International Journal of Advanced Science and Technology
Vol. 45, August, 2012

A Capability Maturity Decision Making System for Educational Quality Based on Rough Computing

R. Manjula, J. Vaideeswaran and D. P. Acharjya

School of Computing Sciences and Engineering, VIT University, Vellore, India
rmanjula@vit.ac.in, jvaideeswaran@vit.ac.in, dpacharjya@gmail.com

Abstract

Educational quality varies in the degree to which they attempt to deal with different completing aspects of quality such as relative importance of various quality attributes of key process areas. With the emerging of information and communication technology, educational quality has drastically changed. This revolution has brought radical change in the way educational data are generated for ease of decision making. It is well established fact that use of information at the right time provides an advantage to educational quality. But the real challenge lies in converting high dimensional data into knowledge. Though present technology helps in creating databases, but most of the data may not be relevant for formulating a quality educational model. In this paper, we propose a new capability maturity decision making model based on rough computing for extracting key process areas and its relevance for the development of quality education. In particular, 769 educational institutions data are considered for validation and the results shows the practical viability of the proposed model.

Keywords: *Rough set, indiscernibility, information system, rule generation, rule validation, approximation space, decision table, quality education, maturity evaluation*

1. Introduction

Education in general and higher education in particular represents too process oriented, intangible, and multiple-stakeholders situation. Most of the performance measurement systems of higher educational institutions do not reflect the full range of interested stakeholders and are not easily linked to the strategic and quality management. Therefore, balanced scorecard approach in order to reinforce the importance of managing rather than just monitoring performance is proposed [8]. Garvetson [7] confirms the importance of the expectations of key stakeholders in the educational process. A general criticism about the level of institutions relates to the methodology that addresses quality in a superficial way whereas it projects a complex image. While obtaining the level of quality education, some categorize institutions and publish level for each category. In some cases, weights to indicators are assigned and weighted scores are used to scale the level of the institutions. But, the indicators considered for leveling the institutions is not unique. Some times, it is observed that levels for the institutions vary based on the methods used to analyze the data [11].

Another limitation in leveling the institutions is arbitrariness in giving weightings. *US News and World Report* (*UNSWR*) gives 16% weight to graduation rate whereas Macleans give 2%. Similarly, *Times Higher Education Supplement* (*THES*) provides 5% score for the proportion of academic staff recruited from overseas and 5% score for international students. However, these indicators used are no way related to quantify the academic quality [12].

Statistical validity is another limitation for quantifying the level of institutions. In addition to statistical validity majority of them fail with normal tests of reliability and validity. On the other end, while considering multiple leveling of the systems, it becomes questionable whether meaningful differences between institutions exist. Guarino [4] attempted these problems by using Bayesian latent variable model to estimate institutions quality based on a set of observable features or characteristics.

Avdjieva and Wilson [9] suggest, higher education institutions are now required to become learning organizations, where internal stakeholders also interpret and access the quality of higher education provision. As a result, many higher education institutions are looking forward to adopt different accreditations for quality improvement in higher education and implementing total quality management practices in order to achieve quality goals. Therefore, education system engineering is an inter-disciplinary concept, which spans the dimensions of academic, infrastructure, facilities, administration etc. Some of the potential benefits include continual improvements in system quality and adhering to global standards. Therefore, the increasing competency in information technology and education sectors necessitates a process maturity evaluation methodology.

The convergence of computers and information technology revolution observed in the recent past has brought a drastic change in the educational system. It is rightly established fact that the use of information at the right time and the right point provides to gain better knowledge in educational assessment methodology. Although knowledge mining from the databases is increasingly important, but the knowledge discovered is not always useful to users. This is because the discovered knowledge does not necessarily fit a user's interest, and may be redundant or inconsistent with a priori knowledge. Therefore, the real challenge lies in converting high dimensional data into knowledge, and to use this knowledge to make quality improvement properly. Though present technologies help in obtaining decisions by creating databases, however it is observed that most of information may not be relevant. So attribute reduction becomes an important factor for handling such large database by eliminating superfluous data to enable decision making in an effective manner. Many traditional tools to mine knowledge from the rapidly growing data are proposed but most of these are crisp, deterministic and precise in character. But real life situations are quite opposite to that. For a complete description of a real time system, often one would require by far more detailed data than a human being could recognize and understand simultaneously for ever. This gives to the extension of the concept of crisp sets so as to model imprecise data that can enhance their modeling power. In general, statistical inferences on the existing data are carried out by different researchers but this tendency gets accentuated by increased interest in making efficient use of organizational data through data mining and data warehousing [3]. Therefore, there is enough scope for consideration of some of the newer techniques such as rough set [13, 14] which has developed in recent past. In this paper we propose a new capability maturity decision making model for educational system using rough set theory as developed and studied by Pawlak and Skowron [15, 16, 17]. The basic aim of this model is to extract rules from empirical data and to use it in key process areas for quality improvement in educational system.

This paper uses the basic idea of rough set theory to discover rules between the attributes present in the dataset. The rest of the paper is organized as follows: Section 2 presents the basics of rough set theory. The proposed capability maturity decision model for getting quality education is presented in Sections 3. In Section 4, a empirical study on educational quality is presented. This is further followed by a conclusion in Section 5.

2. Foundations of Rough Set Theory

The rough set philosophy was initially developed for a finite universe of discourse based on the assumption that with every object of the universe of discourse, we associate some information. Objects characterized by the same information are indiscernible in view of the available information about them. The indiscernibility relation generated in this way is the mathematical foundation of rough set theory. In rough set theory of Pawlak [13, 14] data is organized in a decision table, in which each row represent an object and each column represents an attribute. Let U be a finite nonempty set called the universe of discourse. Suppose $R \subseteq (U \times U)$ is an equivalence relation on U. The equivalence relation R partitions the set U into disjoint subsets. Elements of the same equivalence class are said to be indistinguishable. Equivalence classes induced by R are called elementary concepts. Every union of elementary concepts is called a definable set. The empty set is considered to be a definable set, thus all the definable sets form a Boolean algebra. The order pair (U, R) is called an approximation space. Given a target set $X \subseteq U$, we can characterize X by a pair of lower and upper approximations. We associate two subsets $\underline{R}X$ and $\overline{R}X$ called the R-lower and the R-upper approximations of the set X. It is defined as:

$$\underline{R}U = \bigcup \{Y \in U / R : Y \subseteq X\}$$

and $$\overline{R}X = \bigcup \{Y \in U / R : Y \cap X \neq \phi\}$$

The R-boundary of X, $BN_R(X)$ is given by $BN_R(X) = \overline{R}X - \underline{R}X$. We say X is rough with respect to R if and only if $\overline{R}X \neq \underline{R}X$.

2.1. Information System

Let U be a nonempty finite set of institutions called the universe of discourse and A be a nonempty finite set of attributes called the key process areas. An information system can be represented as a pair $I = (U, A)$. A decision system is any information system of the form $I = (U, A, C, D)$, where $C, D \subseteq A$ be two subsets of attributes, called condition and decision attributes respectively. The equivalence classes of the relations $IND(C)$ and $IND(D)$ are called condition and decision classes respectively. The information system I can also be represented as $I = (U, A, V_x, f_x)$. For every $x \in A, V_x$ is the set of values that attribute x may take and $f_x : U \rightarrow V_x$ is an information function. Thus, f is a function from $(U \times A) \rightarrow V$ [6, 13, 14].

For example, table 1 illustrates an example of information system of five institutions that are characterized with five key processes a_1, a_2, a_3, a_4, a_5 and decision d, where a_1 represents intellectual capital, a_2 represents infrastructure facility, a_3 represents placement performance, a_4 represents industry institute interaction, a_5 represents teaching learning process whereas d represents level of the institution. Columns of the table are labeled by key processes and rows by institutions. Each cell of the table provides key process values. Thus, each row of the table can be seen as information about specific institutions. For example, institution I_4 is characterized in the table by the attribute value set (a_1, very high), (a_2, very high), (a_3, high), (a_4, low), (a_5, very high) which form the information about the institution.

Table 1. Sample Information System

Institutions	a_1	a_2	a_3	a_4	a_5	***D***
I_1	High	Very high	Very high	High	High	Level 3
I_2	Medium	High	High	Low	Medium	Level 1
I_3	High	Medium	Medium	Medium	High	Level 1
I_4	Very high	Very high	High	Low	Very high	Level 2
I_5	Very high	Very high	High	Very high	High	Level 4

2.1. Indiscernibility Relation

Especially at the age of digital data, with every object of the universe of discourse we associate some information and the objects can be accessible through this information only. Therefore, it is very difficult to uniquely identify these objects based on available set of attributes of the objects. This is because of lack of sufficient information about the objects of the universe. Hence, object with the same information can not be distinguished and appear as same. This leads to the concept of indiscernibility [2]. Thus, we need classification of the objects into similarity classes to characterize these objects. The indiscernibility relation generated in this way is the mathematical basis of rough set theory [14]. Any set of all indiscernible objects is called an elementary concept and forms a basic knowledge about the universe. Any union of the elementary concepts is referred to be either crisp (precise) set or rough (imprecise) set.

Let $P \subseteq A$ and $x_i, x_j \in U$. Then we say x_i and x_j are indiscernible by the set of attributes P in A if and only if $f(x_i, a) = f(x_j, a)$ for all $a \in P$. For example, $\{I_1, I_3\}$ are indiscernible for the attribute a_1. Similarly, $\{I_4, I_5\}$ are also indiscernible for the attribute a_1.

3. Proposed Capability Maturity Decision Model

Maturity decision models are useful for quality education that emphasize on key process improvement. In general, it assists institutions in determining the different key process improvements so as to move to the next level in quality education. There are many maturity hierarchical models developed and analyzed by different researchers [1, 5, 10]. A traditional capability maturity model is shown in Figure 1. But these models are lacking in taking decisions to switch from one level to another without any hierarchical structure. It is observed that an organization moves from one level to another in hierarchy based on different focus areas. However, in real life situations it is not true. Some organizations move from one level to other without any hierarchical structure. Therefore, it is essential to identify the importance of each key process areas to attain a certain level. To this end, we use rough computing for extracting key process areas and its relevance for the development of quality education. The main advantage of this model is that, it works for both categorical and numerical data.

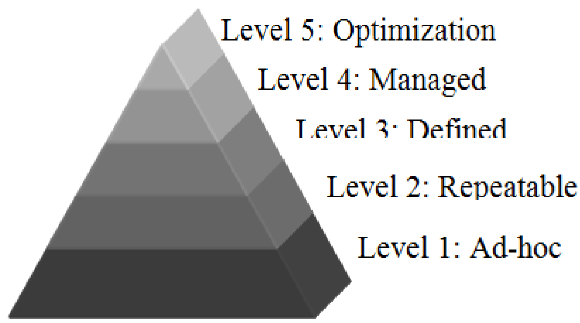

Figure 1. Traditional Development Model

Now, we propose our capability maturity decision model as shown in Figure 2. The model consists of problem definition, data preparation, data partition, rule generation and rule validation. Problem definition and incorporation of prior knowledge are the fundamental steps of any model in which we identify the right problem. Secondly, proper structuring the corresponding objectives and the associated attributes is done. Finally, a target dataset is created or collected on which data mining is to be performed. Before further analysis, a sequence of data cleaning tasks such as consistency check, removing noise and data completeness is done to ensure that the data are as accurate as possible. This step takes the most time needed for the whole knowledge and decision extracting process. We discuss in detail the subsequent steps of the architecture design of the proposed capability maturity decision model.

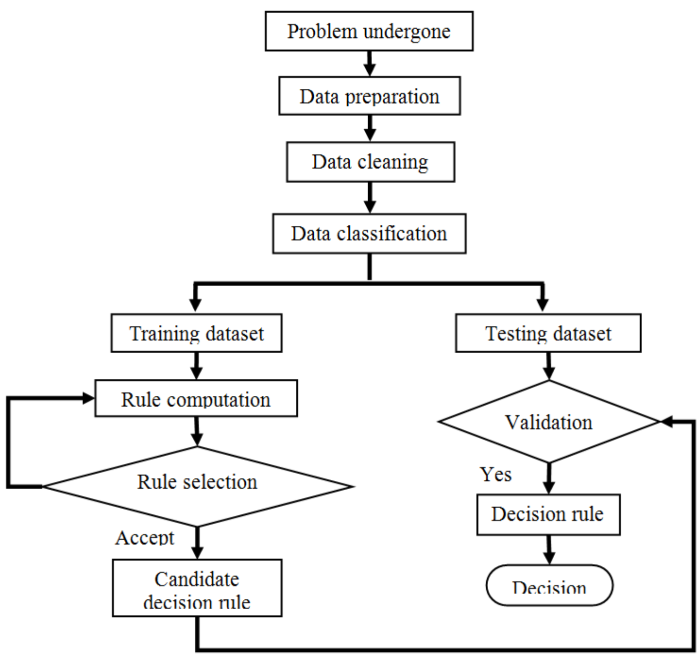

Figure 2. Capability Maturity Decision Model

3.1. Rule Generation and Validation Algorithm

In this section we propose a rule generation and validation algorithm that generates all the possible reducts by eliminating all dispensable attributes and deriving the candidate decision rules from the training dataset. Also, it examines the objects in the testing dataset to estimate the validity of the candidacy rules obtained. We apply the following steps in order to generate and validate the procedure.

Algorithm

Input: Target information system
Output: Validated decision rules

1. Set object number $i = 1$
2. Choose object i from the training data set and compute a set of reducts for all the condition attributes.
3. Replace $i = i + 1$
4. If all objects have been chosen, then go to step 4; else go to step 2.
5. Compute the number of supporting objects for each reduct after combining the identical reducts.
6. Obtain the decision rules from the selected reducts.
7. Compare each decision rule with each new object from the testing data set. Compute the number of objects that support with the rule.
8. Repeat step 7 for all decision rules obtained at step 6.
9. Calculate the accuracy of each rule by using the following equation.

$$\text{Accuracy} = \frac{\text{Total number of supported objects}}{\text{Sum of supported and non-supported objects}}$$

10. If the accuracy is greater than predefined threshold, then go to step 11; else remove the decision rule obtained at step 6.
11. Process terminates and writes the validated rules.

3.2. A Numerical Illustration

In this section we explain our capability maturity decision model with a numerical illustration. Consider a target dataset with 19 objects as shown in Table 2, where the attributes $A_1, A_2, A_3, \cdots, A_6$ are defined based on the problem objective. Before we compute and validate the rules, data cleaning and preprocessing are carried out. In particular, we remove the objects o_8, o_{11}, and o_{12} that contain missing attribute values. We divide randomly the target dataset into two groups called training dataset and testing dataset. The candidate decision rules are obtained from training dataset whereas the testing dataset is used to detect over fitting of the decision rules based on the threshold value as decided by the educational experts. Thus, the testing dataset allows one to estimate the validity of the rules generated from the training dataset. To this end we have divided the target dataset randomly into a training dataset that contained first 9 objects o_1, $o_2, \cdots$, o_7, o_9, o_{10} and the testing data set that contained the remaining 7 objects o_{13}, o_{14}, $\cdots$, o_{19}.

Following steps 2-4 of the algorithm, a set of decision rules based on the attributes A_1, A_2, $\cdots$, A_6 are generated with their supporting objects. The candidate rules generated are given in Table 3. For example, rule 1 is denoted as $\times\times 3\times\times\times 2$. This leads to the following decision rule: IF A_3 = 3, THEN the value of the decision attribute is 2. Similarly, we can also obtain the other decision rules. Finally, following steps 7-10 of the algorithm, the rules obtained are validated with the testing dataset.

Table 2. Numerical Illustration Dataset

Objects	A_1	A_2	A_3	A_4	A_5	A_6	Decision
o_1	1	3	2	2	4	3	1
o_2	2	4	1	2	4	1	1
o_3	2	4	8	1	5	1	1
o_4	1	5	1	2	5	1	1
o_5	1	4	3	1	4	1	2
o_6	1	1	7	1	2	1	2
o_7	1	2	3	1	2	3	2
o_8	2	2	--	--	1	2	2
o_9	1	1	3	1	2	1	2
o_{10}	2	4	2	1	4	3	1
o_{11}	2	--	--	--	--	1	--
o_{12}	--	--	3	--	2	--	1
o_{13}	1	3	3	1	4	1	2
o_{14}	2	1	5	2	1	2	2
o_{15}	1	4	3	1	3	2	2
o_{16}	2	1	7	2	2	1	1
o_{17}	2	2	5	2	2	1	1
o_{18}	1	4	3	1	2	3	2
o_{19}	1	1	7	1	3	3	2

Table 3. Candidacy Rule Computation

Rule	A_1	A_2	A_3	A_4	A_5	A_6	Decision	Supporting objects
[1]	×	×	3	×	×	×	2	o_5, o_7, o_8
[2]	×	×	7	×	×	×	2	o_6
[3]	×	×	8	×	×	×	1	o_3
[4]	×	×	×	×	5	×	1	o_3, o_4
[5]	×	×	×	×	2	×	2	o_6, o_7, o_8
[6]	2	×	×	×	×	×	1	o_2, o_3, o_9
[7]	×	1	×	×	×	×	2	o_6, o_8
[8]	×	×	×	2	×	×	1	o_1, o_2, o_4

Compare the first decision rule obtained in Table 3 with each new object $o_{13}, o_{14}, o_{15}, o_{16}, o_{17}, o_{18}, o_{19}$ from the testing data set. The number of objects that support the first rule is 3 whereas the number of objects that do not support the rule is 0. Thus, the accuracy is given as

$$Accuracy = \frac{\text{Number of supported objects}}{\text{Sum of the supported and non-supported objects}} = \frac{3}{3+0} = 100\%$$

Again, the number of objects that support the rule number 6 is 2 whereas the number of objects that do not support the rule is 1. Therefore, the accuracy is given as

$$Accuracy = \frac{\text{Number of supported objects}}{\text{Sum of the supported and non-supported objects}} = \frac{2}{2+1} = 66\%$$

Similarly, the accuracy of the other candidacy rules may be obtained. Finally, if the accuracy is greater than the predefined threshold value, then the corresponding rule will be selected else the rule will be removed from the candidacy rules. Therefore, on increasing reasonably the threshold value, we can get better knowledge and decision.

4. An Empirical Study on Educational Quality

In this section, we demonstrate how the decision model can be applied to get quality education. We apply rough computing for extracting rules from the dataset obtained from different educational institutions. However, we keep the identity confidential due to some specified official reason. The institutions can be judged by the outcomes that are produced. The quality of the output can be judged by the placement performance of the institute. However, there are various other parameters that influence the quality education. To produce the quality output, the input should be of high quality and the major inputs for an institution are infrastructure facilities required for imparting quality education, intellectual capital, and product realization. Product realization is considered more important than placement because it identifies the degree of placement at top level management, entrepreneurs, and academic extension growth. Academic quality outcome is considered to measure the quality outcome at two different levels namely student and faculty and hence it is essential for quality education. Teaching and learning process is a fundamental attribute for any educational institution and hence is a major part in quality education. In many institutions, the academic quality and maturity are enhanced with other non-academic attributes and hence extracurricular activity is introduced. This helps the students for team building and to develop leadership quality. Apart from this there are many other attributes such as strategy planning, research consultancy and extension, and industry institution interface are also involved in the development of quality education.

This study began by collecting data from different educational institutions. The 769 institutions data were checked for completeness and consistency. We have removed unrelated items in the data in order to avoid unnecessary complexity. Among the institutions, 57 institutions data were removed from the dataset, because none of them had sufficient support. Also, 123 institutions data were removed from the dataset, because of missing attribute values. Similarly, 63 institutions which had junk data were removed from the dataset during analysis. In total, 243 institutions data were removed from the dataset. Here the data is collected only to identify the levels of the institutions and the decisions that lead to that level. Literature and numerical values based on different symptoms were collected and studied. These parameters form the attribute set for our analysis. The level decision of the institutions was also considered, and it becomes our decision variable. The various major parameters that play a vital role in quality education are used in our analysis. The different parameters considered are given in Table 4. We randomly divided the 526 institutions data into the training dataset that contained 289 institutions data (55%) and the testing data set that contained 237 patients (45%). A graphical view of the dataset characteristics is presented in Figure 3.

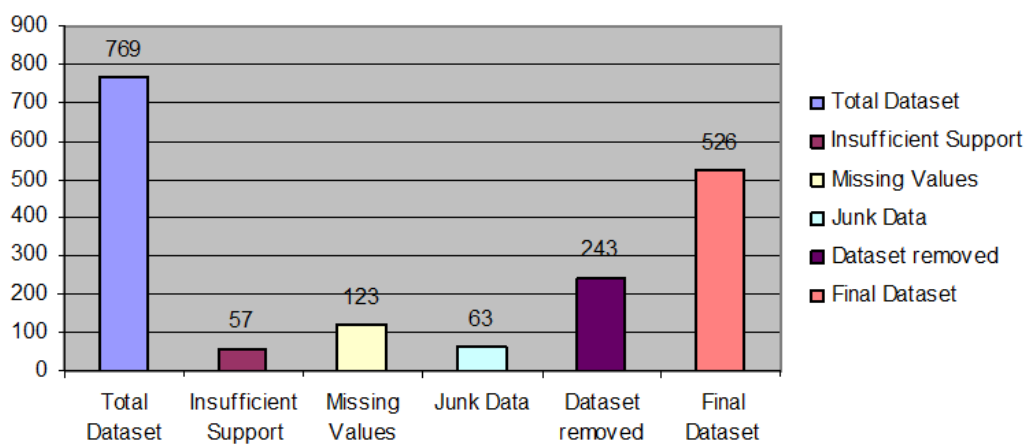

Figure 3. Graphical View of Dataset

The information system as discussed in Table 1 contains data about the universe *U*, attributes and decision attribute. Our basic objective is to obtain rules that effect the decision and will help us to develop a quality education. We normalize the data and formed a classifying rule to categorize the attribute values into different groups. The different cut off points for classifying the attribute values into different categories are made with the help of educational experts. The normalized information is given in Table 4. To make our analysis simple we have assigned some value to each classification group. However, these values are optional and do not affect the analysis. Keeping view at the length of the paper, a sample training dataset is presented in Table 5, where we keep the institutions identity confidential as it does not affect our analysis.

Table 4. Attribute Normalization and Classification

Attributes	Notation	Normalized value	Classification
Intellectual Capital (IC)	A_1	0 – 45 46 – 90 91 – 120 > 120	Low (1) Medium (2) High (3) Very high (4)
Infrastructure (IF)	A_2	0 – 60 61 – 75 76 – 90 > 90	Low (1) Medium (2) High (3) Very high (4)
Placement (PC)	A_3	0 – 70 71 – 85 86 – 95 > 95	Low (1) Medium (2) High (3) Very high (4)
Product Realization (PR)	A_4	0 – 75 76 – 85 86 – 90 > 90	Low (1) Medium (2) High (3) Very high (4)
Academic Quality Outcome (AQO)	A_5	0 – 25 26 – 37 38 – 45 >45	Low (1) Medium (2) High (3) Very high (4)
Extracurricular Activity	A_6	0 – 25 26 – 37 38 – 45 > 45	Low (1) Medium (2) High (3) Very high (4)
Teaching and Learning Process	A_7	0 – 170 171 – 176 177 – 180 > 180	Low (1) Medium (2) High (3) Very high (4)
Strategy Planning	A_8	0 – 35 36 – 40 41 – 45 > 45	Low (1) Medium (2) High (3) Very high (4)
Research, Consultancy, and Extension	A_9	0 – 10 11 – 30 31 – 70 > 70	Low (1) Medium (2) High (3) Very high (4)
Industry Institute Interface	A_{10}	0 – 15 16 – 35 36 – 40 > 40	Low (1) Medium (2) High (3) Very high (4)
Technical & Generic Competencies	A_{11}	0 – 25 26 – 30 31 – 43 > 43	Low (1) Medium (2) High (3) Very high (4)

Table 5. Sample Training Dataset

Objects	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}	A_{11}	Level
I_1	2	1	2	2	3	2	4	2	3	2	1	2
I_2	2	2	3	3	3	2	3	3	3	3	2	3
I_3	2	1	1	1	2	1	3	2	1	1	1	1
I_4	3	3	3	4	3	4	4	4	4	4	4	4
I_5	2	3	2	3	2	1	3	2	3	2	1	3
I_6	4	4	3	4	3	2	3	4	4	4	4	1
I_7	3	2	3	3	3	4	4	4	3	3	3	3
I_8	2	2	1	1	2	2	2	1	1	1	1	1
I_9	2	3	2	3	3	1	3	2	4	4	4	3
I_{10}	3	2	2	1	2	2	3	2	2	1	2	2
I_{11}	4	3	4	4	4	3	3	4	4	4	4	4
I_{12}	2	3	1	2	2	2	3	2	2	1	2	2
I_{12}	1	1	1	1	3	2	4	1	1	1	1	1
I_{14}	3	3	3	2	3	1	4	2	3	3	2	3
I_{15}	3	4	3	4	3	2	4	4	4	4	4	4

4.1. Rule Generation and Validation

We employed the training dataset to derive the rules and the final rules are selected based on validation procedure. The training data set of 289 institutions obtained is further classified into four categories- level 1, 2, 3, and 4. Each category expresses different levels of quality education. The data collected under each level is obtained as 57, 93, 88, and 51 respectively. The unusual outcomes are also removed from each level. Along with the unusual outcomes, we have also removed the identical objects from the training dataset for each level in order to avoid unnecessary analysis. The number of identical objects removed for level 1 is 9 whereas for level 2 is 4. Similarly, number of identical objects removed for level 3 and 4 are 6 and 1 respectively. According to the algorithm, the rules were determined and validated. A decision rule is selected only when it has at least two supporting objects. The total number of candidate rules generated in level 1 for further validation is 6 and is summarized in Table 6. Similarly, the total number of rules generated in level 2 for further validation is 27 and is summarized in Table 7. The total number of rules generated in level 3 for further validation is 15 and is summarized in Table 8. Finally, the total number of rules generated in level 4 for further validation is 16. The number of supporting institutions for each rule generated in level 4 is summarized in Table 9.

Table 6. Candidacy Rules for Level 1

Rule	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}	A_{11}	Level	Supporting institutions
[1]	×	×	×	×	×	×	×	1	×	×	×	1	11
[2]	×	×	×	×	×	×	×	×	1	×	×	1	15
[3]	×	1	1	×	×	×	×	×	×	×	×	1	2
[4]	×	×	×	×	×	×	×	×	×	1	1	1	3
[5]	×	1	×	1	×	×	×	×	×	×	×	1	8
[6]	×	×	1	×	×	×	×	×	×	×	1	1	9

Table 7. Candidacy Rules for Level 2

Rule	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}	A_{11}	Level	Supporting institutions
[1]	2	×	×	×	×	×	4	×	×	×	×	2	4
[2]	×	1	2	×	×	×	×	×	×	×	×	2	2
[3]	×	2	×	×	×	×	×	2	×	×	×	2	5
[4]	×	×	2	×	×	2	×	×	×	×	×	2	4
[5]	3	×	×	×	2	×	×	×	×	×	×	2	6
[6]	3	×	×	×	×	×	3	×	×	×	×	2	4
[7]	×	3	1	×	×	×	×	×	×	×	×	2	3
[8]	×	1	×	×	×	×	×	×	×	2	×	2	4
[9]	×	×	2	1	×	×	×	×	×	×	×	2	2
[10]	×	×	×	2	×	×	×	×	×	1	×	2	3
[11]	×	1	×	×	×	×	×	×	3	×	×	2	3
[12]	×	×	×	×	3	×	×	×	×	2	×	2	3

[13]	×	×	×	×	×	×	4	×	×	2	×	2	2
[14]	2	1	×	×	3	×	×	×	×	×	×	2	5
[15]	2	1	×	×	×	2	×	×	×	×	×	2	5
[16]	×	×	×	2	2	×	×	×	×	×	×	2	2
[17]	×	×	×	2	×	2	×	×	×	×	×	2	4
[18]	×	×	×	×	×	2	×	2	×	×	×	2	3
[19]	2	×	×	×	3	×	×	×	×	×	1	2	3
[20]	3	2		×	×	×	×	×	×	×	2	2	2
[21]	×	1	×	×	3	×	×	2	×	×	×	2	4
[22]	×	1	×	×	×	×	4	2	×	×	×	2	3
[23]	×	2	×	×	×	×	3	×	×	1	×	2	2
[24]	×	×	2	×	3	×	×	×	3	×	×	2	2
[25]	×	×	1	×	×	2	3	×	×	×	×	2	4
[26]	×	×	×	×	3	×	×	×	3	×	1	2	2
[27]	×	×	×	×	×	×	4	2	×	×	1	2	3

Table 8. Candidacy Rules for Level 3

Rule	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}	A_{11}	**Level**	**Supporting institutions**
[1]	×	×	×	3	×	×	×	×	×	×	×	3	2
[2]	×	×	×	×	×	×	×	×	×	3	×	3	3
[3]	×	2	3	×	×	×	×	×	×	×	×	3	2
[4]	×	2	×	×	3	×	×	×	×	×	×	3	11
[5]	×	3	2	×	×	×	×	×	×	×	×	3	4
[6]	×	3	×	×	×	1	×	×	×	×	×	3	6
[7]	×	×	2	×	×	1	×	×	×	×	×	3	6
[8]	×	×	×	×	3	1	×	×	×	×	×	3	4

[9]	×	×	×	×	×	1	×	×	3	×	×	3	5
[10]	2	2	×	×	×	×	3	×	×	×	×	3	10
[11]	2	×	×	×	3	×	3	×	×	×	×	3	7
[12]	×	3	×	×	3	×	×	2	×	×	×	3	9
[13]	×	3	×	×	×	×	×	×	3	×	×	3	3
[14]	×	×	×	×	×	×	3	×	3	×	×	3	8
[15]	×	×	×	×	×	×	×	×	3	×	2	3	2

Table 9. Candidacy Rules for Level 4

Rule	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}	A_{11}	Level	Supporting institutions
[1]	3	×	3	×	×	2	×	×	×	×	×	4	4
[2]	3	×	×	×	3	2	×	×	×	×	×	4	4
[3]	3	×	×	×	×	2	4	×	×	×	×	4	5
[4]	×	×	×	×	×	3	×	×	×	×	×	4	3
[5]	3	2	×	×	×	×	×	×	×	×	×	4	3
[6]	×	4	×	×	×	×	4	×	×	×	×	4	3`
[7]	×	3	×	×	×	×	×	4	×	×	×	4	3
[8]	4	3	×	×	×	×	×	×	×	×	×	4	2
[9]	3	×	×	×	×	×	×	×	4	×	×	4	2
[10]	×	×	×	×	×	×	×	4	×	4	×	4	3
[11]	×	×	4	×	×	×	×	×	×	×	×	4	3
[12]	×	×	3	×	×	2	4	×	×	×	×	4	4
[13]	×	×	×	×	×	2	4	4	×	×	×	4	3
[14]	×	×	×	×	×	4	×	×	×	×	4	4	2
[15]	×	×	×	×	×	×	4	×	4	×	×	4	2
[16]	×	×	×	4	×	×	4	×	×	×	×	4	4

The testing dataset of 237 institutions is considered for further validation of the rules generated through the procedure. Finally, the accuracy of each candidate rule generated is examined to estimate the corresponding validity. We present the total number of supporting, non-supporting institutions and the accuracy of each candidate rule in table 10. Keeping view at the length of the paper, we consider level 1 to show the validation procedure. Since the accuracy of candidate rules 1, 2, and 6 is greater than the defined threshold 70%, these three rules were validated after applying the rule validation procedure whereas candidate rules 3, 4, and 5 are discarded as it falls less than the predefined threshold value. A graphical view of the algorithm for the level 1 is given in Figure 4.

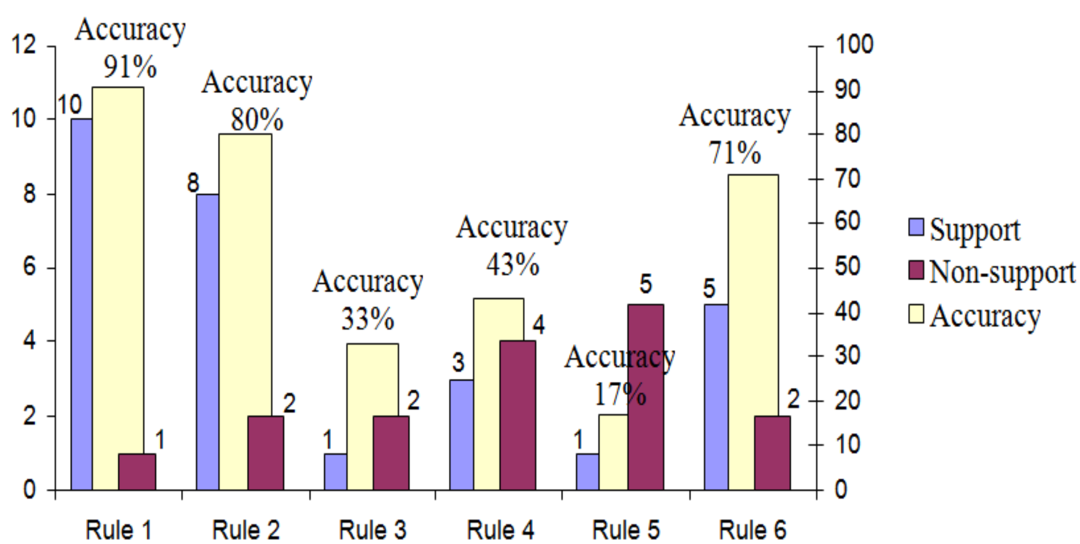

Figure 4. Graphical View of Validation of Level 1

Out of 27 rules generated in level 2, the numbers of rules that fall more than the defined threshold value of 70% after validation are 19. The rules 2, 7, 9, 13, 20, 23, 25, 27 are discarded from table 7 after the validation phase of the algorithm. Similarly, out of 15 rules generated in level 3, the numbers of rules that fall more than the defined threshold value of 70% after validation are 9. The rules 1, 2, 3, 5, 13, 15 are discarded from table 8 after due validation. Again out of 16 rules generated in level 4, the number of rules that fall more than the defined threshold value 70% after validation are 9. The rules 4, 5, 7, 8, 9, 11, 14 are discarded from Table 9 after the validation phase of the algorithm. An overall graphical view of the empirical study on quality education is given in the following Figure 5.

Table 10. Candidacy Rule Validation for Level 1

Rule		Description	Support	Non-support
[1]	IF	strategy planning is low,	10	1
	THEN	we can infer that the institution is in level 1	Accuracy: 91%	
[2]	IF	research, consultancy and extension is low,	8	2
	THEN	we can infer that the institution is in level 1	Accuracy: 80%	
[3]	IF	infrastructure is low; and placement is low	1	2
	THEN	we can infer that the institution is in level 1	Accuracy: 33%	
[4]	IF	industry institute interface is low; and technical and generic competencies is low,	3	4
	THEN	we can infer that the institution is in level 1	Accuracy: 43%	
[5]	IF	infrastructure is low; and product realization is low,	1	5
	THEN	we can infer that the institution is in level 1	Accuracy: 17%	
[6]	IF	placement is low; and technical and generic competencies is low,	5	2
	THEN	we can infer that the institution is in level 1	Accuracy: 71%	

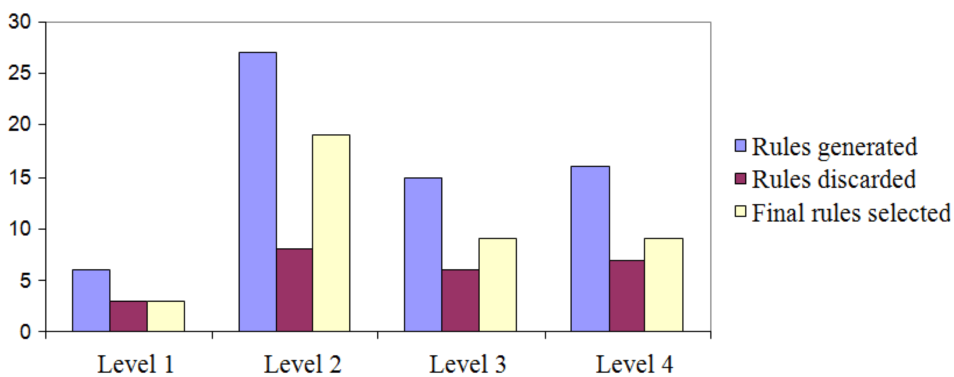

Figure 5. Overall Graphical Views of Rule Generation and Validation

5. Conclusion and Future Extension

Rough computing is a useful data mining tool for finding decisions from the information system. In this paper, we have presented a capability maturity decision making model as an application of rough computing. We have shown how analysis can be done by taking rough computing as a model for mining knowledge. The main objective of this work is to expand the domain application of rough computing and to find the key process areas that influence the educational quality system. The rules generated are also validated based on the threshold value. On increasing the threshold value reasonably, a better educational quality can be achieved. Also, the decision rules generated through the model can act as an expert system that may be used for future strategic decision making. This study is employed based on the empirical data collected from different institutions over a period of time. The total number of rules generated is 64. This is further minimized to 40 by validation procedure with the help of threshold value. From the validated rules the key process affecting the decisions for each level can be identified. The results obtained show practical viability of this data mining approach in the quality education. However, the requirement of indiscernibility relation in rough set is a restrictive condition and may limit the application of the rough set model. Also, to solve problems in identifying institutions providing quality education, we can not successfully use classical methods as it involves uncertainties. Considering all these factors, we plan to develop hybrid models that use some specific combination of rough set, fuzzy set, intuitionistic fuzzy set, formal concept analysis and theory of probability to obtain more specialized decisions and the chief factors affecting the decisions.

References

[1] A. Bicego and P. Kuvaja, "Software process maturity and certification", Journal of Systems Architecture, vol. 42, **(1996)**, pp. 611-620.

[2] B. K. Tripathy and D. P. Acharjya, "Association rule granulation using rough sets on intuitionistic fuzzy approximation spaces and granular computing", Annals. Computer Science Series, Tome, vol. 9, no. 1, **(2011)**, pp. 125-144.

[3] B. Malcolm, C. Bruce and M. Peter, "Knowledge Discovery in Marketing: An approach through rough set theory", European Journal of Marketing, vol. 35, no. 7-8, **(2001)**.

[4] C. Guarino, G. Ridgeway, M. Chun and R. Buddin, "Latent variable analysis: A new approach to university ranking", Higher Education in Europe, vol. 30, no. 2, **(2005)**, pp. 147-165.

[5] C. Lutteroth, A. L. Reilly, G. Dobbie and J. Hamer, "A maturity model for computing education", Ninth Australasian Computing Education Conference, Conferences in Research and Practice in Information Technology, vol. 66, **(2007)**, pp. 107-114.

[6] D. P. Acharjya and B. K. Tripathy, "Rough Sets on Fuzzy Approximation Space and Application to Distributed Knowledge Systems", International Journal of Artificial Intelligence and Soft Computing, vol. 1, no. 1, **(2008)**, pp. 1-14.

[7] J. A. Garretson, M. I. Rapert, S. Smith and A. Velliquette, "The meaning of quality: Expectations of students in pursuit of an MBA", Journal of Education for Business, **(2004)**, pp. 17-24.

[8] J. Cullen, J. Joyce, T. Hassall and M. Broadbent, "Quality in higher education: From monitoring to management", Quality Assurance in Education: An International Journal, vol. 11, no. 1, **(2002)**, pp. 5-14.

[9] M. Avdjieva and M. Wilson, "Exploring the development of quality in higher education", Managing Service Quality: An International Journal, vol. 12, no. 6, **(2002)**, pp. 372-383.

[10] M. Baig, S. Basharat and Manzil-e-Maqsood, "A maturity model for quality improvement in higher education", Proceedings of First International Conference on Assessing Quality in Higher Education (ICAQHE), **(2006)**, Lahore, Pakistan.

[11] M. Rocki, "Statistical and mathematical aspects of rankings: Lessons from Poland", Higher Education in Europe, vol. 30, no. 2, **(2005)**, pp. 173-181.

[12] S. Marginson, "Australian universities in the global context", Campus Review, **(2006)** March 22nd, pp. 8-9.

[13] Z. Pawlak, "Rough Sets: Theoretical Aspects of Reasoning about Data", Kluwer Academic Publishers, London, **(1991)**.
[14] Z. Pawlak, "Rough Sets", International Journal of Computer Information Science, vol. 11, **(1982)**, pp. 341–356.
[15] Z. Pawlak and A. Skowron, "Rudiments of rough sets", Information Sciences, vol. 177, no. 1, **(2007)**, pp. 3–27.
[16] Z. Pawlak and A. Skowron, "Rough sets: some extensions", Information Sciences, vol. 177, no. 1, **(2007)**, pp. 28–40.
[17] Z. Pawlak and A. Skowron, "Rough sets and Boolean reasoning", Information Sciences, vol. 177, no. 1, **(2007)**, pp. 41–73.

Authors

R. Manjula received her B.E degree in computer science and engineering from University of Vishwesvaraya and Engineering, Bangalore, Karnataka, India in 1992; M.E in software engineering from Anna University, Tamilnadu, India. She is presently working as Associate Professor in the school of computing sciences and engineering, VIT University, Vellore, India. She is a Ph. D candidate of school of computing sciences and Engineering, VIT University, India. Her area of interest includes software process modeling, software metrics, software testing, XML-web services and service oriented architecture.

J. Vaideeswaran is working as a Senior Professor in the school of computing sciences and engineering, VIT University, Vellore, Tamilnadu, India. He received his Ph. D degree in computer engineering from Anna University, Chennai, India in 2000. He has published number of papers in international and national journals of repute. His area of research interests includes coding theory, computer architecture, robotics and software engineering.

D. P. Acharjya received his Ph. D in computer science from Berhampur University, India; M. Tech. in computer science from Utkal University, India; M. Phil. from Berhampur University, India; and M. Sc. from NIT, Rourkela, India. He has been awarded with Gold Medal in M. Sc. At present, he is an Associate Professor in the school of computing sciences and engineering, VIT University, Vellore, India. He has authored many national and international journal papers and four books; Fundamental Approach to Discrete Mathematics, Computer Based on Mathematics, Theory of Computation, Rough Set in Knowledge Representation and Granular Computing – On Some Aspects to his credit. He is associated with many professional bodies CSI, ISTE, IMS, AMTI, ISIAM, OITS, IACSIT, CSTA, IEEE and IAENG. He was founder secretary of OITS Rourkela chapter. His current research interests include rough sets, formal concept analysis, knowledge representation, granular computing, data mining and business intelligence.