Rough Set Analysis for Sudan School Certificate

Moawia Elfaki Yahia[1] and Nasrin Dalil Ali Arabi[2]

[1] College of Computer Science and IT,
King Faisal University,
Alahsa 31982, Saudi Arabia
`meyahia@hotmail.com`
[2] Faculty of Mathematical Sciences,
University of Khartoum,
Khartoum 11115, Sudan

Abstract. This paper has applied the decision analysis of the rough set theory to see the affect of each factor in the result of the Sudan School Certificate Examination. Through the analysis, we have compared between the percentage of success in male and female, and we have tried to answer which type of schools has the biggest number of examined students and which state in Sudan has the highest concentration of examined students. To answer these questions and to achieve good results, the paper has applied the concepts of reduction and rule generation.

Keywords: Rough Sets, Decision Analysis, Attribute Reduction, Rule Generation, School Certificate.

1 Introduction

Rough Set Theory is a mathematical tool of fundamental importance to artificial intelligence and cognitive science, and it is highly applicable to the tasks of machine learning and decision analysis [5]. Rough set are useful in the analysis of decisions in which there are inconsistencies [1]. To cope with these inconsistencies, lower and upper approximations of decision classes are defined. Rough set theory can deal with uncertainty and incompleteness in data analysis [7], [6]. It deems knowledge as a kind of discriminability. The attribute reduction algorithm removes redundant information or features and selects a feature subset that has the same discernibility as the original set of features [3], [4], [8], [9]. Rough Sets are efficient and useful tools in the field of knowledge discovery to generate discriminant and characteristic rules [2].

The Sudan School Certificate Examination passed a number of changes to reach the present level. It is basic academic yardstick for admission to higher education institutions in Sudan. Thus, the other certificates are valued against them. In the last view years, there occurred a severe competition amongst students to acquire the higher results and achieve university entrance and the best faculties and specializations [13]. Therefore, it is found that the details of the results of the school certificate examination are fertile ground for the application of the concepts of the theory of Rough Sets on them, and from them to obtain results which assist to know the factors which affect the success of the students. This in turn assists in setting future plans and policies for the raising of the students levels.

In this study we want to know the yardstick for success in the Sudan School Certificate Examination. We have applied the Rough Set data analysis to see when the final result of the student is become pass or fail according to his/her degree in each subject, which school type has the largest number of students to examined from it and which state has the largest number of students to examined from it [10].

2   The Sudan School Certificate

2.1   Historical Background

The Sudan school certificate had passed over many stages until it reach what it is now. Until 1953, secondary school students set for Cambridge Overseas certificate. In 1954, the Sudan Examination Council was established as an independent unit to cooperate with Cambridge University to conduct the examination of the school certificate [12]. It was conditioned from the beginning, that the Sudan Certificate shall be exactly equivalent to Cambridge University Certificate. Thus, the Sudan School Certificate become as provided by Cambridge University Regulations a requirement for enrollment and study for the external degree of the University of London. It also becomes the yardstick for assessment of all the other Arabic and foreign certificates for university entrance in Sudan [12].

2.2   The Structure of General Education in Sudan

The education is based on two stages:
1. The Basic Education stage, which includes the pre-school education (such as the *Khalwas* and Kindergartens). The entrance age is 4 years. Then there is the basic education, which is 8 years and the entrance age is 6 years [13].
2. The Secondary stage, which is includes two types of education:
   a. The New Secondary School, which is 3 years of study and which qualify the student to sit for the Sudan Secondary School Certificate and the competition to study in universities and higher institutes - in both their academic and technical streams - as some development had occurred in the level, which avail wider opportunities for those sitting for the new academic Secondary School Certificate Examination [13].
   b. The Technical Education, in which period is three years. It qualifies the student to sit for the secondary school certificate to join Universities and Higher Institutes. This is represented by four types: Commercial, Industrial, Agricultural and Women Studies [13].

2.3   The Basic of Success in the School Certificate

In the past, the result of the School Certificate Examinations was based on the competitive total of five subjects only to qualify for the entrance to the various university faculties. That was a system which is based on a narrow concept of the secondary school syllabus, which forms the base of the higher education. This concept may result in the neglect of some students of some subjects which are basic (such as Arabic language, Islamic studies and other). This reduces the impact of these evaluations,

resulting in graduating students from secondary and post-secondary levels having limited knowledge in these subjects [12].

Therefore, in the last decade of the past century this system was changed to that of the percentage ratio which the student obtains seven subjects (4 of which are basic: Arabic language, the religious knowledge, English language and elementary mathematics, plus three specialized subjects) [12].

The system of percentage ratio for all the subjects of the syllabus (which also requires success in the above four basic subject), guarantees for the student acquiring the general culture. In addition to achieving objectives of the educational revolution such as *originalization*, keeping away from Westernization and making the Arabic language the origin of education with caring for a foreign language [12].

3   Methodology and Materials

3.1   Data Description

We have selected 1000 objects of actual student's data and it is obtained from random sample of data of Sudan School Certificate Examination in scientific track related to the student. The following is the description of the information system's attributes:

- The sex is numeric, contains the sex of the student and has values of 1 for female or 2 for male.
- Arabic, English, Religion, Mathematics, Physics, Chemistry and Biology are numeric, contain the student's degrees in these subjects and have values between 1 and 100.
- The State code is numeric, contains the code of the state which a student examined from and has values between 1 and 16.
- The School type is numeric, contains the type of school which a student studied in and then examined from and has values between 1 and 5.
- The Result is numeric, shows the final result of a student and has values of 1 for success or 2 for fail.

3.2   Decision Analysis

The Sudan School Certificate Examination data set contains 10 condition attributes and one decision attribute. The condition attributes are: sex, Arabic, English, Islamic, Mathematics, Physics, Chemistry, Biology, State code and the School type. The decision attribute is the final result of the student.

The research tool is *ROSETTA* system, a toolkit developed for analysis of data within the framework of rough set theory [11]. Our data is stored in Access database. Firstly, we have imported the data to ROSETTA. This is done by using the *ODBC* (Object Data Base Connectivity).

Next, we have applied the Reduction concept specially the Manual Reducer. In the Manual Reducer we have selected:

- Arabic, English, Religion, Mathematics, Physics, Chemistry and Biology attributes to find when the final result of a student be passing or fail according to his degree in each subject.
- Sex attribute to find the percentage of success and fail in male and female.
- School type attribute to see which type of schools have the largest number of examined students.
- State code attribute to see which state have the highest concentration of examined students and which one have the lowest.

After that, for each reducer we have generated a Decision Rule to reach the expected result. Moreover, the decision rules give us a good answer about what we are asking for.

4 The Results

In the following, we have presented some decision rules generated from data of Sudan School Certificate. The data set consists of 1000 cases. We have applied the Manual reducer for each attribute. Also we have presented the values of LHS Support, RHS Support, RHS Accuracy and RHS Coverage according to each rule.

Table 1. Rules generated from the Arabic language attribute

Rule	LHS Support	RHS Support	RHS Accuracy	RHS Coverage
ARABIC(1) => RSL(1) OR RSL(2)	975	903, 72	0.926154, 0.073846	1.0, 0.742268
ARABIC(2) => RSL(2)	25	25	1.0	0.257732

From Table 1, we found that:

- Some of the students who pass the Arabic language subject are passing the final result and some of them are failing the final result.
- All of the students who fail the Arabic language subject are failing the final result.

Table 2. Rules generated from the English language attribute

Rule	LHS Support	RHS Support	RHS Accuracy	RHS Coverage
ENGLISH(1) => RSL(1) OR RSL(2)	958	903, 55	0.942589, 0.057411	1.0, 0.56701
ENGLISH(2) => RSL(2)	42	42	1.0	0.43299

From Table 2, we found:
- Some of the students who pass the English language subject are passing the final result and some of them are failing the final result.
- All of the students who fail the English language subject are failing the final result.

Table 3. Rules generated from the Mathematics attribute

Rule	LHS Support	RHS Support	RHS Accuracy	RHS Coverage
MATH(1) => RSL(1) OR RSL(2)	934	903, 31	0.966809, 0.033191	1.0, 0.319588
MATH(2) => RSL(2)	66	66	1.0	0.680412

From table 3, we found:
- Some of the students who pass the Mathematic subject are passing the final result and some of them are failing the final result.
- All of the students who fail the Mathematic subject are failing the final result.

Table 4. Rules generated from the Sex attribute

Rule	LHS Support	RHS Support	RHS Accuracy	RHS Coverage
SEX(2) => RSL(1) OR RSL(2)	319	276, 43	0.865204, 0.134796	0.305648, 0.443299
SEX(1) => RSL(1) OR RSL(2)	681	627, 54	0.920705, 0.079295	0.694352, 0.556701

From table 4, we found:
- The students are passing or failing in the final result for both male and female.

Table 5. Rules generated from the Physics attribute

Rule	LHS Support	RHS Support	RHS Accuracy	RHS Coverage
PHYSICS(1) => RSL(1) OR RSL(2)	832	802, 30	0.963942, 0.036058	0.888151, 0.309278
PHYSICS(2) => RSL(2) OR RSL(1)	168	67, 101	0.39881, 0.60119	0.690722, 0.111849

From table 5, we found:
- Some of the students who pass the Physics subject are passing the final result and some of them are failing the final result.
- Some of the students who fail the Physics subject are passing the final result and some of them are failing the final result.

Table 6. Rules generated from the School type attribute

Rule		LHS Support	RHS Support	RHS Accuracy	RHS Coverage
SCTYPE(3) RSL(1) OR RSL(2)	=>	157	137, 20	0.872611, 0.127389	0.151717, 0.206186
SCTYPE(1) RSL(1) OR RSL(2)	=>	436	405, 31	0.928899, 0.071101	0.448505, 0.319588
SCTYPE(5) RSL(2) OR RSL(1)	=>	55	5, 50	0.090909, 0.909091	0.051546, 0.055371
SCTYPE(2) RSL(1) OR RSL(2)	=>	251	220, 31	0.876494, 0.123506	0.243632, 0.319588
SCTYPE(4) RSL(1) OR RSL(2)	=>	101	91, 10	0.90099, 0.09901	0.100775, 0.103093

From table 6, we found:
- The largest number of the students is examined from governmental schools then from the teachers unions.
- There are success and fail on all the types of the schools.

5 Discussion

Thus, from all the results that obtained about the Sudan School Certificate, we have observed the following points:

1. The success in the Sudan School Certificate depends only on the success in the four basic subjects (Arabic language, English language, Religion, and Mathematic).
2. If the student fails in one of the four basic subjects he or she will fail the final result.
3. For some students, there are fail in the final result although they are passing the three scientific subjects. This is because they are failing in one or more of the basic subjects.
4. For some students, There are pass in the final result although they are failing in one or more of the three scientific subjects. This is because they are passing all the four basic subjects.
5. The percentage of the success on female is greater than the percentage of the success on male (female=62.7%, male=27.6).
6. The largest number of the students is examined from governmental schools and there are success and fail on all the types of the schools.
7. The largest number of the examined students is from Khartoum state and from Al_Jazeira state. This is normal thing because Khartoum is the capital of Sudan and Al_Jazeira state is the second state after Khartoum state.
8. The smallest number of the examined students is from the states of Darfor and Kurdufan.
9. We have made a test to know the relation between the school type and the state attributes. We got out with about 48 rules one of them is:

SCTYPE(4) AND STATECODE(1) => RSL(1) OR RSL(2)

This means if the school type is special and the state is Khartoum then the result is pass or fail. We observed an important point which is, there are 101 students who are examined form special schools and from Khartoum state, 91 of them are passing the final result and the remaining 10 students are failing the final result.

10. From the last point, we found that the largest number of the students is examined from Khartoum state and most of them are examined from special schools.

11. In the case of passing all the four basic subjects and failing in one or more of the three scientific subjects, the final result is useless because it doesn't qualifies the student for competition to enter the university.

6   Conclusion

In this research we have applied Rough Set Theory to Sudan School Certificate Examination. By using Rough Set we didn't achieve information about success and fail in the final result only, but also we obtained general overview about other factors related to students like the type of schools which had the large number of students and the state which had the large number of students. Also we are able to compare between the percentages of success and fail in different factors. From the data analysis we could find that the main rule for the student to pass the final result, he or she should pass all the four basic subjects rather than pass the other three subjects. Also we could find the large number of the students examined from Khartoum state and most of them studied in special schools. In general, the governmental school had the largest number of students. Moreover, the percentage of success in female is greater than in male.

References

1. Pe-a, J.M., Létourneau, S.: Famili: application of Rough Sets Algorithms to Prediction of Aircraft Component Failure. In: 3rd International Symposium on Intelligent Data Analysis, Netherlands (1999)
2. Marwala, T., Crossingham, B.: Bayesian Approach to Rough Set. arXiv: 0704.3433
3. Swiniarski, R., Skowron, A.: Rough set methods in feature selection and recognition. Pattern Recognition Letters 24, 833–849 (2003)
4. Wang, X., Yang, J., Jensen, R., Liu, X.: Rough Set Feature Selection and Rule Induction for Prediction of Malignancy Degree in Brain Glioma. Computer Methods and Programs in Biomedicine 83(2), 147–156 (2006)
5. Komorowski, J., Polkowski, L., Skowron, A.: Rough sets: A Tutorial
6. Greco, S., Matarazzo, B., Slowinski, R.: Rough sets theory for multi-criteria decision analysis. European Journal of Operational Research 129 (2001)

7. Düntsch, I., Gediga, G.: Rough set data analysis. In: Encyclopedia of Computer Science, Marcel Dekker, New York (2000)
8. An, A., Huang, Y., Huang, X., Cercone, N.J.: Feature selection with rough sets for web page classification. In: Peters, J.F., Skowron, A., Dubois, D., Grzymała-Busse, J.W., Inuiguchi, M., Polkowski, L. (eds.) Transactions on Rough Sets II. LNCS, vol. 3135, pp. 1–13. Springer, Heidelberg (2004)
9. Hedar, A., Wang, J., Fukushima, M.: Tabu search for attribute reduction in rough set theory. Soft Computing 12, 909–918 (2008)
10. Arabi, N.: An Application of Rough Set theory to Sudanese Secondary Certificates. Master Thesis, University of Khartoum, Sudan (2008)
11. Hrn, A., Komorowski, J., Skowron, A., Synak, P.: ROSETTA-Part I: System Overview
12. Admission Policies for Higher Education Institution. Ministry of Education, Sudan (1999)
13. General Directorate of Educational Planning. In: Educational Statistics 2001-2002. Ministry of Education, Sudan (2003)