Efficient Rule Set Generation using K-Map & Rough Set Theory (RST)

Durgesh Srivastava1, Shalini Batra2, Sweta Bhalothia3

1,2,3BRCM CET, Bahal, Bhiwani, Haryana, INDIA
{dkumar.bit, shwetabhalothia252 }@gmail.com

Abstract
With the enormous growth of data especially large size data sets, mixed types of data, uncertainty, incompleteness, data change, use of background knowledge, etc. the inconsistency of databases is growing many fold. Many a times the information system may contain a certain amount of redundancy that will not assist in any knowledge discovery and may in fact mislead the process. One of the methods which can be used to deal with these issues is the Rough Set Theory (RST), a mathematical tool used to deal with imperfect knowledge and discover pattern hidden in data. RST deals with uncertainty and vagueness, without any requirement of additional information about data, allowing generation of the sets of decision rules from knowledge. The redundant attributes may be eliminated in order to generate the reduct set or to construct the core of the attribute set. This paper proposes a new approach: Karnaugh map for the reduction of attributes and then uses rough set theory to generate rules.

Keywords: Rough Set Theory (RST), K-Map, Feature Selection, Decision Rules.

1. Introduction
In machine learning, classification is considered an instance of supervised learning; learning where training set of correctly identified observations is available. Data classification is the categorization of data for its most effective and efficient use. To classify means identify to which category or set of categories a new observation belongs, based on the training set of data containing observations whose category membership is known.

Rough Set Theory, proposed in 1982 by Zdzislaw Pawlak, considered as one of the first non-statistical approaches in data analysis, chiefly concerned with the classification is also concerned with analysis of imprecise, uncertain or incomplete information and knowledge. Very large data sets, mixed types of data (continuous valued, symbolic data), uncertainty (noisy data), incompleteness (missing, incomplete data), data change, use of background knowledge etc. are real world issues where rough set theory can be useful. Rough sets can handle uncertainty and vagueness, discovering patterns in inconsistent data. Attribute reduction is very important in data classification where irrelevant and redundant attributes are removed from the decision without any classification information loss [1], [7], [8]. This paper proposes new approaches for attribute reduction. Karnaugh map, also known as K-map, introduced by Maurice Karnaugh in 1953 to simplify Boolean algebra expressions. The K-map reduces the need for extensive calculations by taking advantages of humans pattern recognition capability [4], [9], [10].

This paper is organized as follows. Section 2 describes some basic concepts of Rough Set Theory. Section 3 highlights the concepts of K-map. Section 4, gives the introduction of proposed method. Experimental results are described in Section 5. Section 6 gives the conclusion followed by references.

2. Rough Set Theory (RST)
Rough set theory proposed by Z. Pawlak in 1982, is a new mathematical approach to imperfect knowledge having close connections with many other theories. Despite its connections with other theories, the rough set theory may be considered as own independent discipline. Recently rough sets have been exhaustively used in many applications such as Artificial Intelligence and cognitive sciences, especially in machine learning, knowledge discovery, data mining, expert systems, approximate reasoning and pattern recognition [1], [7].

Basic concepts of Rough Set Theory
An information system is a pair S = (U, P) where U is a non-empty finite set of objects called the universe and P is a non-empty finite set of attributes such that a: U → V_a for every a ∈ P. The set V_a is called the value set of a.

A decision system is any information system of the form S = (U, P U {d}), where d (is not element of P) is the decision attribute. The elements of P are called conditional attributes or simply conditions.

Let Q ⊆ P then Q-indiscernibility relation denoted by IND(Q), is defined as:

\[ IND(Q) = \{(x, x') \in U^2 | \forall a \in Q, a(x) = a(x')\} \]

If (x, x') ∈ IND(Q), then objects x and x’ are indiscernible from each other by attributes from Q. The equivalence classes of the Q-indiscernibility relation are denoted as \[\{x\}_Q\].

The discernibility matrix of S is a symmetric n×n matrix with entries c_ij as given as:

\[ c_{ij} = |\{a \in Q | a(x_i) \neq a(x_j)\}| \]

Each entry thus consists of the set of attributes upon which objects x_i and x_j differ. Since discernibility matrix is symmetric and c_ij = 0 (the empty set) for i≠1,...,n, this matrix can be represented using only elements in its lower triangular part, i.e. for 1 ≤ j ≤ n.

A discernibility function (f_i) for an information system S is a Boolean function of m Boolean variables a_1, a_2, ..., a_m (corresponding to the attribute a_1, a_2, ..., a_m) defined as:

\[ f_i(a_1, ..., a_m) = \bigvee \{ c_{ij} | i \leq j \leq n, a_i \neq a_j \} \]

Where \[ c_{ij} = \{a^* | a \in c_{ij}\} \]. The set of all prime implicants of f_i determines the set of all reducts of Q. The discernibility function is a Boolean function in POS form.

Approximations of set
Let X is a subset of U, i.e. X ⊆ U.

IJETI
www.ijeti.com
Lower Approximation: For a given concept, its lower approximation refers to the set of observations that can all be classified into this concept.
\[ Q^-(X) = \{ x | x \in X \} \]
Upper Approximation: For a given concept, its upper approximation refers to the set of observations that can be possibly classified into this concept.
\[ Q^+(X) = \{ x | x \in X \cap X \neq \emptyset \} \]

Once the reducts have been computed, the rules are easily constructed by overlaying the reducts over the originating decision table and reading off the values.

3. Karnaugh-Map (K-Map)
The Karnaugh map, also known as the k-map, a method to simplify Boolean algebra expressions reduces the need for extensive calculations. In Boolean algebra, any Boolean function can be expressed in canonical form using the dual concepts of min-terms and max-terms. Product of Sums (POS) is a conjunction (AND) of max terms. The canonical forms allow for greater analysis into the simplification of Boolean functions. Here K-map is used for the reduction of attributes by simplifying discernibility function which is a Boolean function in POS form. Since discernibility function is a monotone Boolean function (Boolean function without negation), so negative terms in K-map are not considered [9], [10].

Advantages of K-map:
- It takes less space
- It takes less time

Example: Consider the following map. The function plotted is:
\[ Z = f(A, B) = A \bar{B} + AB \]

- The values of the input variables form the rows and columns, which are the logic values of the variables A and B (with one denoting true form and zero denoting false form) form the head of the rows and columns respectively.
- The above map is a one dimensional type which can be used to simplify an expression in two variables.
- There is a two-dimensional map that can be used for up to four variables, and a three-dimensional map for up to six variables.

Using algebraic simplification,
\[ Z = A \bar{B} + AB \]
\[ Z = A (\bar{B} + B) \]
\[ Z = A \]

Variable B becomes redundant due to Boolean Theorem.

Referring to the map above, the two adjacent 1's are grouped together. Through inspection it can be seen that variable B has its true and false form within the group. This eliminates variable B leaving only variable A which only has its true form. The minimized answer therefore is \( Z = A \).

4. Proposed Method
In real world data varies in size and complexity, which is difficult to analyze and also hard to manage from computational view point. The major objectives of Rough Set analysis are to reduce data size and to handle inconsistency in data. The work presented here deals with uncertainty and extract useful information from the database. The basic methodology of the proposed work is defined under the following steps:

1. Representation of the Information System
2. Discernibility Matrices
3. Discernibility Functions
4. Reduction of Attributes using K-Map
5. New Reduct Table
6. Set Approximation
7. Rule Generation

![Flow Diagram of Proposed Method](image)

5. Experimental Result
The proposed work uses Flu Data Set where the data is discretized by using RST and K-map. The data is split into training set and test set. In Flu Data Set we have taken six patients data. The data about six patients is used as training data. The attributes for this data are headache, muscle pain, temperature and flu where headache, muscle pain and temperature are conditional attributes and flu is decision attribute. Samples of decision rules for Flu data set generated from reduct are given below:

**Step1: Taken Flu Data Set**
Here several patients data set is taken with possible flu symptoms (As shown in Table 1). By using k-map & Rough set approach, data is analyzed, redundant data is eliminated, attributes are reduced and set of rules are developed. Table 1 shows the patient’s data set and respective symptoms.

<table>
<thead>
<tr>
<th>Patient</th>
<th>Headache</th>
<th>Muscle pain</th>
<th>Temperature</th>
<th>Flu</th>
</tr>
</thead>
<tbody>
<tr>
<td>p1</td>
<td>No</td>
<td>Yes</td>
<td>High</td>
<td>Yes</td>
</tr>
<tr>
<td>p2</td>
<td>Yes</td>
<td>No</td>
<td>High</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Step 2: Discernibility Matrix

As discussed above discernibility matrix of an information system is a symmetric n×n matrix with entries c_{ij} where each entry consists of the set of attributes upon which objects x_i and x_j differ. Since discernibility matrix is symmetric c_{ii} = ∅ for i = 1, …, n. This matrix can be represented using only elements in its lower triangular part i.e. for 1 ≤ j < i ≤ n. In the Table 1 the given information system has 6 objects, so the discernibility matrix for this system is a 6×6 matrix. The complete discernibility matrix for above information system is shown in Table 2:

<table>
<thead>
<tr>
<th>p3</th>
<th>Yes</th>
<th>Yes</th>
<th>Very High</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>p4</td>
<td>No</td>
<td>Yes</td>
<td>Normal</td>
<td>No</td>
</tr>
<tr>
<td>p5</td>
<td>Yes</td>
<td>No</td>
<td>High</td>
<td>No</td>
</tr>
<tr>
<td>p6</td>
<td>No</td>
<td>Yes</td>
<td>Very High</td>
<td>Yes</td>
</tr>
</tbody>
</table>

In this table, (H, M, T) denotes Headache, Muscle pain and Temperature respectively [11].

Step 3: Discernibility Function

A discernibility function for an information system is a Boolean function of Boolean variables. With every discernibility matrix one can associate a discernibility function. For above discernibility matrix, discernibility function is given below:

\[
\begin{align*}
\text{f}(H,M,T) &= (H+M).T.((H+M).T) \\
&= (H+M+T).(H+M).T \\
&= (H+T).(M+T).H \\
&= (H+M+T).T \\
&= (H+M+T)
\end{align*}
\]

Each row in the above discernibility function corresponds to one column in the discernibility matrix. Each parenthesized tuple is a sum in the Boolean expression and one-letter Boolean variables correspond to attribute names [11].

Step 4: K-map for this function

Since there are 3 variables (H, M, T), so the K-map for this function has 2^3=8 squares. In K-map for POS function each square represents a max-term. Adjacent squares always differ by just one literal. The K-map for a function is specified by putting a ‘0’ in the square corresponding to a max-term in case of POS form. For example, the first sum in the above discernibility function is H+M. Here the value of H and M is 0, the absence of T indicates that the value of T is 1, so enter a 0 in the square which has value 001. Like this we enter the values for the above function [9].

K-map can be minimized with the help of grouping the terms to form single terms. When forming groups of squares, following points need to be considered:

- Every square containing 0 must be considered at least once.

In K-map for POS function each square of a max-term is marked as 1. In K-map for POS form, we mark these squares as 0. The K-map for the above function is shown in Fig. 2.

Step 5: Reduct Table

After grouping the terms, it is checked that which variable is changing its values and which remains same in group and the variable that is changing its value is dropped. So the output of a group is the terms that do not change its values. In the K-map given below (Fig. 2), two groups are formed: one is represented by red color and other is by green color. In the group represented by red color, M and T are changing their values but H has the same value throughout the group, so output of this group is H, same as the output of the second group is T.

So, Reduct = {H, T}

Step 6: Reduct Table

After the simplification of K-map following reduct set {H, T} is obtained which indicates that there is only one reduct set {H, T} in the data table and it is the core. Thus the attribute Muscle pain can be eliminated from the table. The new reduct table is:
Table 2. Reduct Information system of Flu Disease

<table>
<thead>
<tr>
<th>Patient</th>
<th>Headache</th>
<th>Temperature</th>
<th>Flu</th>
</tr>
</thead>
<tbody>
<tr>
<td>p1</td>
<td>No</td>
<td>High</td>
<td>Yes</td>
</tr>
<tr>
<td>p2</td>
<td>Yes</td>
<td>High</td>
<td>Yes</td>
</tr>
<tr>
<td>p3</td>
<td>Yes</td>
<td>Very High</td>
<td>Yes</td>
</tr>
<tr>
<td>p4</td>
<td>No</td>
<td>Normal</td>
<td>No</td>
</tr>
<tr>
<td>p5</td>
<td>Yes</td>
<td>High</td>
<td>No</td>
</tr>
<tr>
<td>p6</td>
<td>No</td>
<td>Very High</td>
<td>Yes</td>
</tr>
</tbody>
</table>

**Step 6: Set Approximation**

The lower and the upper approximations of a set are interior and closure operations in a topology generated by a indiscernibility relation. Indiscernibility Relation is the relation between two objects or more, where all the values are identical in relation to a subset of considered attributes. The indiscernibility relation for given information system is [6, 11]:

\[ \text{IND}(Q) = \{\{p1\}, \{p2,p5\}, \{p3\}, \{p4\}, \{p6\}\} \]

\[ X = \{x: \text{Flu}(x) = \text{yes}\} = \{p1, p2, p3, p6\} \]

\[ Q = \{ \text{Headache}, \text{Temperature} \} \]

Lower Approximation set \( Q^*(X) \) of the patients that are definitely having Flu are identified as:

\[ Q^*(X) = \{p1, p3, p6\} \]

Upper Approximation set \( Q^+(X) \) of the patients that possibly have Flu are identified as:

\[ Q^+(X) = \{p1, p2, p3, p5, p6\} \]

Boundary region is defined as:

\[ Q_N = \{p2, p5\} \]

Boundary Region (BR), the set constituted by elements p2 and p5 (which cannot be classified, since they possess the same characteristics, but with differing conclusions) differ in the decision attribute.

**Step 7: Rule Generation**

Once the reducts have been computed, the rules are easily constructed by overlaying the reducts over the originating decision table and reading off the values. Given the reduct \( \{ \text{Headache}, \text{Temperature} \} \) in new reduct table, the rule read off the first object is “if Headache is no and Temperature is high then Flu is yes” [4].

Table 3. Set of Rules for FLU disease

<table>
<thead>
<tr>
<th>Sl. No.</th>
<th>Rules</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(Headache, No) and (Temperature, High) → (Flu, Yes)</td>
</tr>
<tr>
<td>2</td>
<td>(Headache, Yes) and (Temperature, Very High) → (Flu, Yes)</td>
</tr>
<tr>
<td>3</td>
<td>(Headache, No) and (Temperature, Very High) → (Flu, Yes)</td>
</tr>
<tr>
<td>4</td>
<td>(Headache, no) and (Temperature, Normal) → (Flu, No)</td>
</tr>
</tbody>
</table>

**Conclusion:**

An efficient rule set generation approach has been presented to classify data by using K-map and Rough Set Theory. In this paper, different steps used to generate the rules have been presented. The features have been reduced by using first three steps discernibility matrix, discernibility function and K-map. After reducing attributes and calculating set approximations, extraction of classification rules becomes easier making it easy to automate the task of classifying data. The boundary region is representing uncertainty. The elements of boundary region cannot be classified, since they possess the same characteristics, but with differing conclusions. The experimental results on a number of data sets indicate that the given approach is very satisfactory.

**References:**
4. Prasanta Gogoi, Ranjan Das, B Borah & D K Bhattacharyya: "Efficient Rule Set Generation using Rough Set Theory for Classification of High..."
Dimensional Data”. International Journal of Smart Sensors and Ad Hoc Networks (IJSSAN) ISSN No. 2248-9738 (Print) Volume-1, Issue-2, 2011