International Journal on Future Revolution in Computer Science & Communication Engineering
Volume: 3 Issue:11
ISSN: 2454-4248
142 – 146

Rule Based Systems to Judge Child's Basic Mathematical Abilities

Neelam Naik
K.G. Shah Dept. of Computer Applications
Dr. BMN College, Matunga
Mumbai, India
neelamnaik08@gmail.com

Seema Purohit
Dept. of Mathematics
Kirti College, Dadar
Mumbai, India
supurohit@gmail.com

***Abstract* —** India's nationwide survey, Annual Status of Education Report (ASER), provides basic and critical information about rural Indian children's foundational reading skills and basic mathematical abilities. The results of such a large survey has a potential to provide a tool to build quality educational practices and policies. The acquired educational skills of child are the results of child's social, economic and educational background which can be described by facilities available at household, parental and school level. The present study uses rough set theory which provides efficient algorithms for finding hidden patterns in data using minimal sets of attributes. The rough set based rule system is developed and tested for its accuracy by using ASER data to predict child's mathematics learning ability.

Keywords- *Rough set theory, J48, ASER, primary education*

*****

I. INTRODUCTION

The availability of quality primary education is the fundamental step towards building strong foundation for physical, social and intellectual development of children. The reduced drop-out rate from schools can be achieved by teaching reading, writing and other basic learning skills during primary education. Basic mathematical abilities during the early age of schooling are considered as indicators of quality education. The methods such as counting, number knowledge and arithmetic operations are considered for evaluation purpose. [13]

Foundation and Expanse of Mathematics and its application to diverse disciplines makes it a queen of sciences, Mathematics helps to inculcate the habit of reasoning and problem solving from their childhood. The early exposure to mathematics develops an analytical mind of child and helps him in better organization of ideas and accurate expression of thoughts. Learning mathematics is very much important for child's intellectual development. [4]

The children's early encounter with mathematics develops curiosity, imagination, flexibility, inventiveness, persistence and confidence in their abilities. Besides, it gives a strong foundation for logic building. These qualities contribute to their future success in life. Even though education makes individual self-reliant and wise, study of mathematics has potential to sharpen the wit and problem-solving abilities. Thus, mathematics skills can be considered as tool to judge overall progress of child in education. [16].

The non-governmental organization Pratham's Annual Status of Education Report (ASER) is a nationwide survey of reading and mathematical achievement of children from rural India. It provides basic and critical information about rural Indian children's (age group 5-16) foundational reading skills and basic mathematics ability. The survey is conducted each year in the middle of the academic year and findings are made public. ASER survey provides genuine information that mainly contribute to forming effective base for governmental educational practices and policies. [18]

Rule based system has played an important role in machine learning and is known for inducing interpretable and comprehensible classifiers. The main rule based techniques are decision trees based rule learner, sequential covering, associative rules and rough set based methods. [7].

A. *Rough Set theory for generating decision rules*

The hidden patterns in data can be discovered by using the concept of Rough set theory. It is used for feature selection, attribute reduction, decision rule generation, and pattern extraction. [7] [17]

The data analysis based on rough set theory starts from data table. It is also called a decision table. The columns of decision table are attributes in dataset. The rows in data table are objects of interest. Entries in data table are attribute values. Each row contains condition attributes and decision attributes. A single row represents a decision rule [23].

Consider a data table denoted by $S = (U, C, D)$ where C and D are disjoint sets of condition and decision attributes respectively and U is a finite set of objects.

Each $x \in U$ determines a sequence $c_1(x),\ldots,c_n(x)$, $d_1(x),\ldots,d_n(x)$, where
$\{c_1(x),\ldots,c_n(x)\} = C$ and $\{d_1(x),\ldots,d_n(x)\} = D$.
The decision rule induced by x in S. It is denoted by
$c_1(x),\ldots,c_n(x) \rightarrow d_1(x),\ldots,d_n(x)$
That is $C \rightarrow D$

B. *About ASER data*

To understand the behavior of different factors and how they are related to inculcation of mathematical ability, available data from ASER survey is used, as it gives almost all the information about children.

ASER conducts the largest household survey of children in India. It is carried out by around 25,000 volunteers. Each year, it surveys around seven lakhs of children in around fifteen thousand villages. The main aim of the survey is to evaluate learning levels among children. The advantage of choosing household as a basic sample survey unit is that children are tested at home rather than in school, allowing all children to be tested rather than just those in school, testing children in school might create a bias since teachers may encourage testing the brighter children in class and a household sample will generate an age distribution of children that can be crosschecked with other data sources like the census.

The dataset contains attributes such as total number of individuals in household, type of material house is made of, household has facilities (electricity connection, toilet,

 ISSN: 2454-4248
142 – 146

television, cable TV, mobile phone, vehicle, newspapers, other reading material), household has a person who has completed std. 12, household has a person who can use computer, parents age and education, government or private school, whether child takes paid tuition, reading level of child and arithmetic level of child.

The available ASER data is sparse and redundant data. It is also observed that there is vagueness in correlation between some of the survey parameters. Rough set theory performs intelligent data analysis for such data. Using an approximation concept, rough set theory is able to remove data redundancies and consequently generate decision rules. [15] Some of the social, economic and educational factors and child's mathematics learned level are used in the analysis.

C. *Problem Identification*

The acquired educational skills by a child dependent on his or hers social, economic and educational background. It includes facilities available at household, parental and school level. Most of the times child's mathematics achievements are highly correlated to individual and instructional factors and moderately correlated to classroom engagement or attentiveness and evaluation factors [2].

It is also observed that the major factors contributing to poor performance in Mathematics are motivation and attitude; along with the socio-economic factors such as education of parents and their economic status; and school based factors such as availability and use of teaching/learning facilities, school type and teacher characteristics [21].

The differences in mathematics achievement among children has highlighted the importance of classroom, teacher and school factors as well [19].

The aim of the present study is to develop rough set based rule system using ASER data and to show correlation between child's social, economic and educational background and mathematical learning ability.

II. LITERATUTRE REVIEW

Dr. Aditi Banerjee, in her paper [3] elaborates global practices regarding effect of practices used to judge children's language and mathematical skills. This study also recommends beneficial utilization of the inputs to promote better learning in students. [3] Ambrish Dongre et. al. estimates effect of sending children to private tutoring and its correlation with learning outcomes. This paper utilizes a large household survey conducted in rural India and finds that the effect of private tutoring is stronger for the students enrolled in government schools compared to the students enrolled in private schools, for children from economically disadvantaged background, and for children whose parents are relatively less educated. [1]

Academic performance is affected by a number of factors including admission grade, social economic status, school background and many more. Students' personal factors and school based factors are the focus of the study described in this paper [10]. The findings revealed that inadequate teaching and learning materials as well as lecture method of instruction are some of the factors which affect students' performance. [10] The factors contributing to the students' low, medium and high achieving performances in mathematics are found in this study. These are the school context, profiles of the students and teachers, descriptions of sample lessons taught including the insider perceptions of the teacher and students, and a discussion of key issues and challenges faced, as well as initiatives to promote the learning of mathematics in each case. [11]

Educators are facing major challenges in maintaining quality in teaching mathematics. The decision design is based on the knowledge of the factors affecting mathematics achievement in children. This study was conducted to identify the factors affecting the math achievement of students and it revealed that instructional strategies and methods, teacher competency in math education, and motivation or concentration were the three most influential factors that should be considered in the design decisions. [19]

Data mining and knowledge discovery is considered as the main module of development of advanced Decision Support Systems. In this paper, rough set theory is used to simplify decision making process. [12] The paper [22] gives basic ideas of rough set theory as a new approach to vague data analysis. This paper [9] uses rule induction algorithms LEM1 and LEM2.

III. EXPERIMENTAL SETUP AND IMPLEMENTATION

A. *About RSES*

Rough Set Exploration System (RSES) is an open source software system for data exploration, classification and knowledge discovery. Many RSES methods are based on rough set theory. RSES utilizes classification algorithms using elements of rough set theory, instance based learning, and artificial neural networks. The construction of classifier is carried out by several steps. First, the data for analysis has to be loaded or imported into RSES. Data is pre-processed before next step is carried out. RSES supports preprocessing methods which make it possible to manage missing parts in data, discretize numeric attributes, and create new attributes. By using classical rough set concepts such as reducts, dynamic reducts and positive region, dependencies in datasets are found out. Knowledge of reducts may lead to reduction of data by removing some of the redundant attributes [8].

B. *Implementaion*

Algorithms used for generating rules are J48 and LEM2. J48 is implementation of the decision tree C4.5 algorithm in WEKA [14]. The algorithm is based on a greedy technique to make decision trees for classification and practices reduced-error pruning. [5]

Rough Set Exploration System 2.2.2 is opened and new project is created into it. The new Table is created into it and it is loaded with educational dataset (Attribute Relation File Format). The split factor 0.6 is applied on this table. Thus, the training dataset is 60% and testing dataset is 40%. The concept of cut and discretization is used here to reduce dataset horizontally. The attributes having continuous values are partitioned into intervals. The training and testing datasets are discretized by using Global method where discretization score is calculated by considering all the objects in data table at every step of discretization.

The algorithm LEM2 (Learning from Examples Module – version 2) is a component of the data mining system LERS (Learning from Examples using Rough Sets) [6]. By selecting discretized training dataset, rules are generated using LEM2 algorithm. Figure 1 shows nodes to indicate data table, discretized data table and rules generated in RSES software.

The attribute math_code is classified in five classes. Class E indicates that child is not able to do any arithmetic, class D

 ISSN: 2454-4248
142 – 146

indicates that child can recognize numbers 1-9, class C indicates that child can recognize number 11-99, class B indicates that child can do two-digit subtraction and class A indicates that child can do division (3-by-1 form).

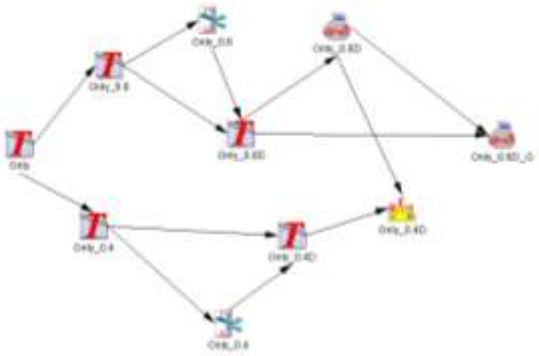

Figure 1: RSES nodes to indicate data table, discretized data table and rules generated

The testing dataset is classified using these rules. The rules generated using training dataset is generalized. The sample rules generated are shown in figure 2.

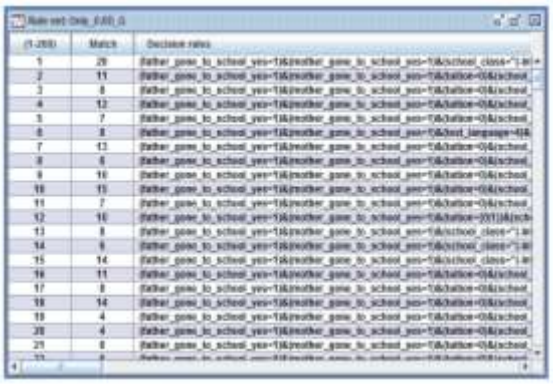

Figure 2: Sample of Decision Rules generated

IV. RESULTS

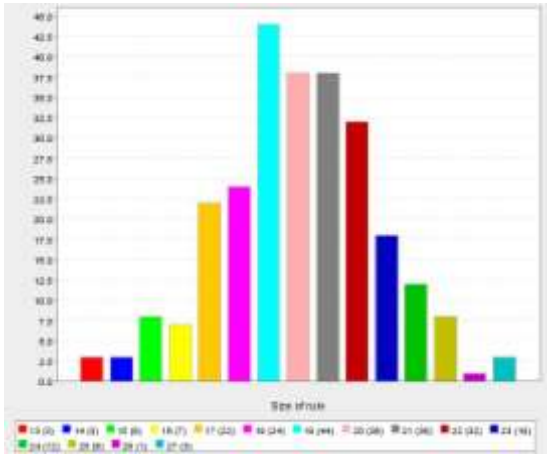

Figure 3: Graph showing sizes of rules generated for LEM2 algorithm

After the application of J48 algorithm for developing decision rules in WEKA, a tree with total 165 leaves, is generated. The application of LEM2 algorithm for deduction of rules in RSES resulted in generation of total 260 rules. The minimal support of rule is 1 while maximum support is 33. An average support is 4.7. The minimum length of rule is 13, maximum length is 27 while average length is 20 as shown in figure 3.

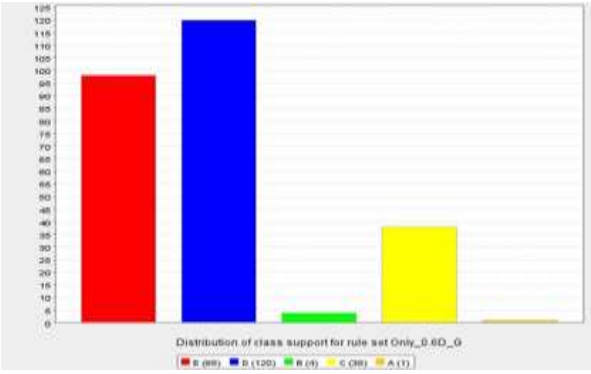

Figure 4: Distribution of class support for rule set for LEM2 algorithm

Class E is the output in 98 rules, class D is the output in 120 rules, class B is the output in 4 rules, class C is the output in 38 rules and class A is the output in 1 rule for the selected sample of dataset as shown in figure 4. The best 5 rules generated by application of J48 in WEKA and LEM2 in RSES are shown in table 1 and table 2 respectively.

Table 1: Rules generated by the application of J48 in WEKA

J48 algorithm	Output class	No. of records
If (school_class <= 1, test_language <= 8, child_age <= 5, tuition <= 0, hh_electricity_today > 0, hh_reading_material <= 0) then math_code = E	E	88
If (school_class <= 1, test_language <= 8, child_age > 5, hh_newspaper <= 0, school_govt > 0, hh_type > 1, hh_motor_vehicle <= 0, hh_type > 2, hh_computer_use <= 0, hh_mobile > 0) then math_code = D	D	74
If (school_class <= 1, test_language <= 8, child_age > 5, hh_newspaper <= 0, school_govt > 0, hh_type > 1, hh_motor_vehicle <= 0, hh_type <= 2, mother_class > 5, hh_toilet <= 0, hh_grad <= 0) then math_code =D	D	45
If (school_class <= 1, test_language <= 8, child_age > 5, hh_newspaper <= 0, school_govt > 0, hh_type > 1, hh_motor_vehicle <= 0, hh_type <= 2, mother_class <= 5, hh_reading_material <= 0, test_sample <= 3, hh_electricity_conn > 0, child_gender > 1) then math_code = E	E	31
If (school_class <= 1, test_language <= 8, child_age > 5, hh_newspaper <= 0, school_govt > 0, hh_type > 1, hh_motor_vehicle <= 0, hh_type <= 2, mother_class > 5, hh_toilet <= 0, hh_mobile > 0, father_class > 3, mother_age > 27, test_sample > 1) then	D	26

 ISSN: 2454-4248
142 – 146

math_code = D		

Table 2: Rules generated by the application of LEM2 in RSES

LEM2 algorithm	Output class	No. of records
If(father_gone_to_school_yes=1, mother_gone_to_scholl_yes = 1, tution=0, school_class<6, hh_electricity_conn=1, hh_newspaper=0, hh_copmuter_use=0|1, child_age>5, test_language=4|15, school_govt=1, school_private=0, hh_grad=0, hh_electricity_today=1,hh_tv=1|0,hh_mobile =1, hh_reading_material=0|1, hh_toilet=0|1,hh_cable_tv=0|1, hh_motor_vehicle=0, child_no=1|2|3|4, total_member=4, test_sample=1|4, father_class=8) then math_code=D	D	33
If(father_gone_to_school_yes=1, mother_gone_to_scholl_yes = 1, school_class <2, tution=0, hh_newspaper=0, test_language=4, school_govt=1, school_private=0, hh_computer_use=0, hh_electricity_today=1, hh_grad=0,hh_reading_material=0, hh_motor_vehicle=0, hh_toilet=1, hh_cable_tv=1, hh_mobile=1, hh_type=1, child_age<6) then math_code = E	E	20
If(father_gone_to_school_yes=1, mother_gone_to_scholl_yes = 1, school_class<2 , tution=0, hh_newspaper=0, test_language=4, school_govt=1, school_private=0, hh_computer_use=0, hh_electricity_today=1, hh_grad=0, hh_reading_material=0, hh_motor_vehicle=0, hh_toilet=1, hh_cable_tv=1,hh_mobile=0, hh_type=1, child_age<6)then math_code = E	E	19
If(father_gone_to_school_yes=1, mother_gone_to_scholl_yes = 1, school_class<2 tution=0, hh_newspaper=0, hh_electricity_connection=1, hh_reading_material=0, hh_computer_use=0, test_languag=4, school_govt=1, school_private=0, hh_elctricity_today=1, hh_motor_vehicle=0, hh_tv=1|0, hh_toilet=1|0, hh_type=3|2|1, hh_mobile=1, child_no=2|3, child_age<6)then math_code=E	E	16
If(father_gone_to_school_yes=1, mother_gone_to_scholl_yes=1, tution=0|1,school_class < 2, hh_electricity_conn=1, hh_newspaper=0, test_language=4|15, school_govt=1, school_private=0, child_age>=6, hh_computer_use=0|1, hh_grad=0, hh_cable_tv=1|0, hh_elctricity_today=1, child_gender=2|1, hh_reading_material=1,	D	16

total_member<=7) then math_code=D		

V. Research Findings

The classification rules generated are mainly for classes D and class E. An accuracy of generating rules in J48 algorithm is 46.4% while accuracy of rough set is 49%. J48 shows maximum support (88) for class E while LEM2 shows maximum support (33) for class D. The generated rules can be used for the prediction of child's mathematics learning level based on social, parental, school and economical parameter.

VI. Conclusion

It is found that conditional attributes, used in decision rules generation, are mainly to decide child's economical background. These attributes are: type of household and availability of facilities at household such as newspaper and vehicle. Parents' educational background also affects the level of mathematics learning of child. Gender of a child is not at all affecting factor to judge mathematics learning ability. Again, it is found that, when J48 algorithm is used, 45.07% government school going children studying in standard I are not able to identify even single digit numbers. And in case of LEM2 algorithm this percentage is 47.11%. Thus, from this study, we can conclude that, on an average 46.09% children of standard I, studying in government school, are not mathematics competent for their age group. Recommendations to improve mathematical ability are

1. Introduce learning by doing approach through playful activities and different physical models while teaching.
2. Design mathematical recreational activities.
3. Conduct activities to make them realize the importance of mathematics.
4. Increase concentration level.
5. Relate concept of mathematics to everyday life.

Acknowledgment

We are really thankful to NGO Pratham and ASER Delhi main center for providing their survey data. We are also thankful to Dr. Manjistha Banerji, Senior Researcher, Research & Statistics Unit, ASER Center, New Delhi for her sincere and timely help.

References

[1] Ambrish Dongre, Vibhu Tewary, "Impact of Private Tutoring on Learning Levels: Evidence from India", AI Working Paper Series, April 2014

[2] Andaya, Olive Joy F., "Factors That Affect Mathematics Achievements of Students of Philippine Normal University - Isabela Campus", Researchers World, 5 (4), 83, 2014

[3] Dr. Aditi Banerjee, "Improving Student's Learning with Correct Feedback: A Model Proposed for Classroom Utility", International Journal of Education and Psychological Research (IJEPR), Volume 3, Issue 4, December 2014

[4] Dr. Roohi Fatima, "Role of Mathematics in the Development of Society", National Meet on Celebration of National Year of Mathematics-2012; organized by NCERT, New Delhi, December 20 – 22, 2012

[5] Gaganjot Kaur, Amit Chhabra, "Improved J48 Classification Algorithm for the Prediction of Diabetes", International Journal of Computer Applications (0975 – 8887) Volume 98 – No.22, July 2014

[6] Grzymala, Busse J.W., "Managing uncertainty in machine learning from examples", Proceedings of 3rd Int. Symp. on Intelligent Systems, Wigry 1994
[7] Honghai FENG, Hongjiang FENG, Yanyan CHEN, and Junhui HUANG, "A Discernibility Degree and Rough Set Based Classification Rule Generation Algorithm (RGD)", Proceedings of the International MultiConference of Engineers and Computer Scientists 2014 Vol I, IMECS 2014, March 12 - 14, 2014, Hong Kong
[8] Jan G. Bazan, Marcin Szczuka, "The Rough Set Exploration System", Transactions on Rough Sets III, LNCS 3400, Springer-Verlag Berlin Heidelberg, pp. 37–56, 2005
[9] Jerzy W. Grzymala-Busse, "Rough Set Theory with Applications to Data Mining", ITTC Annual Report FY 2006
[10] Justice Enu, Osei K Agyman, Daniel Nkum, "Factors influencing students' mathematics performance in some selected colleges of education in Ghana", International Journal of Education Learning and Development", Vol.3, No.3, pp.68-74, April 2015
[11] Masitah Shahrill, Nor Azura Abdullah, Jamilah Yusof, "Factors Affecting Students' Performance in Mathematics: Case Studies of Three Primary Schools", The 7th ICMI-East Asia Regional Conference on Mathematics Education (EARCOME 7), "In Pursuit of Quality Mathematics Education for All", 11-15 May 2015., At Cebu City, Philippines
[12] Mohammad Hossein Fazel Zarandi, Abolfazl Kazemi, "Application of Rough Set Theory in Data Mining for Decision Support Systems (DSSs)", Journal of Industrial Engineering 1 (2008) 25 - 34
[13] Nancy C. Jordan, Joseph Glutting, and Chaitanya Ramineni, "The Importance of Number Sense to Mathematics Achievement in First and Third Grades", Learn Individ Differ. 2010 April 1; 20(2): 82–88. doi:10.1016/j.lindif.2009.07.004
[14] Patrick Ozer, "Data Mining Algorithms for Classification", Thesis submitted at Radboud University Nijmegen, January 2008
[15] Priyanka, Pritika Saini, Niti Sethi, "DECISION SUPPORT IN DATA MINING USING ROUGH SET THEORY", International Journal of Innovative Research in Technology & Science(IJIRTS), ISSN:2321-1156,VOLUME 1, NUMBER 2
[16] Report on Clements & Conference Working Group, 2004
[17] Salvatore Greco, Benedetto Matarazzo, Roman Slowinski, "Rough sets theory for multicriteria decision analysis", European Journal of Operational Research 129 (2001) 1-47
[18] Shaher Banu Vagh, "Validating the ASER Testing Tools: Comparisons with Reading Fluency Measures and the Read India Measures", ASER Study
[19] Stephen Lamb, Sue Fullarton, "Classroom and school factors affecting mathematics achievement: A comparative study of the US and Australia using TIMSS", Australian Council for Educational Research: ACEReSearch, TIMSS Australia Monograph Series, 4-2001
[20] Tuncay Saritas and Omur Akdemir, "Identifying Factors Affecting the Mathematics Achievement of Students for Better Instructional Design", International Journal of Instructional Technology and Distance Learning, Vol. 6. No. 12, ISSN 1550-6908, December 2009
[21] Zachariah Kariuki Mbugua Komen Kibet George Mungiria Muthaa George Reche Nkonke, "Factors Contributing To Students' Poor Performance in Mathematics at Kenya Certificate of Secondary Education in Kenya: A Case of Baringo County, Kenya", American International Journal of Contemporary Research, Vol. 2 No. 6; June 2012
[22] Zdzisław Pawlak, " ROUGH SETS AND DATA MINING", Institute of Theoretical and Applied Informatics, Polish Academy of Sciences, ul. Baltycka 5, 44 100 Gliwice, Poland
[23] Zdzislaw Pawlak, "Rough set theory and its applications", Journal of Telecommunications and Information Technology, March 2002.