CONTEXTUAL BAYESIAN CLASSIFIER

Michal Haindl

Institute of Information Theory and Automation
Czechoslovak Academy of Sciences
Pod vodarenskou vezi 4, 182 08 Prague 8
Czechoslovakia
ISPRS Commission VII

Abstract

A new type of Bayesian classifier is introduced to recognize remote sensing image data. The classifier uses contextual information about the classified pixel surrounding based on autoregression model prediction.

Key Words: Bayesian Classifier

1 INTRODUCTION

The conventional approach to remote sensing picture data classification is to perform test on unknown pixel against all classes using a spectral feature subset and then assign the unknown pixel to one of these classes, not taking respecting the large spatial correlation (Tubbs,2). The interpretation of picture element - pixel does not depend upon any relationship with any other pixel, so we do not use all available information. Not using context has fundamental limitation influence on classification accuracy for machine recognition. On the other hand the use of all context information is paid in unsolvable increase of computational demands. Possible solution is a compromise.

The developed decision rule is the optimal contextual Bayesian classifier with several simplifications. We assume local neighbour pixels to be conditionaly statisticaly independent, their class conditional density to be independent on neighbour labels and we neglect the space arrangement of local neighbour labels. In the first step the thematic map as the output from per-point Bayesian classifier is created. The second step consists of Bayesian classifier in which formula the apriori class probabilities are replaced by noncausal frequency predictor. This predictor is based on the autoregressive model of class frequences estimated from the first step classification.

2 CONTEXTUAL CLASSIFIER

Let us chose some direction of movement on the image plane, for example row scanning from left to right and top to bottom. According to this choice the following index is used throughout this paper:

$$t = (i-1)N' + j \tag{1}$$

where $i, j = 1, 2, \ldots, N'$ is row and column index, respectively. $N' \times N'$ is the size of classified image. Let us denote X_t (multidimensional pixel) in time t of scanning, ω_i $i = 1, \ldots, K$ class indicator, Y_t the K-dimensional vector, which i-th compound is the occurence frequency of ω_i in a window D_t . D_t window is defined as the set of thematic map entries within a given region centered around the class indicator corresponding X_t . The window is defined to be perpendicular with odd number of class indicators in both directions, so that each frequency vector Y can be assigned to the center pixel of its respective window.

Let us denote the set of past thematic map windows

$$D^{(t)} = \{D_t, D_{t-1}, \ldots, D_1\} \tag{2}$$

and $\tilde{D}^{(t)}$ the set of all pixel indexes from $D^{(t)}$. We have chosen the noncausal frequency predictor:

$$\hat{Y}_t = E[Y_t | D^{(t-1)}] \tag{3}$$

Let us denote

$$M_i(X_t) = (X_t - \mu_i)^T \Sigma_i^{-1} (X_t - \mu_i) \tag{4}$$

where μ_i, Σ_i are mean value vectors and covariance matrices (in practise their estimators) of single classes. Now, the modified Bayesian decision rule using the context information about the class membership of surrounding, can be put in the form:

Assign feature vector X_t into class ω_i if

$$\begin{aligned} & M_i(X_t) + \ln|\Sigma_i| - 2\ln \hat{Y}_t(i) \\ = & \min_{j=1,\ldots,K} \{M_j(X_t) + \ln|\Sigma_j| - 2\ln \hat{Y}_t(j)\} \end{aligned} \tag{5}$$

The thematic map for prediction construction is assumed to be the output from per-point Bayesian classifier , then the relation (3) can be put in the form (6)

$$\begin{aligned} \hat{Y}_t = E[Y_t | & \min_{j=1,\ldots,K} \{M_j(X_m) + \ln|\Sigma_j| \\ - & \ 2\ln P_j\} : m \in \tilde{D}^{(t-1)}] \end{aligned} \tag{6}$$

where P_j are some prior class probabilities estimations.

3 THE AUTOREGRESSIVE MODEL OF CLASS FREQUENCES

Several different stochastic models can be used for thematic map class frequences modeling. However the computational complexity is serious limiting factor, therefore the autoregressive model (7) was chosen.

$$Y_t = \sum_{i=1}^{N} A_i Y_{t-1} + \hat{E}_t \tag{7}$$

Where $Y_t \hat{E}_t$ are $K-$dimensional vectors, A_i are KxK matrices of unknown model parameters and N is the order of model. $\hat{E}_t$ is the white noise vector with following properties for $t > N$:

$$E[\hat{E}_t] = 0$$

$$E[\hat{E}_t \hat{E}_{t-i}^T] = 0 \qquad i \neq 0 \qquad i < t \tag{8}$$

$$E[\hat{E}_t Y_{t-1}^T] = 0 \quad 0 < i < t$$

We assume, that probability density of $\hat{E}$ has multidimensional normal distribution independent of previous data and is the same for every time t.

$$E[\hat{E}_t \hat{E}_t^T] = \Omega \tag{9}$$

Ω is a constant covariance $K-$dimensional matrix. The task consist of finding the estimation $\hat{Y}$ (3) in dependance of known process history.

$$Y^{(t-1)} = \{Y_{t-1}, Y_{t-2}, \ldots, Y_1\} \tag{10}$$

To construct estimator (3), we need to derive the conditional probability density

$$p(Y_t | Y^{(t-1)}). \tag{11}$$

Using Bayesian estimation theory (Peterka,1981), we can express (11) in the form of Student's distribution

$$p(Y_t|Y^{(t-1)}) = \pi^{-K/2}\Gamma((\gamma(t) - \beta + K + 1)/2)/\{\Gamma((\gamma(t) - \beta + 1)/2)(1 + Z_t^T V_{z(t-1)}^{-1} Z_t)^{K/2} |\lambda_{t-1}|^{1/2} [1 + (Y_t - \hat{P}_{t-1}^T Z_t)^T \lambda_{t-1}^{-1} (Y_t - \hat{P}_{t-1}^T Z_t)/(1 + Z_t^T V_{z(t-1)}^{-1} Z_t)]^{(\tau(t)-\beta+K+1)/2}\} \tag{12}$$

with conditional mean value (3)

$$\tilde{Y}_t = \hat{P}_{t-1}^T Z_t \tag{13}$$

where $\hat{P}_{t-1}$ is estimation (15) of the $Kx\beta$ matrix (14)

$$P^T = [A_1, \ldots, A_N]. \tag{14}$$

$$\hat{P}_{t-1} = V_{z(t-1)}^{-1} V_{zy(t-1)} \tag{15}$$

$$Z_t = [Y_{t-1}^T, \ldots, Y_{t-N}^T]^T \tag{16}$$

is the $\beta x1$ data vector ($\beta = KN$). The following notation was used in (12):

$$V_{t-1} = \tilde{V}_{t-1} + V_N \tag{17}$$

$$\gamma(t-1) = \gamma(N) + t - 1 - N \tag{18}$$

$$\lambda_{t-1} = V_{y(t-1)} - V_{zy(t-1)}^T V_{z(t-1)}^{-1} V_{zy(t-1)} \tag{19}$$

$$\tilde{V}_{t-1} = \begin{bmatrix} \tilde{V}_{y(t-1)} & \tilde{V}_{zy(t-1)}^T \\ \tilde{V}_{zy(t-1)} & \tilde{V}_{z(t-1)} \end{bmatrix} \tag{20}$$

$$\tilde{V}_{y(t-1)} = \sum_{k=N+1}^{t-1} Y_k Y_k^T \tag{21}$$

$$\tilde{V}_{zy(t-1)} = \sum_{k=N+1}^{t-1} Z_k Y_k^T \tag{22}$$

$$\tilde{V}_{z(t-1)} = \sum_{k=N+1}^{t-1} Z_k Z_k^T \tag{23}$$

V_N is a positive definite matrix and

$$\gamma(N) > N(1+K) - 2 \tag{24}$$

4 NUMERICAL REALIZATION

The predictor (13) can be evaluated using matrix V_t (17) updating and its following inversion. Another possibility is direct updating of $\hat{P}_t$. According to work (Peterka,1981), to ensure the numerical stability of solution , it is advantageous to calculate (15) by the means of the square-root filter REFIL (Peterka,1981), which guarantees the positive definiteness of matrix (17). The filter REFIL updates directly the Cholesky square root of the matrix V_t^{-1} .

The numerical complexity of proposed classifier is larger than the conventional per-point Bayesian one. If we denote the number of arithmetic operations necessary to classify one pixel $h(.)$, then the Bayesian classifier in its most efficient version needs:

$$h(*) = Kd(d+3)/2 \qquad h(+) = K(d-1)(d+2)/2 + 2K$$

The contextual Bayesian classifier is computationally more demanding:

$$h(*) = Kd(d+3)/2 + 3K^2N + 2K^2N^2 + 8KN + 16K$$

$$h(+) = K(d-1)(d+2)/2 + K + 3K^2N + 1.5K^2N^2 + 2.5KN + n$$

Where n is the number of pixels in thematic map window. To avoid overflow problems the smallest possible single class predictor value (13) was chosen to be 0.001.

5 EXPERIMENTAL RESULTS

The contextual classification algorithm was applied to agricultural type of Thematic Mapper subscene from North Moravia . The comparison was made by the Bayesian per-point classifier. The area studied is large cooperative farm situated in Vizovice Hills. The objective of the study was to determine its land-use, land-cover conditions. Ground areas of homogeneous landforms and land cover conditions were identified through a field study in the time of satellite snap shoting. This study and maps formed the ground truth data base for our field definition and class identification. The digital computations were performed using interactive image analysis system developed by the author. The mean vector and covariance matrix of each training field were calculated to develop the spectral signature representative of land-cover classes. The objective of the analysis, was to discriminate among following 15 agricultural classes: water, red clover, white clover, wood, wheat, maize, mixture, millet, grass, harvested rapeseed, rapessed, sugarbeet, wheat II, clover and residential area. The resubstitution estimations of the probability of correct classification are summarised in the table.

class	Contextual Bayesian c.	Bayesian classifier
1	1	1
2	0.8	0.72
3	0.9	0.82
4	0.95	0.81
5	0.98	0.93
6	1	1
7	0.92	0.79
8	1	0.93
9	0.77	0.56
10	0.97	0.97
11	1	0.89
12	0.99	0.91
13	0.85	0.6
14	0.92	0.88
15	0.95	0.68
P	0.93	0.85

We have chosen the autoregressive model of order one ($N = 1$) and the smallest possible thematic map window of nine pixels ($n = 9$). In such a case, the contextual classifier needs $h(*) = 2010$, $h(+) = 1479$ operations, while the conventional approach only $h(*) = 525$, $h(+) = 435$ ones.

6 CONCLUSION

The tested example shows improvement in classification performance. The overall performance of contextual classifier is better in tested example than the non-context perpoint Bayesian one. Similar improvement is seen for all classes to be searched. This result was reached with the smallest possible context window of eight surrounding pixels. The optimal size of window depends on local conditions (average field area) and on the used image resolution. Computational time of presented algorithm was in tested example approximately five times longer than for the standard Bayesian classification one. Further work is still needed to optimize developed software to increase classification speed.

REFERENCES

Peterka,V.,1981.Bayesian approach to system identification. In. Trends and Progress in System Identification. Ed. P. Eykhoff , Pergamon Press, Oxford.

Tubbs,J.D.,1980. Effect of Autocorrelated Observation on Confidence Sets Based upon Chi-Square Statistics. IEEE Trans. Syst. Man. and Cybern., vol. SMC-10(4): 177-180.