102386

A Method for Combining Multispectral and Ancillary Data in Remote Sensing and Geographic Information Processing

Hakil Kim, Philip H. Swain, and Jon A. Benediktsson

School of Electrical Engineering
Purdue University
W. Lafayette, IN 47907 USA

ABSTRACT

This paper introduces a method for combining multispectral and ancillary data in remote sensing and geographic information processing for improvement in digital classification and accuracy of extracting information. The method has the capabilities of representing the uncertainty associated with observed data and providing plausible reasoning in data analysis. This method is viewed from the standpoint of knowledge engineering, and each source is considered as a piece of evidence which provides a certain measure of support to hypotheses. Each body of evidence also has a degree of reliability which represents the relative quality of the corresponding data set in data analysis. The degree of reliability and measures of support of the respective pieces of evidence need to be represented numerically in order for multiple bodies of evidence to be pooled. This paper focuses on formal approachs to the quantitative representations of the degree of reliability and measure of support, and examines existing combining functions of evidence.

This research is supported by the National Science Foundation Grant No. ECS 84-00324.

1. Introduction

As remote sensing and other data acquisition technologies have advanced in recent years, there has been a trend towards exploiting remotely sensed data in conjunction with other ancillary data in geographic information systems to extract more reliable information from multi-attribute data bases. For instance, digital elevation data and slope data make it possible to utilize topographic information together with remotely sensed data for the purpose of land cover analysis. Climatic and meteorological data are used as inputs for crop production estimates [1].

However, ancillary data in geographic information systems present several problems when combined with remotely sensed data in an automated classification. To begin with, since spatial variation of the factor in ancillary data, such as vegetation cover, soil type, or slope aspect, has an effect on the spectral responses obtained from remote sensors, there must be significant but unknown interactions among multiple data sources. So, while it is often reasonable to use the multivariate Gaussian distribution to represent multispectral data alone, this statistical model cannot be extended to accommodate geographic or topographic data combined with remotely sensed data. As well as these unknown interactions, there is a difficulty in describing the various data types which have different units of measurement. The types of data to be combined cannot be assumed to be commensurable. Moreover, some types of data may even be nonnumerical. Such data cannot be treated jointly with other types of data by the conventional multivariate methods. Because of the fact that the quantitative representation of the quality of these data depends on the expertise and intuition of the human analyst, the methods for the analysis of multisource data may be *ad hoc* [1][2].

The initial purpose of this paper is to introduce a new method for combining multispectral data with other ancillary data in remote sensing and geographic information processing. The traditional classification methods for the analysis of data in remote sensing are based on Bayesian probability theory. Bayesian theory cannot represent the uncertain state between TRUE and FALSE because it requires the constraint that the sum of probabilities of an event and its complementary event should be equal to unity. The method described here has the capabilities of representing the uncertainty associated with observed data and providing plausible reasoning in data analysis.

From the standpoint of knowledge engineering, each source of data can be considered as a piece of evidence providing a certain measure of support to each hypothesis. Each piece of evidence also has a degree of reliability which represents how reliable the corresponding data source of the evidence is for data analysis. If degrees of reliability and measures of support are represented numerically, multiple bodies of evidence can be pooled by using combining functions which already exist. The focus of this paper is on a formal approach to the quantitative representation of degrees of reliability and measures of support on the basis of statistical separability and the probability density function of a feature, respectively. By this approach, we present a way of modeling human reasoning under uncertainty - where the problem data or the decision rules are not completely reliable - in pattern recognition and data classification.

2. Functions for Combining Evidence

This section describes various combining functions of evidence which already exist, such as Dempster's rule for combining evidence, the combining function of the measure of belief in *MYCIN*, and the *AND* operator in fuzzy logic, and shows how they can be applied to classification of the data in remote sensing and geographical information processing. These functions must be able to combine multiple pieces of evidence in a consistent manner.

In the sequel, it will be seen that the basic belief measure such as the basic probability measure in Dempster-Shafer theory, the measure of belief (MB) in *MYCIN*, or the measure of membership in fuzzy set theory must be defined before applying one of these combining functions of evidence to real-life problems.

Dempster's rule for combining evidence [3]

Let A represent the hypothesis that an observation X belongs to a certain class ω. Since we are interested only in whether X belongs to ω or not, we can have a frame of discernment Θ which has a single focal element A. This implies that the evidence supports precisely and unambiguously the hypothesis which is a single non-empty subset A of Θ. In this case, we can say that the evidence is limited to supporting A with a certain degree of support.

The degrees of support provided by such evidence are easily specified by formulating a simple support function. If s is the degree of support for A, where $0 \leq s \leq 1$, then the degree of support for a set $B \subseteq \Theta$ is given by;

$$S(B) = \begin{cases} 0 & \text{if } B \text{ does not contain } A \\ s & \text{if } B \text{ contains } A \text{ but } B \neq \Theta \\ 1 & \text{if } B = \Theta \end{cases} \tag{1}$$

The function $S : 2^{\Theta} \rightarrow [0, 1]$ defined above is called a ***simple support function*** focused on A. S is also a belief function with basic probability measures $m(A) = S(A)$, $m(\Theta) = 1 - S(A)$, and $m(B) = 0$ for all other $B \subset \Theta$. Now, we are ready to apply Dempster's rule of combination to simple support functions.

Let S_1 and S_2 be simple support functions based on different pieces of evidence, E_1 and E_2, respectively. Both S_1 and S_2 are focused on A. Assume that S_1 has the basic probability measure $m_1(A) = s_1$ and $m_1(\Theta) = 1 - s_1$, while S_2 has the basic probability measure $m_2(A) = s_2$ and $m_2(\Theta) = 1 - s_2$. Figure 1 is a graphical interpretation of Dempster's rule of combination for these two simple support functions. Only the upper right-hand rectangle, of measure $(1 - s_1)\cdot(1 - s_2)$, fails to be committed to A.

Based on Dempster's rule of combination, the result of orthogonal sum $S = S_1 \oplus S_2$ is another simple support function with basic probability measure as follows:

$$m(A) = 1 - (1 - s_1)\cdot(1 - s_2) \quad (2.a)$$

$$= s_1 + (1 - s_1)\cdot s_2 \quad (2.b)$$

$$m(\Theta) = (1 - s_1)\cdot(1 - s_2) \quad (3)$$

$m(A)$ is the degree of support for A based on the combined evidence.

Combining function of MB in *MYCIN*

MYCIN is a computer-based medical consultation system devised by E.H. Shortliffe [4]. *MYCIN* uses measures of belief and disbelief (MB, MD) to quantify degrees of belief and disbelief of human experts in a hypothesis given a piece of evidence. There are several types of combining functions in *MYCIN*, one of which combines measures of belief (or disbelief) in a hypothesis given multiple bodies of evidence into the combined measure of belief (or disbelief) in the same hypothesis given the combined evidence.

This subsection introduces the combining function for measure of belief.

Let $MB[A,E_1]$ and $MB[A,E_2]$ represent the measures of belief on the hypothesis A given two bodies of evidence, E_1 and E_2 respectively. The combined measure of belief in A based on the new evidence $E_1\&E_2$ is given as:

$$MB[A,E_1\&E_2] = MB[A,E_1] + (1 - MB[A,E_1]) \cdot MB[A,E_2] \tag{4}$$

This function states that the combined measure of belief is stronger than any measure of belief based on a single piece of evidence. This implication accords with our intuition that several concordant indicators reinforce each individual indicator.

Comparing Eq. (2.b) and Eq. (4), it is seen that they produce exactly the same result. The combining function used in *MYCIN* is a special case of Dempster's rule of combination when Dempster's rule is applied to simple support functions.

AND operation in fuzzy logic

Fuzzy set theory was invented by L.A. Zadeh [5] by generalizing classical set theory. This theory has been shown to have a much wider scope of applicability than the ordinary set theory, particularly in the fields of pattern recognition and information processing. Whereas a set in classical set theory is a collection of precisely specified elements in a sample space, a fuzzy set is a collection of objects which do not have a precisely defined criterion of membership. The elements in a fuzzy set are characterized by a membership function, that assigns to each element a grade of membership which is a real number in the interval [0, 1]. When we deal with a set in the ordinary sense of the term, its membership function can take on only two values, 0 and 1.

As fuzzy set theory was evolved from classical set theory, fuzzy logic is an extended version of Boolean [6]. Fractional values of a membership function in fuzzy set theory are interpreted as partial truth values in fuzzy logic. To combine non-integer truth values, fuzzy logic defines the equivalents of the logical operators in ordinary Boolean algebra. We examine in this subsection the *AND* operation in fuzzy logic.

Let A represent the same hypothesis as above, and let $u_1(A)$ and $u_2(A)$ represent the truth values of an unclassified observation X on A given two different bodies of evidence, E_1 and E_2, respectively. By the definition of the *AND* operator in fuzzy logic, the truth value of X on A based on the combined evidence, $u(A)$, is obtained as:

$$u(A) = u_1(A) \ AND \ u_2(A) = MIN\{u_1(A), u_2(A)\} \tag{5}$$

where *MIN* is the ordinary minimum operator.

This operator has the following properties:

(1) The truth value based on the combined evidence depends on the truth values based on each single piece of evidence.

(2) Commutativity, and associativity.

(3) A small increase in either of the component truth values cannot induce a strong increase in the combined truth value.

(4) Complete certainty based on the combined evidence is implied only by complete certainty based on both bodies of evidence.

(5) Complete uncertainty based on either body of evidence implies complete uncertainty based on the combined evidence.

The combining functions of evidence reviewed in this section are applicable only when the basic belief measures which they contain can be defined. The next section shows a possible way of defining those measures based on a feature in data classification.

3. Measure of Support

The term "measure of support" used in this paper refers to general measures including the basic probability measure in Dempster-Shafer theory, the measure of belief in *MYCIN*, and the degree of membership in fuzzy set theory. The measure of support is not necessarily constrained to obey Bayesian probability theory. To apply the combining functions of evidence described in the previous section to real problems, the measure of support based on each piece of evidence must be represented in a mathematical expression, either parametric or non-parametric. In this section, we present a formal approach to the quantitative representation of the measure of support given data sets in remote sensing and geographic information systems.

In our application, Euclidean distance can be used as a feature in data classification to determine the degrees of support based on various data sets such as multispectral, digital elevation, and digital slope data. Euclidean distance is one of the simplest and most intuitive features in pattern classification. More importantly, it is a good feature as a unifier of various types of numerical data to be combined.

Let $\mathbf{d}_i$ be a random variable representing the Euclidean distance from the mean vector to the observation vectors in ω_i. Then, we define the measure of support for the class ω_i as

$$B_i(d) = 1 - P_{d_i}\{d_i \leq d\} = 1 - F_{d_i}(d) \qquad (6)$$

where d denotes the Euclidean distance from the mean vector to a given observation vector X, $P_{d_i}\{d_i \leq d\}$ is the probability of the event $\{d_i \leq d\}$ for samples in ω_i, and $F_{d_i}(d)$ is called the cumulative distribution function of d_i. Since the probability distribution function is the integration of probability density function, the function $B_i(\cdot)$ has the following properties:

(1) $B_i : [0, \infty] \dashrightarrow [0, 1]$

(2) Nonincreasing.

(3) $B_i(0) = 1$, and $B_i(\infty) = 0$.

Given an observation X, d is obtained by the definition of Euclidean distance. As d increases to infinity, $F_{d_i}(d)$ increases to unity. This corresponds to the human intuition that the disbelief in the hypothesis of X belonging to ω_i increases as the Euclidean distance between the mean and X increases. Therefore, if the reciprocal of Euclidean distance is interpreted as the weight of evidence, then $1 - F_{d_i}(d)$ may be considered as the measure of support for the hypothesis of X belonging to ω_i. Although $B_i(\cdot)$ is a nonprobabilistic measure, it cannot be interpreted as a purely heuristic measure because it is derived from the probability density function.

Euclidean distance is a useful distance measure by which satisfactory results in data classification can be produced when the measurement vectors of each class tend to cluster tightly about a typical or representative vector for that class. In most practical cases, however, this method does not seem promising because the measurement vectors of each class are correlated (between components) and dispersed (within components) to various degrees. We need another measure which accounts for correlations and dispersions. The ***Mahalanobis distance*** from an observation vector X to the mean vector M_i of ω_i is defined as [7]

$$d_M = (X - M_i)^T \Sigma_i^{-1} (X - M_i) \quad (7)$$

where Σ_i is the covariance matrix of the class ω_i. This measure is very useful when it is desired to incorporate statistical properties in a distance measure.

For any unclassified measurement X, the measure of support for the hypothesis of X being classified into ω_i is computed by $B_i(d_M)$ instead of $B_i(d)$, where d_M and d are as defined above. In this manner, correlations and dispersions of samples are taken into account in this method.

The results of applying this measure for the classification of multispectral data combined with digital elevation and slope data will be demonstrated during the presentation.

4. Degree of Reliability

Degree of reliability is a relative quality factor for a source of data. Since all the data sets from different sources are in general not equally reliable, we have to represent degrees of reliability numerically to take into account this data quality in data classification. The numerical value of the degree of reliability for a source may be taken as the maximum value of measure of support for a hypothesis based on the body of evidence obtained from the source. The measure of support is a maximum when the hypothesis is absolutely certain based on the body of evidence.

Ideally, we would like to quantify the degree of reliability on the basis of probability of error. But it has been observed that computing probability of error is often not feasible. Statistical separability is an alternative. Using separability information for the quantitative representation of degree reliability seems to be natural if we keep in mind that we are assigning reliability factors for data classification.

For example, the Jeffries-Matusita (J-M) distance is a measure of statistical separability of pairs of classes. Formally, it is defined as follows [8]:

$$J_{ij} = \left\{ \int_X [\sqrt{p(X|\omega_i)} - \sqrt{p(X|\omega_j)}\,]^2 dX \right\}^{\frac{1}{2}} \tag{8}$$

where $p(X|\omega_i)$ is the probability density function for class i. When the classes are assumed to have normal density functions, Eq.(8) reduces to

$$J_{ij} = \sqrt{[2(1 - e^{-\alpha})]} \tag{9}$$

where α is the Bhatacharyya distance which is defined as [9]

$$\alpha = \frac{1}{8}(M_i - M_j)^T \left(\frac{\Sigma_i + \Sigma_j}{2}\right)^{-1} (M_i - M_j) + \frac{1}{2}\log_e\left[\frac{|(\Sigma_i + \Sigma_j)|/2}{\sqrt{|\Sigma_i| \cdot |\Sigma_j|}}\right] \tag{10}$$

The average J-M distance over class pairs is computed by

$$J_{ave} = \sum_{i=1}^{N} \sum_{j=1}^{N} P(\omega_i) P(\omega_j) \, J_{ij} \tag{11}$$

where N is the number of classes, and $P(\omega_i)$ is a priori probability of class ω_i. J_{ave} has a maximum value of $\sqrt{2}$. Therefore we define the degree of reliability by dividing J_{ave} by $\sqrt{2}$ as follows:

$$R = \frac{J_{ave}}{\sqrt{2}} = \sum_{i=1}^{N} \sum_{j=1}^{N} P(\omega_i) P(\omega_j) \cdot \sqrt{1 - e^{-\alpha}} \tag{12}$$

The effect of the negative exponential term is to give an exponentially decreasing weight to increasing differences between the class density functions, which coincides with human reasoning in quantifying the relative quality of the respective data sources. As a result, the degree of reliability proposed here has a saturating behavior. As a matter of fact, in formulating the degree of reliability from the statistical separability, we may employ any separability measure which has the saturating behavior, such as the average transformed divergence.

5. Summary

In this paper we have presented a new method for combining multispectral and ancillary data in remote sensing and geographic information processing. The method is viewed from the standpoint of knowledge engineering so that it has the capabilities of representing the uncertainty associated with observed data and providing plausible reasoning in data analysis. Each source of data is considered as a piece of evidence which provides a certain measure of support to hypotheses concerning data classification. The key aspects of this method are the quantitative representations of measure of support and degree of reliability of the respective data sources.

The function of measure of support is derived from the probability density function of a feature selected for classification purposes. Therefore, although the measure of support is not a probabilistic measure, it need not be interpreted as a completely heuristic measure. Several combining functions of evidence have been examined that pool multiple bodies of evidence into a combined evidence. The degree of reliability is the relative quality of each source of data. It has been proposed that this quality factor can be numerically represented on the basis of the statistical separability information over class pairs.

In conclusion, the method described in this paper is intended to suggest a way of modeling human reasoning under uncertainty in pattern recognition and information processing, especially where the observed data or the decision rule are not 100% reliable. This technique is under implementation, and the result of classification experiments with Landsat MSS data combined with digital elevation and slope data will be demonstrated during the presentation.

REFERENCES

[1] P.H. Swain, J.A. Richards, and T. Lee, "Multisource data analysis in remote sensing and geographic information processing", *Proc. of the 11-th International Symposium on Machine Processing of Remotely Sensed Data*, pp. 211-217, W.Lafayette, IN., June 1985.

[2] C.F. Hutchinson, "Techniques for combining Landsat and ancillary data for digital classification improvement", *Photogrammetric Engineering and Remote Sensing*,

vol. **48**, no. 1, pp. 123-130, 1982.

[3] G. Shafer, *Mathematical Theory of Evidence*, Princeton Univ. Press, 1976.

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

[5] L.A. Zadeh, "Fuzzy sets", *Information and Control*, vol. **8**, pp. 338-353, 1965.

[6] D. Dubois, and H. Prade, *Fuzzy sets and systems : theory and applications*, Academic Press, 1980.

[7] J.T. Tou, and R.C. Gonzalez, *Pattern Recognition Principles*, Addison-Wesley Publishing Co., 1974.

[8] P.H. Swain, and S.M. Davis, eds., *Remote Sensing : The Quantitative Approach*, McGraw-Hill, New York, 1978.

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

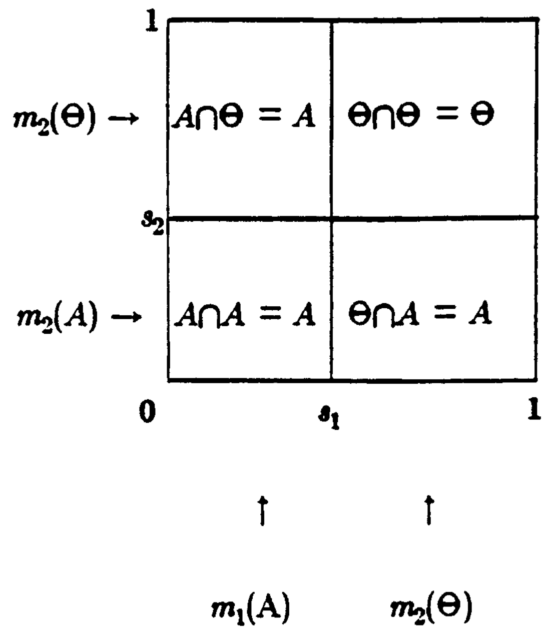

Fig. 1.