High Range Resolution Radar Signal Classification
A Partitioned Rough Set Approach

Dale E. Nelson
Target Recognition Branch
Sensors Directorate, Air Force Research Laboratory
Wright-Patterson AFB, OH 45433 USA

Janusz A. Starzyk
Department of Computer and electrical Engineering
Ohio University, Russ College of Engineering & Tech.,
Athens, OH 45701 USA

Abstract– **In Automatic Target Recognition (ATR) systems there are advantages to developing classifiers based on a portion of the signal. A partitioning technique is introduced in this paper that allows Rough Set Theory to be applied to real-world size problems. Rough Set Theory (RST) is an emerging concept for determining features and then classifiers from a training data set. RST guarantees that once the data has been labeled all possible classifiers (based on that labeling) will be generated. There will be multiple classifiers for each signal partition and multiple partitions for each signal. Classifiers based on a single reduct (classifier) or one partition do not perform well enough to be useful. We fuse all the reducts from all the partitions into one classifier. This fusion of partitioned reducts yields a synergistic result that produces a classifier with a high probability of declaration and good probability of correct classification.**

I. INTRODUCTION

Classification of High Range Resolution (HRR) radar signals is difficult. A typical HRR signal contains 128 range bins with values between 0-255 representing the signal strength. A 3-D object is now being represented by a 1-D signal. This dimensionality reduction introduces ambiguities. In addition, extreme signal variability makes the problem difficult. Because there is no comparable signal that a human has experience classifying, human intuition is of a little help. Therefore, a computerized learning system is required.

Rough set theory is the mathematical foundation for developing a classifier [1-3]. Each HRR range bin is called an attribute in rough set theory and the target class associated with that signal is called the decision attribute. Rough sets provide the mechanism to find the minimal set of attributes required to classify all the training signals. This minimal set of attributes is called a reduct and contains the same knowledge (ability to classify all the training signals correctly) as the original set of attributes in a given information system. Therefore reducts can be used to obtained different classifiers. Rough sets require the data in the range bins to be labeled. Once this labeling has occurred rough set theory guarantees that all possible classifiers will be found! We chose to use a binary labeling based on entropy. This scheme reduces sensitivity to noise and signal registration. Information entropy is used to select the range bins that are most useful in classification and reduce computational time for determining reducts.

Until recently, rough set theory has not been applied to many classification problems because real-world problems are too large [4-5]. The determination of minimal reducts (minimal classifiers) has been proven to be N-P hard. We have developed a method of reducing the time for finding sub-optimum reducts to O(n^2) making it a useful process for finding classifiers in real-world problems. In addition, we have developed a way to fuse results from all reducts to improve classifier performance.

Fusing the results of the reducts for each partition and fusing the reducts for all the partitions improves classifier performance as it was demonstrated on high range resolution radar signal classification problem. On the training set using one partition the probability of correct classification (Pcc) was 89% and the probability of declaration (Pdec) was 93%. Fusing reducts from all partitions the Pcc was 100% and Pdec was 100%. On the training set one would expect 100% performance on both of these parameters. On the test set Pcc for one partition was 79% and Pdec was 90%. When all reducts were fused, Pcc was 92% and Pdec was 99%!

II. ROUGH SET THEORY

It is not the purpose of this paper to be a tutorial of rough set theory. An introduction to rough set theory may be found in [3]. However, some basic concepts need to be introduced. With the binary labeling used the set of all labeled training signals forms a decision table consisting of 1s and 0s. Each row corresponds to a given target type. In many cases it is possible to use a subset of the entire signal to distinguish among different target classes. For example, it may be possible to use range bins 1 through 20 and be able to uniquely classify each signal in the training set. If this subset of range bins cannot be further reduced without loosing its ability to classify the training set, then it is called a **reduct**. This term comes from the idea that we have reduced the size of the table without reducing the information contained in it (i.e.; the ability to uniquely classify all the signals). There may be no reducts (we must use all the range bins) or there may be many reducts. It should be noted that a reduct may not contain another reduct. That is, it must be minimal.

With this preface we now introduce the mathematical formalism. We review basic definitions of rough set theory related to selection of the set of attributes for the purpose of classifying a given set of objects. Discernibility function is formally defined and an alternative characterization of reducts is given which is easier to manipulate for algorithmic purposes. For a full development of this area see [5].

Consider the information system (U, A), where $U = \{x_1,...,x_n\}$ is a nonempty finite set called the **universe**, and $A = \{a_1,...,a_m\}$ is a nonempty set. The elements of *A*, called **attributes** (in our case range bins), are functions

$$a_i : U \to V_i$$

where V_i is called the value set of a_i. In a practical rough set system V_i is a discrete and finite set of values. In the case of a binary labeling used in this work V_i ={0, 1} . The **discernibility matrix** of *A* is the $n \times n$ matrix with i, j^{th} entry

$$c_{ij} = \{a \in A : a(x_i) \neq a(x_j)\}.$$

So an element c_{ij} of a discernibility matrix contains all attributes that differentiate between two given objects x_i and x_j. Let $B \subseteq A$, and let $P(A)$ be the power set of *A*. The **Boolean-valued function** χ_B is

$$\chi_B : P(A) \to \{0,1\}$$
$$: C \mapsto \begin{cases} 1 & \text{when } B \cap C \neq \varnothing \\ 0 & \text{when } B \cap C = \varnothing \end{cases}$$

Let $S_\chi = \{\chi_B : B \in P(A)\}$. Define the binary operator $\wedge$, called conjunction, by

$$\wedge : S_\chi \times S_\chi \to S_\chi$$
$$: (\chi_B, \chi_C) \mapsto \chi_B \wedge \chi_C$$

where

$$\chi_B \wedge \chi_C : P(A) \to \{0,1\}$$
$$: D \mapsto \chi_B(D)\chi_C(D)$$

The associativity property

$$(\chi_B \wedge \chi_C) \wedge \chi_D = \chi_B \wedge (\chi_C \wedge \chi_D)$$

allows us to drop the parenthesis without any possibility of confusion; moreover we can now define $\wedge$ for any finite collection of functions $\{\chi_{B_i}\}_{i=1}^{p}$ by recursion

$$\mathop{\wedge}_{i=1,...,p} \chi_{B_i} = \Big(\mathop{\wedge}_{i=1,...,p-1} \chi_{B_i} \Big) \wedge \chi_{B_p}$$

The **discernibility function** of the information system is

$$f_A : P(A) \to \{0,1\}$$
$$: C \mapsto \left(\mathop{\wedge}_{\substack{1 \le i,j \le n \\ \chi_{c_{ij}} \neq \bar{0}}} \chi_{c_{ij}} \right)(C)$$

where "$\bar{0}$" is the constant function

$$\bar{0} : P(A) \to \{0,1\}$$
$$: C \mapsto 0$$

If f_A is an empty conjunction we define f_A to be the constant zero function. This is an uninteresting case and we assume throughout that f_A is not an empty conjunction.

The condition $\chi_{c_{ij}} \neq \bar{0}$ used in the definition of the discernibility function is equivalent to the condition that $c_{ij} \neq \varnothing$ since

$$\chi_{c_{ij}} \neq \bar{0} \Leftrightarrow \chi_{c_{ij}}(A) = 1 \Leftrightarrow c_{ij} \cap A \neq \varnothing \Leftrightarrow \exists a_k \in c_{ij} \Leftrightarrow c_{ij} \neq \varnothing$$

Using the fact the discernibility matrix is symmetric and that $c_{ii} = \varnothing$ it follows the discernibility function simplifies to $f_A = \mathop{\wedge}_{\substack{1 \le i < j \le n \\ c_{ij} \neq \varnothing}} \chi_{c_{ij}}$. We also know [5] $f_A(A) = 1$.

Let $B \subseteq A$. The ***B-indiscerniability*** relation is

$$Ind(B) = \{(x, y) \in U \times U : (\forall a \in B)(a(x) = a(y))\}$$

The ***B-discernibility*** relation is the complement of $Ind(B)$ in $U \times U$,

$$Dis(B) = U \times U - Ind(B).$$

The following lemma is an immediate consequence of the definition.

Lemma. Let $B \subseteq A$. *Then*

$$Dis(B) = Dis\left(\bigcup_{a \in B} \{a\} \right) = \bigcup_{a \in B} Dis(\{a\}).$$

Consequently, if $a, b \in B$ and $Dis(\{a\}) = Dis(\{b\})$ then

$$Dis(B) = Dis(B - \{a\}) = Dis(B - \{b\}). \qquad \square$$

Essential for the information system are the reducts that describe knowledge represented in this system. A set $B \subseteq A$ is a **discern** in *A* if $Ind(B) = Ind(A)$. A discern is called a **reduct** if $(\forall a \in B)\ Ind(B - \{a\}) \supset Ind(B)$, where "$\supset$" denotes a proper subset relation. The set of all reducts of *A* is denoted $Red(A)$. The reduct generation procedure developed in [6] is based on the expansion of the discernibility function into a disjunction of its prime implicants by applying the absorption or multiplication laws. This procedure is not sufficiently efficient to allow us to use it with real-world size problems. The **core** of the information system is defined as a set $P \subseteq A$ such that

$$P = \bigcap_{B \in Red(A)} B$$

and a set S is a **shell** if

$$\exists B \in Red(A) \quad P \subset S \subset B$$

Let $B \subset A$. A **focused information system** (U, B) represents local properties of the information system. A **focused reduct** F is a reduct of the focused information system, so we have Ind(F)=Ind(B). A focused reduct in general is not a reduct of the original system as it may not differentiate all objects and in general we have $Ind(A) \subseteq Ind(B)$. The **power information system** is defined as a set of all focused information systems.

$$P(U, A) = \{(U, B) : B \in 2^A\}$$

In other words the power information system of a given information system (U, A) is a set of information systems defined on the power set of A. The power information system is more robust than the original information system and can extract useful knowledge from incomplete or corrupted data. We define **a covered** information system as

$$C(U,A)=\{(U,B): B\in C\subseteq 2^{A} \wedge A\subseteq \bigcup_{B\in C} B\}$$

In order to reduce computational cost, focused reducts will be chosen from a covered information system. In general, a covered information system is redundant, which means that $card(A) < \sum_{i} card(A_i)$.

Conjecture. A covered information system yields a combined classification performance of focused reducts exceeding performance of the reducts of the original information system. In addition, the obtained classification is more robust to signal distortion and can work with partially determined signals. □

III. PARTITIONING OF A SIGNAL

In our HRR classification system we have found that using a Haar wavelet transform on the original signal yields a more powerful feature set from which to build a classifier. The disadvantage is that we now must find reducts from a possible 1024 range bins. This is not possible to do in a reasonable time. Therefore we use an entropy measure to select the range bins that have the most information theoretic value in classifying the training signals. Using our algorithms, 50 range bins is a practical limit for an 800 MHz desktop computer.

If we use the entire original signal and its wavelet coefficients we are able to consider only 50 out of 1024 range bins or less than 5% of the range bins. Table I shows the lackluster performance of a classifier built this way. If we partition the signal in two pieces, build a classifier for each partition and fuse the results, our classifier is now based on twice as many range bins. Continuing this reasoning, if we make eight classifiers and fuse their results we will be using 400 range bins and fusing the results of each of these classifiers. We are thus considering more range bins with potentially more information for our classifier.

The next question is how to partition the signal. The first method that comes to mind is to use a block partitioning as illustrated in Fig. 1. This method looks at each portion of the signal in isolation. Normally the ends of a signal contain noise and are not useful. The classifiers based on theses areas tend not to have good performance. However, the fusion equation takes this into account. This partitioning method allows classifiers to be generated which can focus on the more important aspects of the signal. Another advantage of this method is that should a portion of the signal be obscured for any reason, there are still multiple classifier that do no depend on that portion of the signal thus still allowing classification.

Another possible method would be to use an interleaved selection illustrated in Fig 2. The easiest way to explain this method is to describe dividing the signal into two parts. The first partition would consist of all the odd numbered range bins and the other partition would consist of all the even numbered range bins. This concept is easily extended to four and eight partitions. This partitioning scheme reduces the effects of registration of the signals. That is, if the first range bin of a test signal does not match the first range bin of the training signal we might not get classification at all. However, with interleaved partitioning there would be a classifier which would classify. Theoretically this method would allow a misregistration of up to eight range bins.

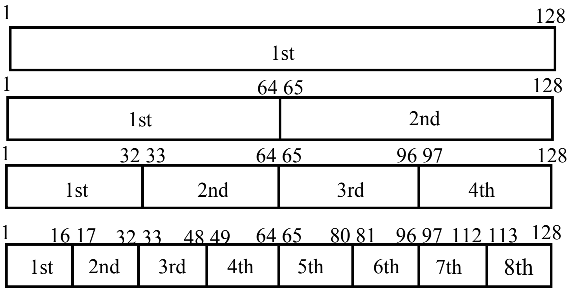

Fig. 1. Block Partitioning

The conjectures put forth regarding misregistration and obscuration have not been tested as of this writing.

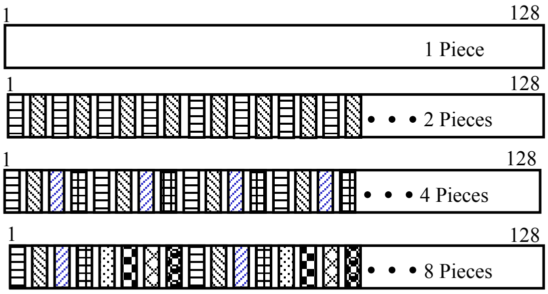

Fig. 2. Interleave Partitioning

IV. FUSION OF MULTIPLE CLASSIFIERS

Having many advisors is valuable in the world (especially if you have a measure of how reliable they are). Therefore, we theorized that by fusing multiple classifiers from multiply partition signals we would produce a better and more robust classifier.

In is not the purpose of this paper to completely describe the fusion process. However, to help you understand the results the fusion equation is presented here. Each classifier votes as to which target it believes the signal belongs to (Pcc_i). All votes for each class are fused and given a score (W_t). The class with the highest weight is the class selected. If no weight is greater than a user set threshold (0.50) then the signal is unclassified. The unclassified signals affect the probability of declaration (Pdec).

$$W_t = 1 - \frac{Pcc_{\max} + \left[\sum_{i=1}^{n} \left(1 - Pcc_i\right) \right] \left(1 - Pcc_{\max}\right)}{\sum_{i=1}^{n} \frac{1}{1 - Pcc_i + \varepsilon}}$$

This equation guarantees the W_t will be at least as large as the greatest *Pcc,* ($Pcc_{\max}$). Further W_t is limited to the range of 0 to 1.

V. RESULTS

Table I shows the test classification results from each partition type and the fusion of differing levels of partitioned classifiers. The column titled Sel. Indicated what partitioning method and which partition is being evaluated.. Entries followed by st, nd, or th indicated that block partitioning was used and that it is the first , second, etc block. Entries where there is no suffix indicated that interleaved partitioning was used. In this cast the entry would mean that the first, second, third, etc element of each block was combined to make a partition. Previously it was mentioned that the first part of a signal and the last part may contain noise an therefore would not be able to perform well. This is confirmed by the zeros in 8-1st, 8-2nd, 8-7th, and 8-8th. Other partitions perform very well (2-1st). The final results support the conjecture that properly fusing many classifiers will result in a better, more robust classifier. The robustness improvement is indicated by the Probability of Declaration being almost 1. This indicates there are very few signals unclassified. Even with this high classification rate, the classifier is still very accurate.

We are currently exploring ways to tradeoff the probability of declaration to achieve a higher probability of correct classification. Additional experiments will be performed to verify the conjectures regarding limited sensitivity to registration and obscuration. Because of the binary labeling scheme, we also believe that this classifier may be resistant to signal noise as well.

VI. REFERENCES

[1] A. Nakamura and G. Jian-Miang, “A modal logic for similarity-based data analysis”, Hiroshima Univ. Technical. Report., 1988.

[2] Z. Pawlak, “Information systems - theoretical foundations”, *Information Systems*, Vol. 6, pp.205-218, 1981.

[3] Z. Pawlak, Rough Sets - Theoretical Aspects of Reasoning About Data, *Kluwer Academic Publ.*, 1991.

[4] J. A. Starzyk, D. E. Nelson, and K. Sturtz, "Reduct Generation in Information Systems", Bulletin of International Rough Set Society, 1999, 3(1/2).

[5] J.A. Starzyk, D.E. Nelson, and K. Sturtz, " A Mathematical Foundation for Improved Reduct Generation in Information Systems", Journal of Knowledge and Information Systems, March 2000.

[6] A. Skowron, C. Rausser, “The Discernibility Matrices and Functions in Information Systems, *Fundamenta Informaticae*, 15(2), pp.331-362, 1991.

TABLE I. TEST CLASSIFICATION RESULTS

Div.	Sel.	Pcc	Pdec	Pcc	Pdec	Pcc	Pdec	Pcc	Pdec
1	1	0.7922	0.90381						
2	1	0.79095	0.79726						
2	2	0.79381	0.75705	0.81229	0.94424				
2	1st	0.91704	0.73964						
2	2nd	0.43967	0.7954	0.79694	0.94942	0.88116	0.992		
4	1	0.72369	0.80949						
4	2	0.74261	0.85531						
4	3	0.7574	0.83313						
4	4	0.63639	0.75311	0.82266	0.9971				
4	1st	0	0						
4	2nd	0.88051	0.80846						
4	3rd	0.65241	0.76575						
4	4th	0	0	0.83239	0.94983	0.882	0.999		
8	1	0.63722	0.81882						
8	2	0.58947	0.78317						
8	3	0.59995	0.75808						
8	4	0.51747	0.65858						
8	5	0.49859	0.80618						
8	6	0.50673	0.72367						
8	7	0.51172	0.7073						
8	8	0.53631	0.64511	0.77047	0.99979				
8	1st	0	0						
8	2nd	0	0						
8	3rd	0.66444	0.68138						
8	4th	0.81829	0.51451						
8	5th	0.70485	0.41439						
8	6th	0.37494	0.82546						
8	7th	0	0						
8	8th	0	0	0.75351	0.97388	0.84906	0.999	0.923	0.999