061489

COMBINING MULTISPECTRAL AND ANCILLARY DATA IN REMOTE SENSING USING INTERVAL-VALUED PROBABILITIES

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)494-1742, FAX: (317)494-6440

ABSTRACT

This paper presents a method for classifying multisource data in remote sensing and geographic information systems using interval-valued probabilities. In this method, each data source is considered as an information source which provides a body of statistical evidence. In order to integrate information obtained from multiple data sources, the method adopts Dempster's rule for combining multiple bodies of evidence. Preliminary experiments have been undertaken to illustrate the use of the method in a supervised ground-cover classification on multispectral data combined with digital elevation data. They demonstrate the ability of the method in capturing information provided by inexact and incomplete evidence when there are not enough training samples to estimate statistical parameters.

1 INTRODUCTION

Since quantitative approaches began to be applicable to remote sensing data analysis by developments of the digital computer and sensor systems in 1960s, information concerning the surface of the earth and its environment has been largely extracted from the multispectral data obtained by a single sensor. In recent years, as remote sensing and other data acquisition technologies have advanced, there has been a trend towards exploiting remotely sensed multispectral data in conjunction with other ancillary data in geographic information systems for the purpose of extracting more reliable information from multi-attribute data bases [1-3].

Most traditional classification methods for the analysis of multispectral data have been formulated based on the multivariate statistical theory. Ancillary data in geographic information systems, however, present a couple of problems when combined with purely spectral data in an automated classification. Firstly, while it is often reasonable to use the multivariate Gaussian distribution to represent multispectral data alone, this parametric model may not be applicable to geographic or topographic data. Secondly, various data sources are in general not equally reliable [4]. These problems have been the motivation for the development of classification techniques where disparate data sources are assessed separately, and individual assessments are combined by some means.

The primary objective of our research is to develop a general, computer-based method of classification for multisource data in remote sensing and geographic information systems. In the method, the body of evidence provided by each data source is represented by interval-valued (rather than point-valued) probabilities, which is a generalization of the ordinary additive probabilities, so that uncertainty can be included as a measure. In order to aggregate the information from multiple data sources and to propagate the uncertainty throughout the combination of information, this method requires a function for combining non-additive probabilities.

There are two basic problems in multisource data analysis using interval-valued probabilities (IVP): (1) how to represent bodies of evidence by IVP, and (2) how to combine IVP to give an overall assessment of multiple bodies of evidence. In this paper, after introducing the axiomatic definition of IVP, we will describe how to construct IVP for a body of statistical evidence based on the Likelihood Principle. Then we will examine Dempster's rule for combining multiple bodies of evidence in the sense of the desirable properties which agree with human intuition.. For the purpose of demonstrating the concepts, this approach will be applied to the problem of ground-cover classification on multi-spectral data in conjunction with digital elevation data.

2 AXIOMATIC DEFINITION OF IVP

Interval-valued probabilities are, in general, a more adequate scheme to express one's state of knowledge in the sense of handling uncertain and/or incomplete evidential information. They can be thought as a generalization of ordinary probability, with the lower and upper extremes of the interval corresponding to an event being bounds for the unknown probability of the event.

There have been various works introducing the concepts of IVP in the areas of philosophy of science and statistics. For example, Koopman [5] derives the upper and lower probabilities from intuition. Smith [6] induces an interval-valued probability from betting odds. Good [7] considers an interval-valued probability system on the analogy of the outer and inner measures of a nonmeasurable set. And Dempster [8] formulates a system of upper and lower probabilities induced by a multivalued mapping. Although the mathematical rationales behind the above approaches are different, there are some properties of IVP which are commonly required. The axiomatic approach to IVP is based on those common properties, so that it can avoid conceptual ambiguities.

DEFINITION [9] Suppose Θ is a finite set of exhaustive and mutually exclusive events. Let β denote a Boolean algebra over the subsets of Θ. The interval-valued probability $[P_*, P^*]$ is defined by the set-theoretic functions:

lower probability function $P_* : \beta \rightarrow [0, 1]$ (2.1)

upper probability function $P^* : \beta \rightarrow [0, 1]$ (2.2)

satisfying the following properties:

I) $P^*(A) \geq P_*(A) \geq 0$ for all $A \in \beta$ (2.3)

II) $P^*(\Theta) = P_*(\Theta) = 1$ (2.4)

III) $P^*(\Phi) = P_*(\Phi) = 0$ (2.5)

IV) $P^*(A) + P_*(\bar{A}) = 1$ for any $A \in \beta$ (2.6)

V) For any $A, B \in \beta$ and $A \cap B = \Phi$,

$$P_*(A)+P_*(B) \leq P_*(A \cup B) \leq P_*(A)+P^*(B) \leq P^*(A \cup B) \leq P^*(A)+P^*(B) \quad (2.7)$$

Given a system of IVP over β, the actual probability measure, $P(A)$, of any subset A of Θ is assumed to lie in the interval $[P_*, P^*]$ such that

$$P_*(A) \leq P(A) \leq P^*(A) \quad (2.8)$$

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

3 REPRESENTATION OF IVP

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 IVP is how to represent a given body of statistical evidence as IVP. According to Eq. (2.6), specifying either P^* or P_* is enough for constructing IVP.

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 χ given θ, and consider them as providing a body of statistical evidence for β.

Our intuitive feeling is that an observation $x \in \chi$ seems to more likely belong to those elements of Θ which assign the greater chance to x. In other words, x makes $\theta \in \Theta$ more plausible than $\theta' \in \Theta$ whenever $p_\theta(x) > p_{\theta'}(x)$. Thus, IVP for statistical evidence may be determined by the likelihood functions of observed data.

Based on the above principle, Shafer [10] proposed the following equation:

$$P^*(\{\theta\}|x) = C \cdot p_\theta(x) \qquad \text{for all } \theta \in \Theta \quad (3.1)$$

where C is a normalizing constant which does not depend on θ. In fact, we can have various IVP depending on C. For example, Shafer's linear plausibility function is defined as:

$$P^*(A|x) = \frac{\max_{\theta' \in A} p_{\theta'}(x)}{\max_{\theta \in \Theta} p_\theta(x)} \quad \text{for all } A \in \beta \text{ and } A \neq \Phi \quad (3.2)$$

The corresponding lower probability function is given as:

$$P_*(A|x) = 1 - \frac{\max_{\theta \in \bar{A}} p_\theta(x)}{\max_{\theta \in \Theta} p_\theta(x)} \quad \text{for all } A \in \beta \qquad (3.3)$$

In particular, when the set A is singleton, say $\{\theta'\}$, the function in Eq.(3.2) gives the relative likelihood of θ' to the most likely element in Θ.

4 DEMPSTER'S RULE FOR COMBINING IVP

The role of rules for combining evidence is to integrate the conditional knowledge about a proposition based on each single piece of evidence into combined knowledge based on the total evidence.

Various subjective Bayesian updating rules have been obtained by applying one or two of statistical independence assumptions to Bayes' rule [11,12]. However, there have been some controversies over the inconsistency between independence assumptions and their updating rules[13-15].

Dempster's rule [8] has been known to be mathematically well defined and accord with human intuition in combining multiple sources of evidence under uncertainty. The only condition that Dempster's rule requires is that the bodies of evidence to be combined must be entirely distinct. Combining entirely distinct bodies of evidence may be considered as a fusion of the individual observations made by independent observers on the same experiment. The meaning of independence here is that one's observation does not have effect on any of the other's, which is quite different from the conventional independence definitions in probability theory.

Suppose there are two entirely distinct bodies of evidence, E_1, E_2, which provide a proposition θ with lower probabilities, p_1, p_2, and its negation $\bar{\theta}$ with q_1, q_2, respectively, where $p_i+q_i \leq 1$, i.e., they are not additive. Then, the respective IVP for θ and $\bar{\theta}$ based on E_i are $[p_i, 1-q_i]$ and $[q_i, 1-p_i]$. Dempster's rule gives the lower probabilities for θ and $\bar{\theta}$ based on the combined evidence as follow:

$$p_*(\theta|E_1,E_2) = \frac{p_1 \cdot p_2 + p_1 \cdot (1-p_2-q_2) + p_2 \cdot (1-p_1 \cdot q_1)}{1-p_1 \cdot q_2 - q_1 \cdot p_2}$$

$$= 1 - \frac{(1-p_1) \cdot (1-p_2)}{1-p_1 \cdot q_2 - q_1 \cdot p_2} \qquad (4.1)$$

$$p_*(\bar{\theta}|E_1,E_2) = \frac{q_1 \cdot q_2 + q_1 \cdot (1-p_2-q_2) + q_2 \cdot (1-p_1 \cdot q_1)}{1-p_1 \cdot q_2 - q_1 \cdot p_2}$$

$$= 1 - \frac{(1-q_1) \cdot (1-q_2)}{1-p_1 \cdot q_2 - q_1 \cdot p_2} \qquad (4.2)$$

The upper probability for each proposition is obtained by using Eq.(2.6).

Let $\oplus$ denote the operation of combining IVP using Dempster's rule. It has the following properties:

1 It is commutative and associative; the order of evidence in combination does not effect the final result.
2 $[p_*, p^*] \oplus [0, 1] = [p_*, p^*]$; $[0, 1]$ plays the role of *identity* to this rule.
3 $[0, 0] \oplus [1, 1]$ is undefined.
4 $[0, 0]$ and $[1, 1]$ are annihilators; for an interval $[p_*, p^*] \neq [0, 0]$, $[p_*, p^*] \oplus [1, 1] = [1, 1]$, and for an interval $[p_*, p^*] \neq [1, 1]$, $[p_*, p^*] \oplus [0, 0] = [0, 0]$
5 The width of the combined interval is no larger than those of the intervals before the combination; the width of IVP corresponds to the measure of uncertainty, and it seems intuitively reasonable that the amount of the measure of uncertainty gets smaller as we gather more evidential information.

5 EXPERIMENT AND DISCUSSIONS

For the purpose of demonstrating the concepts, this approach was applied to the problem of ground-cover classification on multispectral data in conjunction with digital elevation data.

Table 1 describes the set of data sources for the experiment. The image in this data covers a forestry site around the Anderson River area of British Columbia, Canada. Source 1 consists of 4-band airborne multispectral scanner data in the visible region. Sources 2 and 3 are synthetic aperture radar imagery in the shallow mode and the steep mode, respectively. The spectral band column of sources 2 and 3 explains the band, and the transmit and receive type of SAR images. For example, XHV means that the image is obtained in X-band (λ=3cm) of microwave by horizontal polarization transmit and vertical polarization receive. The last source is the digital elevation data (DEM).

Table 2 is the statistical correlation matrix between spectral bands of the multiband image and the other sources. Correlations between every pair of bands (except XHV and XHH bands of SAR SHAL) from different sources are relatively low compared to those from the same source. When the data can be assumed to be normally distributed, their uncorrelatedness implies statistical independence. Thus, if XHV and XHH bands are excluded from the second source, we may assume that the data sources are globally independent. In remote sensing, however, it is believed that different remote sensors provide physically independent observations. Hence, allowing the second source to keep XHV and XHH bands does not violate the assumption of independence of evidence underlying Dempster's rule.

We have defined 4 information classes out of 9 cover types, and they are listed in Table 3. We assume that the classes have Gaussian distributions on DEM also. In Table 4, separabilities between classes based on DEM alone are computed by J-M distance. They are large enough not to be ignored, and DEM seems information-bearing from the cover type classification's point of view.

Table 1. Data set for Experiment.

Source index	Data type	Input channel	Spectral band(μm)
1	A/B MSS	4	.50-.55
		5	.55-.60
		6	.60-.65
		7	.65-.69
2	SAR SHAL		XHV
			XHH
			LHV
			LHH
3	SAR STEEP		XHV
			XHH
			LHV
			LHH
4	DEM		

Table 2. Statistical correlation matrix between bands and sources.

		A/B MSS				SAR SHAL				SAR STEEP			
		4	5	6	7	XHV	XHH	LHV	LHH	XHV	XHH	LHV	LHH
A/B MSS	4	1.000	.996	.984	.981	.975	.744	.102	.074	.039	-.141	.076	-.084
	5		1.000	.992	.990	.961	.742	.099	.069	.046	-.128	.082	-.075
	6			1.000	.998	.955	.672	.089	.060	.045	-.122	.078	-.068
	7				1.000	.951	.684	.093	.062	.048	-.122	.079	-.068
SAR SHAL	XHV					1.000	.677	.103	.075	.018	-.170	.061	-.101
	XHH						1.000	.147	.103	.075	-.102	.097	-.076
	LHV							1.000	.323	.209	.187	.165	.163
	LHH								1.000	.144	.147	.097	.087
SAR STEEP	XHV									1.000	.391	.559	.378
	XHH										1.000	.339	.471
	LHV											1.000	.347
	LHH												1.000

Table 5 summarizes the classification results after using 20 training pixels for each class. The maximum posterior classification on the composite of all four data sources (MPCC) provides a small increase in overall accuracy, but decreases average accuracy by a considerable amount compared to accuracy of source 1 alone. Meanwhile, our method (MSDC) increases both overall and average classification accuracies.

The experiment demonstrates 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 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.

Table 3. Information Classes in Experiment.

Class No.	Cover types	Pixel count	% of total
1	douglas fir w/ lodgepole pine	5423	24.10
2	hemlock w/ cedar	3173	14.10
3	douglas fir w/ other species	1309	5.82
4	clearcuts	12600	55.99

Table 4. Separability between classes on DEM

class	2	3	4
1	.574	.927	.630
2		1.000	.690
3			.719

Table 5. Classification results

	Class					
Source	1	2	3	4	Overall	Average
1	60.35	64.70	71.89	81.98	73.75	69.73
2	72.30	42.96	25.90	79.86	69.71	55.26
3	52.20	19.57	3.97	72.68	56.28	37.11
4	51.91	71.57	0.00	81.98	68.51	51.37
MPCC	48.85	52.44	63.48	91.35	74.02	64.03
MSDC	69.52	68.23	88.69	82.70	77.84	77.29

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. One of the features of the method is the capability of plausible reasoning under uncertainty in pattern recognition and information processing, especially where observed data are not 100% reliable.

ACKNOWLEDGEMENTS

This research is 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] T. Lee, J.A. Richards, and P.H. Swain, "Probabilistic and evidential approaches to multisource data analysis," *IEEE Trans. on Geos. and Remote Sensing*, vol.GE-25, pp283-293, 1987.

[2] A.R. Jones, J.J. Settle, and B.K. Wyatt, "Use of digital terrain data in the interpretation of SPOT-1 HRV multispectral imagery," *Int. J. Remote Sensing*, vol.9, no.4, pp669-682, 1988.

[3] C.F. Hutchinson, "Techniques for combining Landsat and ancillary data for digital classification improvement", *Photogrammetric Engineering and Remote Sensing*, vol.48, no.1, pp123-130,1982.

[4] H. Kim, P.H. Swain, and J.A. Benediktsson, "A method for combining multispectral and ancillary data in remote sensing and geographic information processing", *Proc. 1986 IEEE Inter. Conf. on S.M.C.*, vol.2, pp1543-1547, 1986.

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

[6] 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.

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

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

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

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

[11] R. Duda, J. Gaschnig, and P. Hart, "Model design in the PROSPECTOR: consultant system for mineral exploration", in D. Michie, (eds)., *Expert Systems in the Micro-electronic Age*, Edinburg Univ. Press, Edinburg, pp157-167, 1979.

[12] E.H. Shortliffe, *Computer-Based Medical Consultations : MYCIN*, American Elsevier, 1976.

[13] R.W. Johnson, "Independence and Bayesian updating methods", *Artificial Intelligence*, vol.**29**, pp217-222, 1986.

[14] C. Glymour, "Independence assumptions and Bayesian updating", *Artificial Intelligence*, vol.**25**, pp95-99, 1985.

[15] E. Pednault, S. Zucker, and L. Muresan, "On the independence assumption underlying subjective Bayesian updating", *Artificial Intelligence*, vol.**16**, pp213-222, 1981.

WOCON-INFOR 89

PROCEEDINGS OF WORLD CONFERENCE ON INFORMATION PROCESSING ⁄ COMMUNICATION

Editors: Kyung Whan Lee and Chisu Wu

Hilton Hotel
Seoul, Korea
June 14-16, 1989