Presented at the IAPR TC7 Workshop on
Multisource Data Integration in Remote Sensing
June 14-15, 1990

A Method for Classification of Multisource Data using Interval-Valued Probabilities and Its Application to HIRIS Data

H. Kim and P.H. Swain

School of Electrical Engineering
and
Laboratory for Applications of Remote Sensing
Purdue University
West Lafayette, IN 47907, U.S.A.
Tel: (317)494-0212, FAX: (317)494-0776
E-mail address: swain@ecn.purdue.edu

ABSTRACT

A method of classifying multisource data in remote sensing is presented. The proposed method considers each data source as an information source providing a body of evidence, represents statistical evidence by interval-valued probabilities, and uses Dempster's rule to integrate information based on multiple data sources.

The method is applied to the problems of ground-cover classification of multispectral data combined with digital terrain data such as elevation, slope, and aspect. Then this method is applied to simulated 201-band High Resolution Imaging Spectrometer (HIRIS) data by dividing the dimensionally huge data source into smaller and more manageable pieces based on the global statistical correlation information. It produces higher classification accuracy than the Maximum Likelihood (ML) classification method when the Hughes phenomenon is apparent.

1 INTRODUCTION

The importance of utilizing multisource data in ground-cover classification lies in the fact that it is generally correct to assume that improvements in terms of classification accuracy can be achieved at the expense of additional independent features provided by separate sensors. However, it should be recognized that information and knowledge from most available data sources in the real world are neither certain nor complete. We refer to such a body of uncertain, incomplete, and sometimes inconsistent information as "evidential information."

The objective of the current research is to develop a mathematical framework within which various applications can be made with multisource data in remote sensing and geographic information systems. The methodology described in this paper has evolved from "evidential reasoning," where each data source is considered as providing a body of evidence with a certain degree of belief. The degrees of belief based on the body of evidence are represented by "interval-valued (IV) probabilities" rather than by conventional point-valued probabilities so that uncertainty can be embedded in the measures.

There are three fundamental problems in the multisource data analysis based on IV probabilities: (1) how to represent bodies of evidence by IV probabilities, (2) how to combine IV probabilities to give an overall assessment of the combined body of evidence, and (3) how to make decisions based on IV probabilities.

The paper describes a formal method of representing statistical evidence by IV probabilities based on the Likelihood Principle. In order to integrate information obtained from individual data sources, the method presented in the paper uses Dempster's rule for combining multiple bodies of evidence. Although IV probabilities together with Dempster's rule provide an innovative means for the representation and combination of evidential information, they make the decision process rather complicated. We need more intelligent strategies for making decisions. This paper also focuses on the development of decision rules over IV probabilities.

2 AXIOMATIC DEFINITION OF IV PROBABILITY

Interval-valued probabilities can be thought as a generalization of ordinary point-valued probabilities. The endpoints of IV probabilities are called the "upper probability" and the "lower probability."

There have been various works introducing the concepts of IV probabilities in the areas of philosophy of science and statistics [1][2][3][4]. Although the mathematical rationales behind those approaches are different, there are some properties of IV probabilities which are commonly required. The axiomatic approach to IV probabilities is based on those common properties, so that it can avoid conceptual ambiguities.

DEFINITION [5] Suppose Θ is a finite set of exhaustive and mutually exclusive events. Let β denote a Boolean algebra of the subsets of Θ. The IV probability $[\mathcal{L}, \mathcal{U}]$ is defined by the set-theoretic functions:

$$\mathcal{L} : \beta \rightarrow [0, 1] \tag{2.1}$$

$$\mathcal{U} : \beta \rightarrow [0, 1] \tag{2.2}$$

satisfying the following properties:

I) $\mathcal{U}(A) \geq \mathcal{L}(A) \geq 0$ for any $A \in \beta$ (2.3)

II) $\mathcal{U}(\Theta) = \mathcal{L}(\Theta) = 1$ (2.4)

III) $\mathcal{U}(A) + \mathcal{L}(\overline{A}) = 1$ for any $A \in \beta$ (2.5)

IV) For any $A, B \in \beta$ and $A \cap B = \emptyset$,

$$\mathcal{L}(A)+\mathcal{L}(B) \leq \mathcal{L}(A \cup B) \leq \mathcal{L}(A)+\mathcal{U}(B) \leq \mathcal{U}(A \cup B) \leq \mathcal{U}(A)+\mathcal{U}(B) \tag{2.6}$$

Given a system of IV probabilities over β, the actual probability measure, P(A), of any subset A of Θ is assumed to lie in the interval $[\mathcal{L}, \mathcal{U}]$ such that

$$\mathcal{L}(A) \leq P(A) \leq \mathcal{U}(A) \tag{2.7}$$

The degree of uncertainty about the actual probability of A is represented by the width, $\mathcal{U}(A)-\mathcal{L}(A)$, of the interval. In particular, $\mathcal{U}(A)=\mathcal{L}(A)=P(A)$ when there is complete knowledge of the probability of A. In this case, the IV probability becomes an ordinary additive probability. And $\mathcal{L}(A)+\mathcal{L}(\overline{A})=0$ when there is absolutely no knowledge of the probability of A.

The basic probability assignment m defined in Shafer's mathematical theory of evidence[6] has the following relations with the IV probabilities:

$$\mathcal{L}(A) = \sum_{B \subseteq A} m(B) \tag{2.8}$$

$$m(A) = \sum_{B \subseteq A} (-1)^{|A-B|} \mathcal{L}(B) \quad \text{for all } A \subset \Theta \tag{2.9}$$

$$\mathcal{U}(A) = \sum_{B \cap A \neq \emptyset} m(B) \tag{2.10}$$

3 REPRESENTATION OF STATISTICAL EVIDENCE BY IV PROBABILITY

When a body of evidence is based on the outcomes of statistical experiments known to be governed by any probability model, it is called "statistical evidence." One of the basic problems for any theory of IV probabilities is how to represent a given body of statistical evidence by IV probabilities.

DEFINITION [6] An upper probability function $\mathcal{U}$ is said to be "consonant" if its focal elements are nested, i.e., if for $A_i \subseteq \Theta$ $(i=1,\ldots,r)$ such that $m(A_i) > 0$ for all i and $\sum_{i=1}^{r} m(A_i)=1$, $A_i \subset A_j$ for any $i < j$, where m is the basic probability assignment of $\mathcal{U}$.

Suppose the observed data in a statistical experiment are governed by a probability model $\{p_\theta : \theta \in \Theta\}$, where p_θ is a conditional probability density function on a sample space X given θ. Our intuitive feeling is that an observation $x \in X$ seems to more likely belong to those elements of Θ which assign the greater chance to x.

Based on the above intuition along with the consonance assumption of the upper probability function, Shafer[6] proposed the linear plausibility function defined as:

$$\mathcal{U}(A|x) = \frac{\max_{\theta' \in A} p_{\theta'}(x)}{\max_{\theta \in \Theta} p_\theta(x)} \quad \text{for all } A \in \beta \text{ and } A \neq \emptyset \quad (3.1)$$

The corresponding lower probability function is given as:

$$\mathcal{L}(A|x) = 1 - \frac{\max_{\theta' \in \bar{A}} p_{\theta'}(x)}{\max_{\theta \in \Theta} p_\theta(x)} \quad \text{for all } A \in \beta \qquad (3.2)$$

In particular, when the set A is singleton, say $\{\theta'\}$, the function in Eq.(3.1) gives the relative likelihood of θ' to the most likely element in Θ.

4 DEMPSTER'S RULE FOR COMBINING IV PROBABILITIES

Dempster's rule is a generalized scheme of Bayesian inference to aggregate bodies of evidence provided by multiple information sources. Let m_1 and m_2 be the basic probability assignments associated respec-tively with the belief functions $\mathcal{Bel}_1$ and $\mathcal{Bel}_2$ which are inferred from two entirely distinct bodies of evidence E_1 and E_2. For all A_i, B_j, and $X_k \subset \Theta$, Dempster's rule (or Dempster's orthogonal sum) gives a new belief function denoted by

$$\mathcal{Bel} = \mathcal{Bel}_1 \oplus \mathcal{Bel}_2 \qquad (4.1)$$

The basic probability assignment associated with the new belief function is defined as:

$$m(X_k) = (1-\mathcal{k})^{-1} \sum_{A_i \cap B_j = X_k} m_1(A_i) \cdot m_2(B_j)$$

$$\text{for any } X_k \neq \emptyset \qquad (4.2)$$

where $\mathcal{k}$ is the measure of conflict between $\mathcal{Bel}_1$ and $\mathcal{Bel}_2$ defined as:

$$\mathcal{k} = \sum_{A_i \cap B_j = \emptyset} m_1(A_i) \cdot m_2(B_j) \qquad (4.3)$$

Dempster's rule computes the basic probability of X_k, $m(X_k)$, from the product of $m_1(A_i)$ and $m_2(B_j)$ by considering all A_i and B_j whose intersection is X_k. Once m is computed for every $X_k \subset \Theta$, the belief function is obtained by the sum of m's committed to X_k and its subsets. The denominator $(1-\mathcal{k})$ normalizes the result to compensate for the measure committed to the empty set so that the total probability mass has measure one. Consequently, Dempster's rule discards the conflict between E_1 and E_2 and carries their consensus to the new belief function.

Dempster's rule is both commutative and associative. Therefore, the order or grouping of evidence in combination does not affect the result, and a sequence of information sources can be combined either sequentially or pairwise.

5 DECISION RULES FOR IV PROBABILITIES [7]

Consider a classification problem where an arbitrary pattern $x \in X$ from an unknown class is assigned to one of n classes in Θ. Let $\lambda(\theta_i|\theta_j)$ be a measure of the "loss" incurred when the decision θ_i is made and the true pattern class is in fact θ_j, where i, j = 1, ..., n. Also, let $\hat{\theta}(x)$ denote a decision rule that tells which class to choose for every pattern x. We define the "upper expected loss" and the "lower expected loss" of making a decision $\hat{\theta}(x)=\theta_i$ as:

$$\ell_i^*(x) = \sum_{j=1}^{n} \lambda(\theta_i|\theta_j)\ \mathcal{U}_x(\theta_j) \qquad (5.1)$$

$$\ell_{*i}(x) = \sum_{j=1}^{n} \lambda(\theta_i|\theta_j)\ \mathcal{L}_x(\theta_j) \qquad (5.2)$$

where $\mathcal{U}_x$ and $\mathcal{L}_x$ are respectively the upper and the lower probabilities for x being actually from θ_j.

The "Bayes-like rule" is the one which minimizes both the upper and the lower expected losses, i.e.,

$$\hat{\theta}(x)=\theta_i \text{ if } \ell_i^*(x) \le \ell_j^*(x) \text{ and } \ell_{*i}(x) \le \ell_{*j}(x)$$

$$\text{for } j=1,\dots,n \qquad (5.3)$$

A problem with the above decision rule is that there does not always exist θ which satisfies the condition in Eq.(5.3), which can lead to ambiguity. In such an ambiguous situation, one may withhold the decision and wait for a new piece of information. Otherwise, the ambiguity may be resolved by resorting to the following rule, so-called "minimum average expected loss rule":

$$\hat{\theta}(x)=\theta_i \text{ if } \frac{\ell^*_i(x)+\ell_{*i}(x)}{2} \le \frac{\ell^*_j(x)+\ell_{*j}(x)}{2} \quad \text{for } j=1,\dots,n \qquad (5.4)$$

As an alternative to the Bayes-like rule, there are two other rules by which a decision is made according to individual measures of the interval, that is, either the upper expected loss or the lower expected loss:

(A) minimum upper expected loss rule:

$$\hat{\theta}(x)=\theta_i \text{ if } \ell^*_i(x) \le \ell^*_j(x) \quad \text{for } j=1,\dots,n \qquad (5.5)$$

(B) minimum lower expected loss rule:

$$\hat{\theta}(x)=\theta_i \text{ if } \ell_{*i}(x) \le \ell_{*j}(x) \quad \text{for } j=1,\dots,n \qquad (5.6)$$

Although the above two rules always produce decisions and there is no ambiguous situation in making a decision according to the rules, they do not utilize all of the information represented by the IV probabilities. The performance of these rules are compared with the minimum average expected loss rule in the experiments by applying them to problems of ground-cover classification based on remotely sensed and geographic data.

6 EXPERIMENTAL RESULTS

The experiments have been performed over two different image data sets. In the experiments, the classification accuracies of the multisource data (MSD) classification based on the proposed method were compared with those of Maximum Likelihood (ML) classifications based on the stacked vector approach.

Table 1 describes the set of data sources for the first experiment. The image in this data set consists of 256 lines by 256 columns and covers a forestry site around the Anderson River area of British Columbia, Canada. Source 1 is 11-band Airborne Multispectral Scanner (A/B MSS) data. Sources 2 and 3 are Synthetic Aperture Radar (SAR) imagery in Shallow mode and Steep mode, respectively. Sources 4 through 6 provide digital terrain data.

In this experiment, 6 classes were defined as listed in Table 2, and 100 pixels per class were used for training data, which is between 4% and 8% of the total pixels of the classes in the test fields. The training samples are uniformly distributed over the test

Table 1. Anderson River Data Set.

Source Index	Data Type	Spectral Region	Input Channel	Spectral Band(μm)
1	A/B MSS	Visible	1	.38 - .42
			2	.42 - .45
			3	.45 - .50
			4	.50 - .55
			5	.55 - .60
			6	.60 - .65
			7	.65 - .69
		Near IR	8	.70 - .79
			9	.80 - .89
			10	.92 - 1.10
		Thermal	11	8 - 14
2	SAR	Shallow		XHV
				XHH
				LHV
				LHH
3	SAR	Steep		XHV
				XHH
				LHV
				LHH
4	Topo-graphic	Elevation		
5		Aspect		
6		Slope		

Table 2. Information Classes for Test of Anderson River Data Set.

Class Index	Cover Types	Tree Sizes	No. of Pixels	% of Total
1	Douglas Fir 2 (df2)	31 - 40*m*	2246	21.72
2	Douglas Fir 3 (df3)	21 - 30*m*	1501	14.52
3	DF+Other Species 2 (df+os2)	31 - 40*m*	1352	13.08
4	DF+Lodgepole Pine 2 (df+lp2)	21 - 30*m*	1589	15.37
5	Hemlock+Cedar (hc)	31 - 40*m*	1587	15.35
6	Forest Clearings (fc)		2064	19.96
Total			10339	100.0

fields so that they may be considered as good representatives of the total samples.

We have observed that some of the classes defined in Table 2 cannot be assumed to be normally distributed in the topographic data. Thus it was decided to adopt a nonparametric approach such as the "Nearest Neighbor" (NN) method [8] in computing probability measures while the optical and radar data sources were assumed to have Gaussian probability density functions.

First, the ML classification based on the stacked vector approach was carried out for various sets of the data sources, adding one source at a time to the A/B MSS data in the order Elevation, SAR-Shallow, SAR-Steep, Aspect, and Slope. Then the MSD classification based on the proposed method was performed using different decision rules. Tables 3 and 4 compare the results for the training samples and the test samples, respectively. Even though the compounded data in the ML classification were treated as having Gaussian distributions, the ML and the MSD methods produced similar results for the training samples. This is not surprising because the ML method uses conventional additive probabilities assuming that the knowledge concerning the actual unknown probabilities is complete, which is reasonable as far as the training samples are concerned.

Table 3. Results of Classifications over Training Samples of Anderson River Data.

	Decision Rule	Sources					
		1	1, 4	1, 2, 4	1 - 4	1 - 5	1 - 6
ML		82.50	88.67	91.67	92.00	92.83	93.50
MSD	MUEL	-	89.83	92.00	92.50	93.17	94.33
	MLEL	-	88.67	91.17	91.33	92.33	93.67
	MAEL	-	88.50	91.00	91.67	91.67	93.50

Comparing the performance of the three decision rules, the minimum upper expected loss (MUEL) rule was superior to the other rules, the minimum lower expected loss (MLEL) rule and the minimum average expected loss (MAEL) rule. It is not known in general which rule is the best. Further research is needed to determine whether guidelines can be devised for selection of the decision rule.

Table 4. Results of Classifications over Test Samples of Anderson River Data.

	Decision Rule	Sources					
		1	1, 4	1, 2, 4	1 - 4	1 - 5	1 - 6
ML		74.16	77.77	79.13	78.93	79.80	81.01
MSD	MUEL	-	80.60	82.39	82.69	83.02	84.54
	MLEL	-	78.45	81.42	81.67	82.24	83.65
	MAEL	-	78.21	80.95	82.05	81.88	83.16

In the second experiment, the proposed method was applied to the classification of HIRIS data by decomposing the data into smaller pieces, i.e., subsets of spectral bands. The data set used in this experiment is simulated HIRIS data obtained by RSSIM [9]. RSSIM is a simulation tool for the study of multispectral remotely sensed images and associated system parameters. It creates realistic multispectral images based on detailed models of the ground surface, the atmosphere, and the sensor. Table 5 provides a description of the simulated HIRIS data set.

Figure 1 is a visual representation of the global statistical correlation coefficient matrix of the data. The image is produced by converting the absolute values of coefficients to gray values between 0 and 255. Based on the correlation image, the 201 bands were divided into 3 groups in such a way that intra-correlation is maximized and inter-correlation is minimized. Table 6 describes the multisource data set after division. Note that the spectral regions of the input channels in Source 3 coincide with the water absorption bands.

With 225 training samples (a third of the total samples) for each class, the ML classification and the MSD classification using the minimum upper expected loss rule were performed over the total samples for various sets of the sources, and the results are listed in Table 7.

The results of the ML method apparently show effects of the Hughes phenomenon;

the accuracy goes down as the dimensionality of the source increases while the number of training samples is fixed. In particular, the accuracy decreases by a considerable amount when all features are used. Presence of the Hughes phenomenon causes the ML method to be particularly sensitive to a bad source, Source 3 in this case. Meanwhile, the proposed MSD classification method always shows robust performance and gives consistent results.

Table 5. Description of Simulated HIRIS Data Set.

Name	Finney County Data Set
Data Type	201-band HIRIS data simulated by RSSIM
Spectral Region	0.4 - 2.4*m m*
Spectral Resolution	0.01*m m*
Image Size	45 lines × 45 columns (2025 samples)
Information Classes	Winter Wheat, Summer Fallow, Unknown

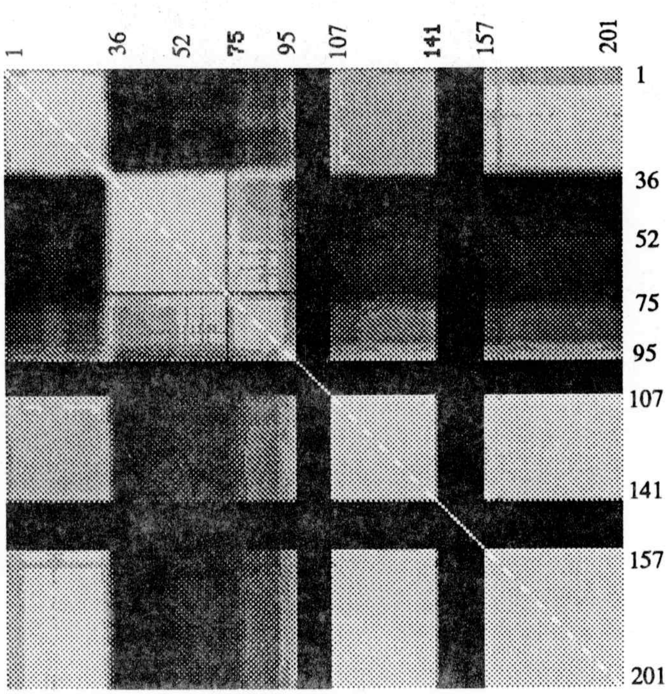

Figure 1. Global Statistical Correlation Coefficient Image of Simulated HIRIS Data.

Table 6. Divided Sources of HIRIS Data Set.

Source Index	Input Channels	No. of Features
Source 1	1- 35, 107 - 141, 157 - 201	115
Source 2	36 - 95	60
Source 3	96 - 106 (1.35 - 1.45μm) 142 - 156 (1.81 - 1.95μm)	26

Table 7. Results of Classifications over Test Samples of Simulated HIRIS Data Set.

	Sources				
	S1	S2	S3	S1, S2	All
ML	75.75	75.60	45.83	74.56	65.14
MSD	-	-	-	77.83	77.63

6 CONCLUSIONS

In this paper we have investigated how interval-valued probabilities can be used to represent and aggregate evidential information obtained from various data sources. Overall concepts of interval-valued probabilities have been employed to develop a new method of classifying multisource data in remote sensing and geographic information systems. The experiments demonstrate the ability of our method to capture uncertain information based on inexact and incomplete multiple bodies of evidence. The basic strategy of this method is to decompose the relatively large size of evidence into smaller, more manageable pieces, to assess plausibilities and supports based on each piece, and to combine the assessments by Dempster's rule. In this scheme, we are able to overcome the difficulty of precisely estimating statistical parameters, and to integrate statistical information as much as possible.

ACKNOWLEDGEMENTS

This research is partially supported by the National Aeronautics and Space Administration under Contract No. NAGW-925.

The SAR/MSS Anderson River data set was acquired, processed and loaned to Purdue University by the Canadian Center

for Remote Sensing, Department of Energy, Mines and Resources, of the Government of Canada.

REFERENCES

[1] B.O. Koopman, "The axioms and algebra of intuitive probability", *The Annals of Mathematics*, vol.**41**, pp269-292, 1940.

[2] I.J. Good, "Subjective probability as the measure of a non-measurable set", in E.Nagel, Suppes and Tarski (eds.), *Logic, Methodology and the Philosophy of Science*, pp319-329, Stanford University Press, 1962.

[3] C.A.B. Smith, "Consistency in statistical inference and decision", *J. Roy Stat. Soc., Ser.B*, vol.**23**, pp1-25, 1961.

[4] A.P. Dempster, "Upper and lower probabilities induced by a multivalued mapping", *Ann. Math. Stat.*, vol.**38**, pp325-339, 1967.

[5] P. Suppes, "The measurement of be-lief", *J. Roy. Stat. Soc., Ser.B*, vol.**36**, pp160-191, 1974.

[6] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, 1976.

[7] H. Kim, "A Method of Classification for Multisource Data in Remote Sensing based on Interval-Valued Probabilities", Ph.D. Thesis, School of Electrical Engineering, Purdue University, West Lafayette, IN 47907 U.S.A., August 1990.

[8] Fukunaga, K., *Intro. to Statistical Pattern Recognition*, Academic Press, 1972.

[9] Kerekes, J.P., and D.A. Landgrebe, "Modeling, Simulation, and Analysis of Optical Remote Sensing Systems", TR-EE 89-49, School of Electrical Engineering, Purdue University, West Lafayette, IN U.S.A., 1989.