



Determination of Optimal Pair of Thresholds in Three-Way Decisions using Objective Functions

Remesh K M¹, Latha R Nair²

¹Division of Computer Engineering, School of Engineering, CUSAT, Kochi, India, remeshkm@hotmail.com

²Division of Computer Engineering, School of Engineering, CUSAT, Kochi, India, latha5074@gmail.com

ABSTRACT

The use of a single threshold for deciding on acceptance or rejection in a probabilistic two-way decision-making process is inappropriate when the available information is uncertain or incomplete. Probabilistic three-way decisions use a pair of thresholds to trisect the universe allowing a more effective and sensible decision-making process. Determination of optimal thresholds in the probabilistic three-way decision-making model minimizes the overall uncertainty of the three regions. This assures the quality of trisections and guides to more valuable and accurate decisions. Quality of trisection can be measured with the use of objective functions. In this paper, the optimal pair of thresholds is determined experimentally by using Shannon entropy and chi-square static as objective functions and are analyzed and compared.

Key words: Rough Set Theory, Probabilistic Rough Sets, Three-way decisions, Shannon entropy, Chi-square statistic.

1. INTRODUCTION

Data gathered from real-world systems may have uncertain or incomplete information. Decision making on real-world data thus becomes complex as it may lead to incorrect decisions or decisions with errors. In such situations, decision making is typically formulated by selecting a threshold as a trade-off between errors and decisions [1][2][11]. The uncertainty aroused due to bisection can be resolved by a probability to represent uncertainty and relay on a threshold to make an appropriate decision as accept or reject. This is known as probabilistic two-way decisions [1][11]. In this case, the decision space is divided into two regions of acceptance and rejection based on a threshold. These regions are also called positive regions and negative regions. A hypothesis is formulated and if the probability of the hypothesis is less than the threshold value, the hypothesis is rejected and the information is in the negative region. If the probability of the hypothesis is greater than the threshold value, the hypothesis is accepted and the information is in the positive region [1][11].

Here, the decisions are not definite. It is a trade-off between incorrect-acceptance and incorrect-rejection due to insufficient information or incorrect data. There are only two choices viz accept with errors or reject with errors [1][11]. Now, the major concern is in minimizing the errors aroused due to incorrect-acceptance and incorrect-rejection to a low level. The problem of incorrect-acceptance and incorrect-rejection can be overwhelmed to a certain extent by the introduction of a third region in the decision space called non-commitment region or boundary region. To trisect the information space, a pair of thresholds (α , β) satisfying the condition $0 \leq \beta < \alpha \leq 1$ is required. If the probability of the hypothesis is below β , decide to the negative region, if the probability of the hypothesis is in between the two thresholds α and β , decide to the boundary region, otherwise, if the probability is higher than the threshold α , then decide to the positive region [1][11].

In this paper, the optimal pair of thresholds were determined using two different objective functions viz Shannon Entropy and Chi-square statistic on a variety of data sets. The size of the original data set was reduced using “Reduct sizereduction” algorithm proposed by the authors [14] and the optimal pair of thresholds for the resultant data set thus obtained was determined using the above mentioned objective functions. The results obtained were compared and analyzed.

The minimal set of attributes that characterize the whole knowledge of an information system is known as the reduct of the information system [14]. All the attributes in a reduct are significant. Removal of attributes having negligible significance in the information system does not adversely affect the classification accuracy and the overall performance of the information system [14]. The “Reduct sizereduction” algorithm removes attributes having negligible significance thereby reducing the number of attributes to characterize the knowledge in the information system.

In this paper, two algorithms using Shannon entropy and Chi-square statistic are presented to determine the optimal pair of thresholds

2. RELATED STUDY

Rough Set Theory is a mathematical tool to deal with vague, imprecise, inconsistent, and uncertain knowledge [3]. The basic concept of Rough Set Theory is the indiscernibility between objects. Two objects are said to be indiscernible if all the knowledge characterized by the two objects are the same. Yan Zhang in his study [21], uses game-theoretic rough sets for the study and analysis of the changes in probabilistic thresholds and its impact on the Gini coefficient of rough set regions. For this purpose, he has formulated a competitive game between regions and adopted an interactive learning mechanism such that effective thresholds can be obtained. Xiaofei Deng and Yiyu Rao in their study [11] proposed an information-theoretic approach by using Shannon Entropy as a measure of uncertainty to determine the optimal pair of thresholds. The problem of determining the optimal pair of thresholds in three-way decisions is formulated as the minimization of Shannon entropy of the three regions. Xiaofei Deng and Yiyu Rao in their study [1] presented a general framework for the determination of a pair of thresholds in three-way decisions as an optimization problem. This study also presented a heuristic algorithm based on a gradient-descent approach for the determination of the optimal pair of thresholds. Nouman Azam, Yan Zhang and Jing Tao Yao in their study [9] use Game-Theoretic Rough Set (GTRS) model to resolve the issues encountered when the rough set-based three-way decision is extended to multiple criteria decision making (MCDM). MCDM uses multiple functions, and the difference in evaluation functions is addressed by implementing a game with multiple evaluation functions as game players and the differences in results are resolved by determining a tradeoff between evaluation functions. Conditions under which different game outcomes could contribute to a game solution is the choice structure for selecting the three types of decision choices. Cong Gao and Yiyu Yao in their study [10] established the chi-square statistic as an objective function to quantify the goodness of a three-way approximation. A large value of the Chi-square statistic shows a strong correlation between the trisection and classification. Nouman Azam and Jing Tao Yao in their study [16] proposed variance-based criteria for the determination of thresholds in three-way decisions. The authors introduced variance-based criteria viz within region variance, between region variance and the ratio of these two variances for determining thresholds in three-way decisions. The authors formulated the problem of finding an optimal pair thresholds as optimization functions.

3. THREE-WAY DECISIONS

Most of the decision-making problems have to deal with incomplete, imprecise, and vague information and leads to decision making under uncertainty. To reduce the cause of uncertainty, a three-way decision-making strategy is adopted over the years, in which a non-commitment decision option is exercised [9]. Thus, the decision-making process leads to

three decision actions viz accept the objects to the Positive region, reject the objects to the Negative region and non-commitment objects to the Boundary region [10]. Many application areas especially medical decision making, psychology, social judgment theory, management sciences, and machine learning has been using three-way decisions over the years [2].

Let U is a finite set of objects and C is a concept such that $C \subseteq U$. Based on the Pawlak rough set model, the three regions in the three-way decision model are defined as [9]:

$$\begin{aligned} POS(C) &= \{x \in U \mid [x] \subseteq C\} \\ NEG(C) &= \{x \in U \mid [x] \subseteq C^c\} \\ BND(C) &= \{x \in U \mid \neg([x] \subseteq C) \wedge \neg([x] \subseteq C^c)\} \end{aligned} \quad (1)$$

Where $[x]$ represents an equivalence class of the object

x and C^c represents the complement of concept C .

The rough set being a powerful tool for dealing with imprecise and vague information is much useful in three-way decisions.

3.1 Probabilistic Three-way Decisions

Suppose U represents a finite set of objects called the universe and C represents the target set such that $C \subseteq U$. For a pair of thresholds (α, β) , such that $0 \leq \beta < \alpha \leq 1$, the concept C is approximated for acceptance, rejection, and non-commitment using the following three-way decision rules:

Rule for Acceptance : If $\Pr(C \mid [x]) \geq \alpha$ accept $x \in C$;

Rule for Rejection : If $\Pr(C \mid [x]) \leq \beta$ reject $x \in C$;

Rule for Non-commitment : If $\beta < \Pr(C \mid [x]) < \alpha$, neither accept or reject $x \in C$.

Accordingly, the positive, negative, and boundary regions are defined as follows:

$POS_{(\alpha, \beta)}(C) = \{x \in U \mid \Pr(C \mid [x]) \geq \alpha\} \Rightarrow \underline{P}_{(\alpha, \beta)}(C)$; the lower approximation of P

$NEG_{(\alpha, \beta)}(C) = \{x \in U \mid \Pr(C \mid [x]) \leq \beta\} \Rightarrow \overline{P}_{(\alpha, \beta)}(C)^c$; the compliment of upper approximation of P

$BND_{(\alpha, \beta)}(C) = \{x \in U \mid \beta < \Pr(C \mid [x]) < \alpha\} \Rightarrow (POS_{(\alpha, \beta)}(C) \cup NEG_{(\alpha, \beta)}(C))^c$ (2)

3.2 Evaluation function

The three fundamental issues concerning the evaluation and designated values in three-way decision theory are [4][9] :

a. Decision measurement

The first issue is to define a set of values each for measuring satisfiability for acceptance and non-satisfiability for rejection. A single set may be used for measuring both satisfiability and non-satisfiability.

b. Evaluation functions

The second issue is the construction and interpretation of evaluation functions for evaluating objects. An evaluation of a set of criteria set apart satisfiability or non-satisfiability of objects. Evaluation functions may be constructed and

interpreted in a sensitive and practically operable manner. Examples of evaluation are cost, risk, errors, benefits, profits, user satisfaction, etc.

c. Decision conditions

The third issue is the determination and interpretation of the two thresholds that trisect the decision regions. These sets of thresholds should reflect a perspective understanding of acceptance and rejection.

4. THRESHOLD DETERMINATION USING SHANNON ENTROPY AS OBJECTIVE FUNCTION

The amount of information contained in a variable is termed as the entropy of that variable. According to Shannon, the information content of anything can be measured in bits. Shannon entropy quantifies the amount of information in a variable. The degree of uncertainty in any probability distribution or a partition can be measured using the Shannon entropy as [1][11][13]:

$$H = -\sum_i (P_i * \log P_i) \tag{3}$$

where P_i is the probability of the distribution or partition i . Shannon entropy can be used to measure the uncertainty level. A pair of thresholds (α, β) , such that $0 \leq \beta < 0.5 \leq \alpha \leq 1$, trisects the data space into three partitions:

$$\pi_{(\alpha, \beta)}(C) = (POS_{(\alpha, \beta)}(C), NEG_{(\alpha, \beta)}(C), BND_{(\alpha, \beta)}(C)).$$

The uncertainty concerning each region $\pi(C) = \{C, C^c\}$ is [11] :

$$H(\pi_C | POS_{(\alpha, \beta)}(C)) = - \Pr(C | POS_{(\alpha, \beta)}(C)) \log \Pr(C | POS_{(\alpha, \beta)}(C)) - \Pr(C^c | POS_{(\alpha, \beta)}(C)) \log \Pr(C^c | POS_{(\alpha, \beta)}(C)).$$

$$H(\pi_C | NEG_{(\alpha, \beta)}(C)) = - \Pr(C | NEG_{(\alpha, \beta)}(C)) \log \Pr(C | NEG_{(\alpha, \beta)}(C)) - \Pr(C^c | NEG_{(\alpha, \beta)}(C)) \log \Pr(C^c | NEG_{(\alpha, \beta)}(C)).$$

$$H(\pi_C | BND_{(\alpha, \beta)}(C)) = - \Pr(C | BND_{(\alpha, \beta)}(C)) \log \Pr(C | BND_{(\alpha, \beta)}(C)) - \Pr(C^c | BND_{(\alpha, \beta)}(C)) \log \Pr(C^c | BND_{(\alpha, \beta)}(C)). \tag{4}$$

where $\Pr(C | POS_{(\alpha, \beta)}(C))$ is the conditional probability of an object in C such that the object is in the probabilistic positive region and so on. Conditional probabilities are evaluated as follows [11]:

$$\Pr(C | POS_{(\alpha, \beta)}(C)) = \frac{|C \cap POS_{(\alpha, \beta)}(C)|}{|POS_{(\alpha, \beta)}(C)|}$$

$$\Pr(C^c | POS_{(\alpha, \beta)}(C)) = 1 - \Pr(C | POS_{(\alpha, \beta)}(C))$$

$$\Pr(C | NEG_{(\alpha, \beta)}(C)) = \frac{|C \cap NEG_{(\alpha, \beta)}(C)|}{|NEG_{(\alpha, \beta)}(C)|}$$

$$\Pr(C^c | NEG_{(\alpha, \beta)}(C)) = 1 - \Pr(C | NEG_{(\alpha, \beta)}(C))$$

$$\Pr(C | BND_{(\alpha, \beta)}(C)) = \frac{|C \cap BND_{(\alpha, \beta)}(C)|}{|BND_{(\alpha, \beta)}(C)|}$$

$$\Pr(C^c | BND_{(\alpha, \beta)}(C)) = 1 - \Pr(C | BND_{(\alpha, \beta)}(C)) \tag{5}$$

The overall entropy of the three partitions is [11] :

$$H(\pi_C | \pi_{(\alpha, \beta)}(C)) = \Pr(POS_{(\alpha, \beta)}(C))H(\pi_C | POS_{(\alpha, \beta)}(C)) + \Pr(NEG_{(\alpha, \beta)}(C))H(\pi_C | NEG_{(\alpha, \beta)}(C)) + \Pr(BND_{(\alpha, \beta)}(C))H(\pi_C | BND_{(\alpha, \beta)}(C)).$$

which is the expected overall entropy and is called the conditional entropy of $\pi_{(C)}$ given $\pi_{(\alpha, \beta)}(C)$

The probabilities of the three regions can be computed as [11]:

$$\Pr(POS_{(\alpha, \beta)}(C)) = \frac{|POS_{(\alpha, \beta)}(C)|}{|U|}$$

$$\Pr(NEG_{(\alpha, \beta)}(C)) = \frac{|NEG_{(\alpha, \beta)}(C)|}{|U|}$$

$$\Pr(BND_{(\alpha, \beta)}(C)) = \frac{|BND_{(\alpha, \beta)}(C)|}{|U|} \tag{6}$$

The Shannon entropy of a partition $\pi_C = \{C, C^c\}$ is [11]:

$$H(\pi_C) = -\frac{|C|}{|U|} \log \frac{|C|}{|U|} - \frac{|C^c|}{|U|} \log \frac{|C^c|}{|U|}, \text{ where } C \subseteq U \text{ and } C \neq \emptyset. \tag{7}$$

Upon trisecting the partition using the thresholds (α, β) , the reduction of uncertainty may be computed as:

$H(\pi_C) - H(\pi_C | \pi_{(\alpha, \beta)}(C))$. Here, $H(\pi_C)$ is independent of (α, β) . Thus, to obtain the maximum uncertainty reduction, minimize $H(\pi_C | \pi_{(\alpha, \beta)}(C))$. Thus, the optimal pair of thresholds can be found out by the optimization problem:

$$(\alpha^*, \beta^*) = \arg \min_{(\alpha, \beta)} H(\pi_C | \pi_{(\alpha, \beta)}(C)). \tag{8}$$

Where (α^*, β^*) are the optimal pair of thresholds.

4.1 A Method for Optimal Threshold Determination using Shannon Entropy

Assume that each equivalence class of a training example is a partition and want to predict the class of an arbitrary object. If the dimensions of all the classes are the same then the prediction to which the object belongs is the most uncertain and hence the information entropy attains the maximum value ($\log n$). If there is only one class, the information entropy attains the minimum value (0). Thus, the problem to find out the pair of optimal thresholds is to identify a pair of thresholds such that the information entropy is minimal.

An algorithm to determine the optimal pair of thresholds using Shannon entropy as an objective function is proposed here. The approach followed in this method is to determine Shannon entropy for various pairs of thresholds and identify a pair with the minimum entropy as the optimal pair of thresholds. The values of thresholds are varied by systematically applying stepwise increment/decrement to the

existing values within the limit $0 \leq \beta < \alpha \leq 1$. The proposed algorithm is given below:

Algorithm Optimal_Thresholds_by_Entropy_Minimization

Input: A set of training examples U

The target concept such that $C \subseteq U$

Step-size by which α and β are to alter

Output: A set of optimal thresholds (α, β)

Begin :

1. Determine equivalent classes of U
2. For each equivalent class i, repeat steps 3 to 4
3. Determine the probability of the class :

$$\Pr(X_i) = \frac{\text{Number of elements in the class}}{\text{Total number of elements in the data set}}$$
4. Determine the conditional probabilities of the class :

$$\Pr(C | X_i) = \frac{\text{Number of elements in the class that belongs to the concept}}{\text{Total number of elements in the class}}$$
5. Sort $\Pr(X_i), \Pr(C | X_i)$ and $\Pr(C^c | X_i)$ in descending order according to $\Pr(C | X_i)$
6. Initialize $\alpha = 1, \beta = 0, \text{min} = 1$;
7. **Do**
 8. Compute probabilities and conditional probabilities of the three regions
 9. Compute Shannon entropy of the three regions
 10. Compute overall uncertainty of the three regions
 11. If (overall uncertainty < min) then

$$\text{min} = \text{overall uncertainty}$$

$$\alpha' = \alpha$$

$$\beta' = \beta$$
 - end
 12. Modify $\alpha = \alpha - \text{step-size}, \beta = \beta + \text{step-size}$
13. **While** ($(\beta < 0.5)$ and $(\alpha \geq 0.5)$)
14. Return (α, β)

End.

4.2 Experimental Results

Experiments were conducted using the algorithm on various data sets downloaded from the UCI Machine Learning Data Repository. Determination of an optimal pair of thresholds was done for each data set in two different ways say Case 1 and Case 2.

1. In Case 1, the above algorithm is directly applied to each of the data set and obtained an optimal pair of thresholds.
2. In Case 2, the attribute reduction algorithm “Reduct sizedreduction” proposed by the same authors [14] was used to reduce the number of attributes in the data set, and a revised data set is obtained with less number of attributes wherever possible. Then, the above “Optimal_Thresholds_by_Entropy_Minimization” algorithm is applied to the revised data set and obtained the optimal pair of thresholds.

Even though the algorithm outputs optimal thresholds in one run by identifying the minimum of the entropy for different pairs of α and β , entropy for each pair is tabulated in Table 1 and Table 2 for analysis purposes.

Case 1:

The data set “mammography” downloaded from UCI Machine Learning Data Repository is used and applied to the algorithm. There are 961 samples and 6 attributes in the data set. The conditional entropy for various threshold values, α and β are tabulated in Table 1.

Table 1: Conditional entropy

$\beta \rightarrow$ $\alpha \downarrow$	0.4	0.3	0.2	0.1	0.0
1.0	0.2066	0.1796	0.1626	0.1576	0.1542
0.9	0.1993	0.1754	0.1546	0.1587	0.1621
0.8	0.2127	0.1851	0.1678	0.1613	0.1642
0.7	0.2305	0.2009	0.1825	0.1709	0.1783
0.6	0.2562	0.2239	0.2040	0.1826	0.1991
0.5	0.2723	0.2647	0.2439	0.2134	0.2383

Case 2:

In this case, the algorithm “Reduct sizedreduction” is applied to the data set “mammography” and the possible attributes were removed from the data set. There were only 4 attributes in the resultant data set as against 6. This data set is then applied to the “Optimal_Thresholds_by_Entropy_Minimization” to determine conditional entropy. The conditional entropy thus obtained for various threshold values α and β are tabulated in Table 2.

Table 2: Conditional entropy after attribute reduction

$\beta \rightarrow$ $\alpha \downarrow$	0.4	0.3	0.2	0.1	0.0
1.0	0.6206	0.6243	0.6698	0.6921	0.9003
0.9	0.5585	0.5573	0.5566	0.5622	0.6171
0.8	0.5556	0.5541	0.5507	0.5553	0.6016
0.7	0.5556	0.5541	0.5507	0.5553	0.6016
0.6	0.5680	0.5663	0.5607	0.5643	0.6012
0.5	0.5745	0.5727	0.5661	0.5693	0.6017

4.3 Observations

A pair of thresholds with minimum conditional entropy is the optimal pair of thresholds. Here in Case 1, the minimum conditional entropy is 0.1546 for the thresholds $\alpha = 0.9$ and $\beta = 0.2$, and these pairs of thresholds are the optimal pair of thresholds in Case 1. The positive region ranges from 0.9 to 1, the negative region is from 0 to 0.2 and the boundary is between 0.3 to 0.8.

In Case 2, there are two instances in the table 2, with minimum conditional entropy 0.5507 for thresholds $\alpha = 0.7$ and 0.8 and $\beta = 0.2$. The pair of thresholds with $\alpha = 0.7$ and $\beta = 0.2$ being the affirmative one, can be selected as optimal

thresholds. Thus, the positive region is from 0.7 to 1, the negative region is from 0 to 0.2 and the boundary is only between 0.3 to 0.6. The size of the boundary region is reduced in turn reducing the deferment area as compared with Case 1.

5. THRESHOLD DETERMINATION USING CHI-SQUARE STATISTIC AS AN OBJECTIVE FUNCTION

Chi-square statistics can also be used as an objective function for the determination of a pair of threshold values. A contingency table of three-way decisions shows connections between the actual classification $\{C, C^c\}$ and the three-way approximations

$\pi_{(\alpha, \beta)}(C) = (POS_{(\alpha, \beta)}(C), NEG_{(\alpha, \beta)}(C), BND_{(\alpha, \beta)}(C))$ as shown in table 3 below [10]:

Table 3: Contingency Table of Three-way Decisions

	$POS_{(\alpha, \beta)}(C)$	$BND_{(\alpha, \beta)}(C)$	$NEG_{(\alpha, \beta)}(C)$	Total
C	n_{CP}	n_{CB}	n_{CN}	n_C
C^c	n_{C^cP}	n_{C^cB}	n_{C^cN}	n_{C^c}
Total	n_P	n_B	n_N	n

n_{CP}, n_{CB} etc are the numbers of objects in the corresponding region and category of a class and n_C, n_B etc are the marginal totals.

Chi-square (χ^2) statistic can be used to test the independence of two variables. The (χ^2) statistic can be computed from a contingency table as:

$$\chi^2 = \sum \frac{(observed - expected)^2}{expected} \tag{9}$$

where “observed” is the actual observed number in the contingency table and “expected” is the expected number. n_C and n_P can be computed as:

$$\begin{aligned} n_C &= n_{CP} + n_{CB} + n_{CN} \text{ and} \\ n_P &= n_{CP} + n_{C^cP} \end{aligned} \tag{10}$$

The expected value of the positive region = Probability of the concept * Probability of the positive region of the concept * total number of objects.

$$= Pr(C) * Pr(POS_{(\alpha, \beta)}(C)) * |U| = \left(\frac{n_P}{n} \frac{n_C}{n}\right) n = \frac{n_C n_P}{n} \tag{11}$$

Similarly, the expected values for the negative and boundary regions may be calculated.

The observed value of the positive region is n_{CP} obtained from the contingency table. The divergence under the independence assumption between the observed and the

expected values can be measured as $(n_{CP} - n_C n_P / n)^2$. If the observed value is close or equal to the expected value, then, $(n_{CP} - n_C n_P / n)^2$ is close to or even equal to 0. That is $(n_{CP} - n_C n_P / n)^2 / (n_C n_P / n)$ is also close to or even equal to 0. Thus, if the value of the chi-square statistics is large then it is presumed that the dependency is strong. Statistically, significant chi-square statistics mean that $\{C, C^c\}$ and $\pi_{(\alpha, \beta)}(C)$ are correlated or dependent and a larger chi-square statistic indicates a stronger correlation [10].

To find optimal thresholds, it is required to find a pair of thresholds that provide the strongest correlation or obtain a pair thresholds with the maximum chi-square statistic (χ^2). Thus, the optimal pair of thresholds can be found out by the optimization problem [10]:

$$(\alpha^*, \beta^*) = \arg \max_{(\alpha, \beta)} \chi^2_{(\alpha, \beta)} \tag{12}$$

Where (α^*, β^*) are the optimal pair of thresholds.

5.1 An Algorithm for Optimal Pair of Threshold Determination Using Chi-square Statistic

An algorithm using the Chi-square statistic as an objective function to determine the optimal pair of thresholds is proposed in this paper. Chi-square statistic for various pairs of thresholds is determined in this method and the pair with the maximum chi-square statistic is chosen as the optimal pair of thresholds. The values of thresholds were varied by systematically applying stepwise increment/decrement to the existing values within the limit $0 \leq \beta < \alpha \leq 1$. The proposed algorithm is given below:

Algorithm Optimal_Thresholds_by_Chi-square_Statistic

Input: A set of training examples U
 The target concept such that $C \subseteq U$
 Step-size by which α and β are to alter

Output: A set of optimal thresholds (α, β)

Begin :

1. Determine equivalent classes of U
2. For each equivalent class i, repeat steps 3 to 4
3. Determine the probability of the class :

$$Pr(X_i) = \frac{\text{Number of elements in the class}}{\text{Total number of elements in the data set}}$$
4. Determine the conditional probabilities of the class :

$$Pr(C | X_i) = \frac{\text{Number of elements in the class that belong to the concept}}{\text{Total number of elements in the class}}$$

$$Pr(C^c | X_i) = 1 - Pr(C | X_i)$$
5. Sort $Pr(X_i), Pr(C | X_i)$ and $Pr(C^c | X_i)$ in descending order according to $Pr(C | X_i)$
6. Initialize $\alpha = 1, \beta = 0, \max = 0;$
7. **Do**

8. Construct the contingency table for the observed numbers
9. Construct the contingency table for the expected values
10. Compute the chi-square statistic

$$\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

11. If ($\chi^2 > \max$) then

$$\max = \chi^2$$

$$\alpha' = \alpha$$

$$\beta' = \beta$$

end

12. Modify $\alpha = \alpha - \text{step-size}$, $\beta = \beta + \text{step-size}$

13. While (($\beta < 0.5$) and ($\alpha \geq 0.5$))

14. Return (α' , β')

End.

5.2 Experimental Results

Experiments using the algorithm described above were conducted on various data sets downloaded from the UCI Machine Learning Data Repository. Determining the optimal pair of thresholds was done for each data set as described in the case of determining of optimal pair of thresholds using Shannon entropy.

The data set “mammography” is used for the analysis purpose here as in the previous case. The Chi-statistic for various threshold values α and β have tabulated in Table 4 as in Case 1 above.

Table 4: Chi-square statistic

$\beta \rightarrow$ $\alpha \downarrow$	0.0	0.1	0.2	0.3	0.4
0.5	782.96	782.96	783.08	778.34	776.43
0.6	807.46	807.46	808.43	804.86	793.08
0.7	810.54	810.54	812.56	810.89	802.60
0.8	808.42	808.42	811.20	811.02	805.40
0.9	807.21	807.21	810.21	810.44	805.55
1.0	807.21	807.21	810.21	810.44	805.55

As in Case 2 above, the chi-square values for various pairs of thresholds are tabulated in Table 5 using the above algorithm. Less significant attributes of the data set “mammography” are removed as above and given to the algorithm to determine chi-square values.

Table 5: Chi-square statistic after attribute reduction

$\beta \rightarrow$ $\alpha \downarrow$	0.0	0.1	0.2	0.3	0.4
0.5	457.98	488.20	492.04	490.99	489.90
0.6	460.74	497.50	502.09	502.57	501.58
0.7	453.45	502.47	508.53	512.46	511.77
0.8	453.45	502.47	508.53	512.46	511.77
0.9	433.61	492.59	499.97	507.49	507.19
1.0	91.43	348.08	376.83	440.92	446.23

5.3 Observations

Optimal thresholds are determined as the pair of thresholds with maximum chi-square statistics. In Table 4, the chi-square value 812.56 is the highest among others. Thus, the pair $\alpha = 0.7$ and $\beta = 0.2$ are the optimal pair of thresholds. The positive region ranges from 0.7 to 1, the negative region is from 0 to 0.2 and the boundary is between 0.3 to 0.6.

In the case of Table 5, highest chi-square value is obtained in two cases with $\alpha = 0.7$, $\beta = 0.2$ and $\alpha = 0.8$, $\beta = 0.2$. As in the previous case $\alpha = 0.7$, $\beta = 0.2$ is selected as optimal thresholds. The positive region ranges from 0.7 to 1, the negative region is from 0 to 0.2 and the boundary is between 0.2 to 0.7. Thus, the positive region is from 0.7 to 1, the negative region is from 0 to 0.2 and the boundary is only between 0.3 to 0.6. From these, it is evident that the sizes of the boundary regions remain the same in both cases.

6. CLASSIFICATION ACCURACY

The Classification accuracy of the positive and negative regions was compared for all the pairs of threshold values. The best classification accuracy was found for the pair of thresholds $\alpha = 0.7$, $\beta = 0.2$ and $\alpha = 0.8$, $\beta = 0.2$, mitigating the selection as the optimal pair of thresholds. The incorrectly classified instances for the above thresholds were also found to be the lowest when compared with all the other pairs of thresholds.

7. CONCLUSION

In a decision making strategy that involves insufficient, uncertain information, both the Shannon entropy and Chi-square statistic can be effectively used as objective functions to determine the optimal pair of thresholds. Optimal pair of thresholds is computed as an optimization problem in both the algorithms proposed in this paper and can be used as a general framework for the computation of optimal thresholds. The information-theoretic rough set model using Shannon entropy adopts entropy minimization and a pair with minimum value is the optimal pair. Whereas, a high value of chi-square statistic indicates a high correlation and thus the pair with the maximum chi-square statistic is the optimal pair. The classification accuracy of the positive and negative regions obtained through the two methods reiterates the obtained pair of thresholds as the optimal pair.

REFERENCES

1. Xiaofei Deng, Yiyu Yao. **A Multifaceted Analysis of Probabilistic Three-way Decisions**, *Fundamenta Informaticae* 132 (2014) 291–313 291, DOI 10.3233/FI-2014-1045, IOS press
2. Hafeez Ur Rehman, Nouman Azam, JingTao Yao, Alfredo Benso. **A three-way approach for protein function**, *PLOS ONE*, DOI:10.1371/journal.pone.0171702 February 24, 2017

3. Z Pawlak. **Rough Set Theory and its Applications**, *Journal of Telecommunications and Information Technology*, March 2002
4. Yiyu Yao. **An Outline of a Theory of Three-way Decisions**, Springer, Heidelberg (2012) RSCTC 2012. LNCS (LNAI), vol. 7413, pp. 1-17.
https://doi.org/10.1007/978-3-642-32115-3_1
5. Qinghua Zhang, Qin Xie, Guoyin Wang a. **A survey on rough set theory and its applications**, *ScienceDirect, CAAI Transactions on Intelligence Technology 1 (2016)* 323e333, Doi:10.1016/j.trit.2016.11.001
6. Yuhua Qian, Jiye Liang, Yiyu Yao, Chuangyin Dang. **MGRS: A multi-granulation rough set**, *Information Sciences* 180 (2010) 949–970 ELSEVIER
7. Yuhua Qian, Shunyong Li, Jiye Liang, Zhongzhi Shi, Feng Wang. **Pessimistic rough set based decisions: A multigranulation fusion strategy**, *Information Sciences* 264 (2014), ELSEVIER , 196–210
8. Nouman Azam, JingTao Yao. **Analyzing uncertainties of probabilistic rough set regions with game-theoretic rough sets**, *International Journal of Approximate Reasoning* 55 (2014) 142–155, ELSEVIER
9. Nouman Azam, Yan Zhang, JingTao Yao. **Evaluation Functions and Decision Conditions of Three-way Decisions with Game-theoretic Rough Sets**, *European Journal of Operational Research*, DOI:10.1016/j.ejor.2016.12.048
10. Cong Gao, Yiyu Yao. **Determining Thresholds in Three-Way Decisions with Chi-Square Statistic**, *Springer International Publishing AG 2016V*. Flores et al. (Eds.): IJCRS 2016, LNAI 9920, pp. 272–281, 2016. DOI: 10.1007/978-3-319-47160-0 25
11. Xiaofei Deng, Yiyu Yao. **An Information-Theoretic Interpretation of Thresholds in Probabilistic Rough Sets**, *RSKT 2012*, LNAI 7414, pp. 369–378, 2012. © Springer-Verlag Berlin Heidelberg 2012
12. Wentao Li, Witold Pedrycz, Xiaoping Xue, Xiaoyan Zhang, Bingjiao Fan, Binghan Long. **Information measure of absolute and relative quantification in double-quantitative decision theoretic rough set model**, *Journal of Engineering*, 2018, Vol. 2018 Iss. 16, pp. 1436-1441
13. M.S.R. Naidu, P. Rajesh Kumar, K. Chiranjeevi. **Shannon and Fuzzy entropy based evolutionary image thresholding for image segmentation**, *Alexandria Engineering Journal (2017)*, doi:10.1016/j.aej.2017.05.024
14. Remesh K M, Latha R Nair. **A Novel Method to Reduce Attributes from Reducts**, *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887, Volume 7 Issue II, Feb 2019- Available at www.ijraset.com
15. Yiyu Yao, **Decision-Theoretic Rough Set Models**, *Rough Set and Knowledge Technology, Second International Conference, RSKT 2007*, LNAI 4481, pp.1-12,2007
16. Nouman Azam, Jing Tao Yao. **Variance Based Determination of Three-Way Decisions Using Probabilistic Rough Sets**, *Springer International Publishing AG 2016 V*. Flores et al. (Eds.): IJCRS 2016, LNAI 9920, pp. 209–218, 2016. DOI: 10.1007/978-3-319-47160-0 19
17. Yiyu Yao and Bing Zhou. **Naive Bayesian Rough Sets**, *RSKT 2010*, LNAI 6401, pp. 719-726, 2010.
https://doi.org/10.1007/978-3-642-16248-0_97
18. Wentao Li, Witold Pedrycz, Xiaoping Xue, Xiaoyan Zhang, Bingjiao Fan, Binghan Long. **Information measure of absolute and relative quantification in double-quantitative decision-theoretic rough set model**, *Journal of Engineering*, 2018, Vol. 2018 Iss. 16, pp. 1436-1441, eISSN 2051-3305, doi: 10.1049/joe.2018.8315
19. Anne-Laure Boulesteix. **Maximally Selected Chi-Square Statistics and Binary Splits of Nominal Variables**, *Biometrical Journal* 48 (2006)5.838-848, DOI:10.1002/bimj.200510191.
20. Yan Zhang. **Optimizing Gini Coefficient of Probabilistic Rough Set Regions Using Game-Theoretic Rough Sets**, *26th IEEE Canadian Conference of Electrical and Computer Engineering*, 978-1-4799-0033-6/13