44-48} Upper approximation for Engineer in Career choice 1: {3-14, 15-17, 20, 22-24, 26-30,33- 35, 37, 39-41, 44-48, 49, 50}

Lower Approximation for Manager in Career choice 2 :{ 4, 6, 10, 12, 14, 16, 17, 20, 23, 26, 30, 35, 37, 39, 40, 45, 46, 48}

Upper Approximation for Manager in Career choice 2: {4, 6, 10, 12, 14, 16, 17, 20, 23, 26, 30, 35, 37, 39, 40, 45, 46, 48, 49, and 50}

From the table 2, we found out the accuracy of each career. Those Careers who are not having indiscernibility relation have 100% accuracy in the rules. But those who are having indiscernibility relation having less accuracy as for career professor the accuracy is 33%, Engineer 94%, Govt.jobs 66%, Manager 90.04%. This accuracy is for generating the rules for prediction.

Equivalence class:

 $\{1,10,25,35\},\{21,33,42,48,56\},\{15,59,72\},\{65,82\},\{45, 61, 76\}$ $\{30, 41, 58, 85\}, \{26, 49, 83,88\}, \{37,77,86\}$ $\{5,10,17,22,36,52,70,84\},\{2,12,22\},\{4,19,62,79,90\},\{8,18, 31,66,50,71\},$ $\{9,63,75\}.$ Above mentioned are all equivalence class.

Vol.-3(5), PP(181-186) May 2015, E-ISSN: 2347-2693

		D. D. L. L. D.		A			Norman					No	C	C
	D		cision Mak	Attitude B							Verbal Cor B	B		Career 2
1		С	C		D	В	В	C	В	C			Manager	Engineer
2	D	A	A	D	С	В	В	A	В	С	A	B	Govt. jobs	
3	D	В	В	A	В	В	С	В	В	В	В	В		Journalist
4	D	В	В	С	С	В	В	В	В	С	В	B		Manager
5	D	В	В	A	С	В	В	В	В	A	В	B		Socila Wor
6	D	В	В	С	В	B	B	В	В	В	B	B	Engineer	Manager
7	С	C	В	D	В	В	С	В	В	В	A	A		Social Wor
8	C	A	A	A	В	В	D	A	В	A	В	C		Proffesor
9	С	A	В	A	В	D	D	В	D	В	В	С		Proffesor
10	С	D	D	D	A	В	В	D	В	A	С	С		Manager
11	C	D	A	A	В	В	В	A	В	В	A	В	Engineer	Journalist
12	С	D	В	В	A	С	В	В	С	В	B	B	Engineer	Manager
13	С	В	В	В	A	A	В	В	A	A	В	С		Journalist
14	С	В	В	В	B	В	A	В	В	C	A	B		Manager
15	B	A	A	A	C	В	A	A	В	B	В	B	Govt. jobs	
16	B	A	A	D	A	В	В	A	В	С	A	B		Manager
17	B	A	A	С	A	В	В	A	В	В	C	A		Manager
18	В	A	В	A	В	B	В	В	B	A	A	D	U	Social Wor
19	B	A	A	В	C	C	A	A	C	B	B	B	Govt. jobs	
20	В	A	С	В	С	В	D	С	В	D	С	A		Manager
21	В	A	В	С	В	В	A	В	В	С	В	B	Manger	Govt. jobs
22	В	D	A	A	В	В	С	A	В	В	D	В	Engineer	Sales Pers
23	В	D	В	D	В	В	A	В	В	В	В	B		Manager
24	В	D	D	С	В	A	В	D	A	В	С	С		Proffesor
25	В	D	В	С	С	В	A	В	В	A	D	С	Govt. jobs	
26	В	В	D	В	A	С	В	D	С	С	В	B	Engineer	Manager
27	В	В	В	A	С	A	С	В	A	В	В	B	Engineer	Accountar
28	В	В	В	В	В	В	В	В	В	В	D	A	Engineer	Journalist
29	B	B	B	A	B	В	D	B	В	В	B	B		Accountar
30	B	В	C	B	C	A	A	C	A	В	В	C		Manager
31	B	В	A	C	A	B	D	A	B	A	В	B		Social Wor
32	B	B	D	B	B	В	A	D		C	B	B	Manager	Govt. jobs
33	B	c	B	B	C	B	D	B	B	D	C	A	Engineer	Accountar
34	B	C	B	A	B	A B	A	B	A B	C	B	B		Proffesor
35 36	A	A		A B	C B	B	A	D		B	A	A B	Engineer	Manger
36	A	A	A	A	D	B	A B	A	B	B C	A	A	Govt. jobs	
37	A	A	A C	B	C	B		A C		C C				Manger Court jobs
	A		C	С	B		B C	C C	B	В	B	B D		Govt. jobs
39 40	A	A C	B	A	B	B D	A	B	D	B	D	B		Manager Manager
40	A		D	B	C B	B	A C	D					Engineer	Manager
41 42		C C	A	B	C C		B		B	A	A B	A C		Proffesor
42	A	B		В	B	A	B	A		A	B	C	Engineer	Proffesor
43	A	B	A	A	A	B	C B	A	B	A	B	В	Engineer	Govt. jobs
44	A	B	B	B	B	D	B		D		A	B		Journalist
45	A	D	D	D	B	B	A	B	B	A B	A D	B		Manager
46	A	D	C D	A	A	B	A C	C	B	B	A	B		Manager Socila Woi
													Engineer	
48 49	A	D	B	A	A	B	A	B	B	D D	C C	A		Manager Court jobs
	A A		B	A			A			D	C C	A	Proffesor	Govt. jobs
50	A	D	В	A	A	В	A	В	В	U	L	A	Engineer	wanager

Table 2: Normalized Dataset

The rules for Career choice 1 and Career choice 2 are17 and 29 respectively. These rules are generated in Rose tool and formulation is done manually. In 17 rules 17th rule is approximation rule and in 29 rules 29th rule is approximation rule. Approximation rule is for indiscernibility relation in a dataset.

A. Rules for Career choice 1 are mentioned below:

Rule 1: If A5=2 and A10=2, then Decision = 1. Rule 2: if A7=2 and A10=1, then Decision =1.



© 2015, IJCSE All Rights Reserved

Rule 3: if A2=1 and A5=4, then Decision =1. Rule 4: if A4=2 and A10=3, then Decision =1. Rule 5: if A1=2 and A12=1, then Decision =1. Rule 6: if A2=2 and A4=1, then Decision =1. Rule 7: if A5=1 and A9=3, then Decision =1. Rule 8: if A7=3 then Decision =1. Rule 9: if A6=1 then Decision =1. Rule 10: if A2=1 and A5=1, then Decision =1. Rule 11: if A1=3 then Decision =1. Rule 12: if A3=3, A9=2and A11=2 then Decision =2.

Vol.-3(5), PP(181-186) May 2015, E-ISSN: 2347-2693

Rule 13: if A5=2, A6=2, A10=3 and A11=2 then Decision =2.

Rule 14: if A2=1, A5=3 and A8=1 then Decision =3.

Rule 15: if A11=4 and A12=3, then Decision =3.

Rule 16: if A4=3 and A7=4, then Decision =7.

Rule 17: if A5=1 and A10=4, then Decision =1 or Decision =6.

Rule 17 is approximation rule.

B. Rules for Career choice 2 are mentioned below:

Rule 1: If A2=1, A5=3 and A8=1, then Decision=1.

Rule 2: If A4=3, A5=4 then Decision =1.

Rule 3: If A11=4, A12=3 then Decision=1.

Rule 4: If A1=4, A4=3, A8=2 then Decision =2.

Rule 5: If A2=1, A12=1 then Decision =2.

Rule 6: If A8=2, A11=1, A12=2 then Decision =2.

Rule 7: If A5=2, A7=3, A10=2 then Decision =2.

Rule 8: If A5=1, A7=2, A12=2 then Decision =2.

Rule 9: If A10=2, A12=4 then Decision =2.

- Rule 10: If A3=3, A12=3 then Decision =2.
- Rule 11: If A3=4, A4=4 then Decision =2.
- Rule 12: If A2=1, A10=3, A11=2 then Decision =3.

Rule 13: If A4=2, A5=2, A11=2 then Decision =3.

Rule 14: If A6=1, A7=2, A8=2 then Decision =4. Rule 15: If A2=2, A7=3, A9=2 then Decision =4.

Rule 16: If A7=2, A1=3, A2=2 then Decision =4.

Rule 17: If A1=3, A2=4, A11=1 then Decision =4.

Rule 18: If A7=3, A10=2, A11=1 then Decision =5.

Rule 19: If A1=2, A12=4 then Decision =5.

- Rule 20: If A1=4, A2=2, A10=3 then Decision =5.
- Rule 21: If A4=3, A7=4 then Decision =5.
- Rule 22: If A7=4, A12=3 then Decision =6.
- Rule 23: If A2=3, A9=1 then Decision =6.
- Rule 24: If A6=1, A8=4 then Decision =6.
- Rule 25: If A2=3, A3=4 then Decision =6.
- Rule 26: If A1=2, A7=4 and A8=2 then Decision =7.
- Rule 27: If A7=3, A9=1 then Decision =7.
- Rule 28: If A2=4, A7=3 and A11=4 then Decision =8.

Rule 29: If A5=1, A10=4 then Decision =2 or Decision =3.

Rule 29 is an approximation rule.

VI. CONCLUSION

This paper provides Rough set theory as a solution for data mining problems, which we face in our day to day life. For career dataset, we have found out 17 rules for career choice 1 and 29 rules for career choice 2. This means that those students, which come under these rule, will choose the recommended career. The accuracy of all other career is 1.0 except for career professor, engineer, Govt, jobs and Manager the accuracy is 0.33, 0.94, 0.66and 0.90 respectively.

REFERENCES

- [1] Xinjuan Zhang, Jing Zhang; Research on the Graduate Career Decision-making Based on Rough Set Theory and Decision Tree Method.
- [2] Niharika Upadhyay and Pragati Jain; APPLYING ROUGH SET THEORY IN FEEDBACK ANALYSIS ;eISSN 2278-1145.
- [3] Xinjuan Zhang, Jing Zhang; Research on the Graduate Career Decision-making Based on Rough SetTheory and Decision Tree Method.
- [4] Aboul Ella Hassanien, Jafar H. AliRough Set Approach for Generation of Classification Rules of Breast Cancer Data.
- [5] Hameed Al Qaheri, Ajith Abraham; A Generic Scheme for Generating PredictionRules Using Rough Set.
- [6] Zdzislaw Pawlak; Rough set theoryand its applications.
- [7] Kein E. Voges; Research Techniques Derived From Rough Sets Theory:Rough Classification and Rough Clustering
- [8] Sivia Rissino, Germano Lambert Torres; Rough Set Theory – Fundamental Concepts, Principals, Data Extraction, and Applications
- [9] Chien Chung Chan, Dept. Of mathematics and computer science; Rough Sets Theory and Its Applications
- [10] Zdzisław Pawlak; Rough Sets; Institute of Theoretical and Applied Informatics, Polish Academy of Sciences, ul. Bałtycka 5, 44 100 Gliwice, Poland.
- [11] Lokesh S. katore, Dr.(Prof). Umale Sir; Novel approach to recommend career option to engineering students using improved analytics.
- [12] Xuecai Bao, Tangsheng Wang, Jianfeng Hu; Analysis of Individual Career Decision-making based on partial least-squares regression model" 2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery.
- [13] Lokesh S. katore, Dr. J.S. Umale; Comparative Study of Recommendation Algorithms and Systems using WEKA; ISBN: 973-93-80884-86-4
- [14] Sudheep Elayidom, Dr. Sumam Mary Idikkula, Joseph Alexander, Anurag Ojha, Applying Data mining techniques for Placement chance prediction" 2009 International Conference on Advances in Computing, Control, and Telecommunication Technologies
- [15] Ying Cao, Lei Zhang; Research about the College Students Career Counseling Expert System based on Agent"
- [16] Jyi-Shane Liu and Ke-Chih Ning; Applying Link Prediction to Ranking Candidates for High-Level Government Post" 2011 International Conference on Advances in Social Networks Analysis and Mining
- [17] Bert w. westbrook, Eleanor e. Sanford, Predictive and Construct Validity of Six Experimental Measures of



Career Maturity" Journal of Vocational Behavior 27, 338-355 (198.5)

- [18] Eliza Byington *, Will Felps, Why do IQ scores predict job performance?An alternative, sociological explanation" Research in Organizational Behavior 30 (2010) 175–202.
- [19] Mercedes Inda 1, Carmen Rodríguez , José Vicente Peña, Gender differences in applying social cognitive career theory in engineering students" Journal of Vocational Behavior 83 (2013) 346–355
- [20] Thomas Zellweger , Philipp Sieger, Frank Halter; Should I stay or should I go? Career choice intentions of students with family business background" Journal of Business Venturing 26 (2011) 521–536

AUTHORS PROFILE

Mr.Lokesh Katore Qualification: ME (Comp) Persuing

Ms.Bhakti Ratnaparkhi Qualification: ME (Comp) Persuing



Dr.J.S.Umale Designation : Professor (Dean Academics) Qualification : Ph.D. (Computer)

Mrs.N.Upadhyay Qualification: M.Sc. (Maths), persuing P.Hd.





