Geostatistical Fusion of Spectral and Spatial Information in Remote Sensing Data Classification

No-Wook Park, Kwang-Hoon Chi
Geoscience Information Center
Korea Institute of Geoscience and Mineral Resources
30 Gajeong-dong, Yuseong-gu, Daejeon 305-350, Korea
nwpark@kigam.re.kr

Byung-Doo Kwon
Dept. of Earth Sciences, Seoul National University
San 56-1, Shillim-dong, Kanwak-gu, Seoul 151-742, Korea

Abstract: This paper presents a geostatistical contextual classifier for the classification of remote sensing data. To obtain accurate spatial/contextual information, a simple indicator kriging algorithm with local means that allows one to estimate the probability of occurrence of certain classes on the basis of surrounding pixel information is applied. To illustrate the proposed scheme, supervised classification of multi-sensor remote sensing data is carried out. Analysis of the results indicates that the proposed method improved the classification accuracy, compared to the method based on the spectral information only.
Keywords: Geostatistics, Spatial Information, Classification.

1. Introduction

Since the remote sensing data have the inherent spatial nature, an effective way to improve the accuracy of the classification of remote sensing data is to incorporate the spatial information into spectral information in data processing procedures. In traditional way for the classification of remote sensing data, most commonly used methods are purely based on the spectral information extracted from remote sensing data, regardless of single source data or multi-source data. As a result, noisy features such as isolated pixels are often shown in the classification result. For example, the backscattering signal from spatially adjacent objects in a SAR image has often interference relics and appears to be overlapped. In these cases, the mean value of the pixel intensity of the target objects are usually close to each other, while the standard deviations are very different. Therefore, the classification including SAR images based on the information provided by individual pixels cannot generally produce satisfactory results due to speckle. In this case, if we consider the pixels in context with other measurements, more complete information might be derived.

To overcome this type of drawbacks, many "contextual" classification methods have been proposed and tested [1]. One of the most widely used methods to integrate contextual information in the classification is Markov random field (MRF) developed on the basis of statistical properties of the data [2], [3]. Though MRF has been successfully applied, it is difficult to formulate an effective method which can correctly infer the parameters for the given model, since the method contains many parameters which are difficult to interpret.

Related to the processing of spatial data, geostatistics provides us with a collection of statistical tools to model the spatial variability [4]. Originally, geostatistics was devised to estimate statistical properties of unsampled points for delineating ore deposit models. Nowadays, geostatistics is increasingly used to infer the local and spatial uncertainty and integrate various data. Despite its great potential of spatial data processing, geostatistics has seldom been used in remote sensing, since remote sensing data already provide exhaustive information. However, if we regard the classification procedure as the prediction of ground properties of unsampled points, we can incorporate the spatial coordinates of sparse ground data into remote sensing data that are exhaustive sources of indirect information on the ground properties by using some geostatistical algorithms. Recently, geostatistical algorithms incorporated to the Bayesian classification procedure have been proposed and applied to hyperspectral data classification [5] and multi-sensor data fusion [6]. However, sophisticated geostatistical algorithms have not been tested in remote sensing data classification yet.

In this paper, we present a new contextual classification method based on geostatistics. An indicator-based geostatistical algorithm is proposed and tested for integration of spectral and spatial information in the remote sensing data classification. The geostatistical algorithm proposed here is simple indicator kriging with local means. This algorithm has its merit to incorporate various independent information (e.g. spectral information and spatial information) afterward and improve the confidence in a classification stage. Supervised land-cover classification using multi-sensor remote sensing data is applied to illustrate application of the method.

2. Methods

1) Classification Based on Geostatistics

The basic concept of the proposed method is that the unsampled location is likely to be allocated to the same land-cover classes as the nearest observation. To incorporate these spatial patterns into the classification procedure, geostatistical algorithms that allow one to account for hard and soft probabilities with a neighborhood can be applied. In traditional remote sensing data classification tasks, the training data which represent the spectral signatures of land-cover classes can be regarded as the hard data which are precise measurement of the land-cover class of interest. Meanwhile, spectral information derived from remote sensing data can be regarded as soft data that provide indirect information on the land-cover classes. Throughout this paper, the hard and soft data (information) refer to the ground-based training data and spectral feature derived from remote sensing data, respectively.

The basic paradigm of geostatistics is based on a random function model, whereby a set of unknown values is regarded as a set of spatially dependent random variables [4]. Once a random function model has been chosen, the next step is to infer its spatial patterns from the available information. The spatial patterns are generally described in terms of a variogram, dissimilarity of observations as a function of the separation distance and direction. By computing the variogram, spatial information via the covariance of each class can be incorporated into the data processing.

For categorical attributes such as land-cover classes, quantitative information on the spatial correlation between different categories can be handled by the indicator algorithm [4]. The indicator approach provides a non-parametric distribution estimated directly at two possible outcomes: 0 and 1. One major advantage of the indicator approach is the ability to process different type of information (hard and soft data) together, regardless of their origins.

Suppose that $\{\omega_k,\ k = 1,2,\cdots,K\}$ is a set of *K* mutually exclusive land-cover classes with *n* ground observations (training data) $\{\omega(X_\alpha),\ \alpha = 1,2,\cdots,n\}$ that are considered as precise measurements (hard data) of the class ω_k prevailing at X_α.

Using both hard and soft data, we aim to assign the land-cover class to any unsampled locations. Once the uncertainty has been modeled, a single land-cover prevailing at each location is determined. The uncertainty is modeled by the conditional probability distribution function (pdf) of the discrete random variable $\omega(X)$ which is conditional to the surrounding information.

$$\begin{aligned} p(X;\omega_k \mid h+s) &= p(\omega(X) = \omega_k \mid h+s) \\ &= E\{i(X;\omega_k) \mid h+s)\} \end{aligned} \tag{1}$$

where "$\mid h+s$" expresses the conditioning to the hard data (*h*) and soft data (*s*) retained in the neighborhood of *X*. The notations *i* and *E* express an indicator variable and an expectation of it, respectively.

Within the indicator framework, the hard data and soft data are firstly coded into a set of *K* local prior probabilities.

Precise measurements of a land-cover category ω_k at hard data location X_α are coded into a set of *K* binary indicator data defined as:

$$i(X_\alpha;\omega_k) = \begin{cases} 1 & if\ \omega(X_\alpha) = \omega_k \\ 0 & otherwise \end{cases} \tag{2}$$

Spectral or backscattering feature information derived from remote sensing data provides prior probabilities of occurrence for the class ω_k at location *X*.

$$y(X;\omega_k) = p(\omega_k \mid s) \tag{3}$$

This set of local soft indicator data has values between 0 and 1. In order to get the soft indicator data, we can apply traditional feature extraction methods widely used in the remote sensing community. If we deal with multi-source/multi-sensor data, various data fusion methods within a probabilistic framework can be applied.

After finishing the prior indicator coding, the next step is to update different types of prior probabilities mentioned above into posterior pdf values. When dealing with multi-source/multi-sensor data, we should integrate all different sources of soft information to build the best (i.e. most data-charged) prior, then update it with the hard data.

Updating hard and soft prior probabilities into posterior pdf values is carried out through several algorithms. In this paper, we apply simple indicator kriging with local means to perform such updating.

2) Simple Indicator Kriging with Local Means

With spectral information providing information about the local prior probability $p(\omega_k \mid s)$ of the land-cover class ω_k prevailing at location *X*, the simple indicator kriging estimate can be written as follows:

$$\begin{aligned} i(X_\varepsilon;\omega_k) = &\sum_{\alpha=1}^{n} \lambda_\alpha(X;\omega_k)[i(X_\alpha;\omega_k) - p(\omega_k \mid s)] \\ &+ p(\omega_k \mid s) \end{aligned} \tag{4}$$

Eq. (4) represents an updating of the prior probability $y(X_\alpha;\omega_k)$ at location X_α using the neighboring hard indicator data.

The kriging weights $\lambda_\alpha(X_\alpha;\omega_k)$ are obtained by solving a simple indicator kriging system:

$$\sum_{\beta=1}^{n} \lambda_\beta C_I(X_\beta - X_\alpha;\omega_k) = C_I(X - X_\alpha;\omega_k), \alpha = 1,\cdots,n \tag{5}$$

where $C_I(X_\beta - X_\alpha;\omega_k)$ is the covariance of the residual random function.

Using Eq. (4) and Eq. (5), we estimate the unknown residual from the residual data using simple indicator kriging and to add the resulting estimate to the prior local mean $p(\omega_k \mid s)$. After obtaining the posterior probabilities for all land-cover classes, each pixel is assigned to certain class, which maximizes the posterior probabilities.

One should notice the following issues when applying this algorithm to land-cover classification. When integrating ground-based hard data with soft remote sensing data, the behavior of spectral/soft information is crucial for evaluating the contribution of soft information in land-cover classification. For this algorithm, however, there is no other free parameter to impose a higher weight on soft data. The influence of the soft information is limited to the mean of the prior local cdf at location *X*. Intrinsically, this algorithm considers the soft information from remote sensing data as an average spatial variation of the primary land-cover class. If soft information does not allow a significant discrimination of the land-cover classes, the estimate would then revert to the simple kriging estimate with constant stationary mean. Though this algorithm does not directly control the spatial variability of the land-cover classes, it is much simpler to implement in practice, compared to soft cokring or MRF.

3. Experiments

To illustrate the proposed method, we applied the method to the multi-sensor remote sensing data set (data set grss-dfc-0006 [7]). The data set includes 6 optical ATM images and 9 SAR images of the P, L, and C bands with full polarizations. The study area is an agricultural site that consists of five agricultural classes (i.e. sugar beets, stubble, bare soil, potatoes and carrots). As for the training and reference data, we preprocessed them by applying a stratified random sampling scheme to whole ground truth data in the study area. The proportion of training data is 5% in the study area.

In our approach, the training data were used for two purposes. The first use was to compute spectral information (i.e. conditional probability) for each remote sensing image. Second, they were treated as the hard data to update the prior probabilities derived from the remote sensing images into posterior pdf values.

To extract the spectral information from the remote sensing images, we separately processed the optical images and the SAR images. As for the optical images, a traditional parametric maximum likelihood classifier was adopted. The smoothed kernel method was adopted for classifying the SAR images. Then the Bayesian probabilistic fusion approach was applied to obtain the final fused spectral information.

After the indicator coding of the training data, at each training datum, the residual values were computed by subtracting the soft indicator datum from the collocated hard indicator datum. The variogram of residuals was then computed and modeled. Then the residual values at all pixels were estimated using simple indicator kriging and the neighboring hard training data. The posterior probability was obtained by adding the soft indicator datum to the simple kriging estimate. Finally, each pixel was allocated to the land-cover class with the largest posterior probability of occurrence.

Compared with the classification result based only on spectral information, the proposed method showed more homogeneous results in each class region (not shown here). The results obtained here mainly arose from the consideration of spatial information, in addition to the spectral information.

To investigate the effects of training data density on final classification accuracy, we repeated the experiments using training data of 2.5% and 1% proportions in the study area, respectively. According to decrease of the proportions of the training data, the overall accuracy was decreased. However, the proposed algorithm showed the higher overall accuracy than the spectral information based classification in all cases.

4. Conclusions

For integrating spatial information with spectral information in remote sensing data classification, we present and investigate the applicability of the geostatistical algorithm, simple kriging with local means. Compared to the traditional spectral based classification, the proposed algorithm could account for the spatial variability and improve the classification accuracy.

Acknowledgement

The data set grss-dfc-0006 was provided by the IEEE GRSS Data Fusion Committee. This research was partly supported by the Korea Remote Sensing Land Research Center.

References

[1] Tso, B. and P.M. Mather, 2001. *Classification methods for remotely sensed data*, Taylor & Francis, London.

[2] Geman, S. and D. Geman, 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Trans. Pattern Anal. Machine Intell.*, 6: 721-741.

[3] Solberg, A.H.S., T. Taxt, and A.K. Jain, 1996. A Markov random field model for classification of multisource satellite imagery, *IEEE Trans. Geosci, Remote Sensing*, 34: 100-113.

[4] Goovaerts, P., 1997. *Geostatistics for natural resources evaluation*, Oxford University Press, NY.

[5] Goovaerts, P., 2002. Geostatistical incorporation of spatial coordinates into supervised classification of hyperspectral data, *J. Geographical Systems*, 4: 99-111.

[6] Park, N.W., K.H. Chi, and B.D. Kwon, 2002. Combining geostatistical indicator kriging with Bayesian approach for supervised classification, *Proc. ISRS 2002*, Korea, pp.382-387.

[7] **URL:** IEEE GRSS Data Fusion Committee, data set grss-dfc-0006. Available at: http://www.dfc-grss.org.