

## Application of maximum likelihood classification Based on minimal risk in crop interpretation

Yubin Song<sup>1,a</sup>, Xiufeng Yang<sup>1,a</sup>, Yancang Wang<sup>1,a</sup>, Zihui Zhao<sup>1,a</sup>,  
Xuhong Ren<sup>1,a</sup>, Longfang Duan<sup>1,a</sup>

<sup>1</sup>North China Institute of Aerospace Engineering, Langfang 065000, China

<sup>a</sup>ssyybb2006@126.com

**Keywords:** classification, minimal risk, crop, interpretation.

**Abstract.** In crop interpretation by remote sensing, Gray distribution of crop is overlapped in some intervals. The non-target crops fall into the target crop, which would greatly increase the workload in post classification. To reduce these classification errors, and improve accuracy of clarification, maximum likelihood classification based on minimal risk is used. And the relationship between extraction rate and accuracy were analyzed. Experiments show that this method can improve the accuracy of extracting target crops, reduce the workload of the post classification, and improve efficiency.

### Maximum Likelihood Classification Based on Minimal Risk

Theory of maximum likelihood classification Based on minimal risk is different from maximum likelihood classification Based on minimum error rate by the classification principles. Both methods use priori probability and probability density function to get posterior probability, which is used to classify pixels. Supposes are as follows:

(1) There are  $n$  bands in remote sensing image, each pixel can be described by  $\mathbf{x}, \mathbf{x} = (x_1, x_2, \dots, x_n)^T$

(2) Remote sensing image would be classified into two classes ( $w_1$  and  $w_2$ ).

(3) The class-conditional probability density functions  $p(\mathbf{x} | w_i)$ ,  $i=1, 2$ , are described as the distribution of the feature vectors in each of the classes.

The rules of maximum likelihood classification based on minimum error rate can be stated as:

$$\begin{cases} p(w_1 | \mathbf{x}) > p(w_2 | \mathbf{x}), \mathbf{x} \in w_1 \\ p(w_2 | \mathbf{x}) > p(w_1 | \mathbf{x}), \mathbf{x} \in w_2 \end{cases}$$

Risk is introduced to influence the rules of maximum likelihood classification based on minimum error rate. by the significance of the risk, we can get risk as follows.

Table 1 decision and risk

By $\alpha_1$ , risk of classifying $\mathbf{x}$ to $w_1$	By $\alpha_1$ , risk of classifying $\mathbf{x}$ to $w_2$	By $\alpha_2$ , risk of classifying $\mathbf{x}$ to $w_1$	By $\alpha_2$ , risk of classifying $\mathbf{x}$ to $w_2$
$\lambda(\alpha_1, w_2)$	$\lambda(\alpha_1, w_2)$	$\lambda(\alpha_2, w_1)$	$\lambda(\alpha_2, w_2)$

Through table 1, the risk will be clear when the decision is made. by the physical meaning of the decision, if take decision  $\alpha_1$  and classify  $\mathbf{x}$  to  $w_1$ , the risk will be zero; if take decision  $\alpha_2$  and classify  $\mathbf{x}$  to  $w_2$ , the risk will be zero too, then we can get decision an risk improved as table 2:

Table 2 decision and risk improved

By $\alpha_1$ , risk of classifying $\mathbf{X}$ to $w_1$	By $\alpha_1$ , risk of classifying $\mathbf{X}$ to $w_2$	By $\alpha_2$ , risk of classifying $\mathbf{X}$ to $w_1$	By $\alpha_2$ , risk of classifying $\mathbf{X}$ to $w_2$
0	$\lambda(\alpha_1, w_2)$	$\lambda(\alpha_2, w_1)$	0

or class  $w_j$ , make decision  $\alpha_i$ , the risk will be  $\lambda(\alpha_i, j)$ , then we get the risk with conditions:

$$R(\alpha_i | \mathbf{x}) = \sum_{j=1}^M \lambda(\alpha_i, w_j) p(w_j | \mathbf{x})$$

The rule of maximum likelihood classification Based on minimal risk can be stated as:

$$R(\alpha_k | \mathbf{x}) = \min\{R(\alpha_i | \mathbf{x})\}, i = 0, 1$$

By derivation, the effect to decision of risk can stated as follows:

$$\begin{cases} \frac{\lambda(\alpha_2, w_1)}{\lambda(\alpha_1, w_2)} \leq \frac{p(\mathbf{x} | w_2)}{p(\mathbf{x} | w_1)} \frac{p(w_2)}{p(w_1)}, \mathbf{x} \in w_1 \\ \frac{\lambda(\alpha_2, w_1)}{\lambda(\alpha_1, w_2)} > \frac{p(\mathbf{x} | w_2)}{p(\mathbf{x} | w_1)} \frac{p(w_2)}{p(w_1)}, \mathbf{x} \in w_2 \end{cases}$$

According to the assumption,  $p(w_1)$ ,  $p(w_2)$ ,  $p(\mathbf{x} | w_2)$ ,  $p(\mathbf{x} | w_1)$  are known, so  $p(\mathbf{x} | w_2)p(w_2)/p(\mathbf{x} | w_1)p(w_1)$  is a function of  $\mathbf{x}$ . The range of  $\mathbf{x}$  will be change as the ratio of  $\lambda(\alpha_2, w_1)$  and  $\lambda(\alpha_1, w_2) \cdot p(w_1), p(w_2), p(\mathbf{x} | w_2), p(\mathbf{x} | w_1)$  can be estimated by the sample.

## Test and Analysis

Study area is part of the remote sensing image, which is the integration of the multi-spectral image and panchromatic image. the resolution of multi-spectral image is 2.1 meters, the resolution of panchromatic image is 0.8 meters. The resolution of image fused is 0.8 meters.

The content of this experiment aim to get the information of wheat in remote sensing image fused. Suppose wheat belongs to  $w_1$  and the other belong to  $w_2$ , remote sensing image is classified into  $w_1$  and  $w_2$ .

Result will be different with change of the risk. When the ratio of  $\lambda(\alpha_2, w_1)$  and  $\lambda(\alpha_1, w_2)$  varies from 1 to 7, the result will be like as follows:

Two types of errors were defined as follows:

- (1) the first type of error:  $\mathbf{x}$  is assigned to  $w_1$ , but  $\mathbf{x} \in w_2$
- (2) the second type of error:  $\mathbf{x}$  is assigned to  $w_2$ , but  $\mathbf{x} \in w_1$

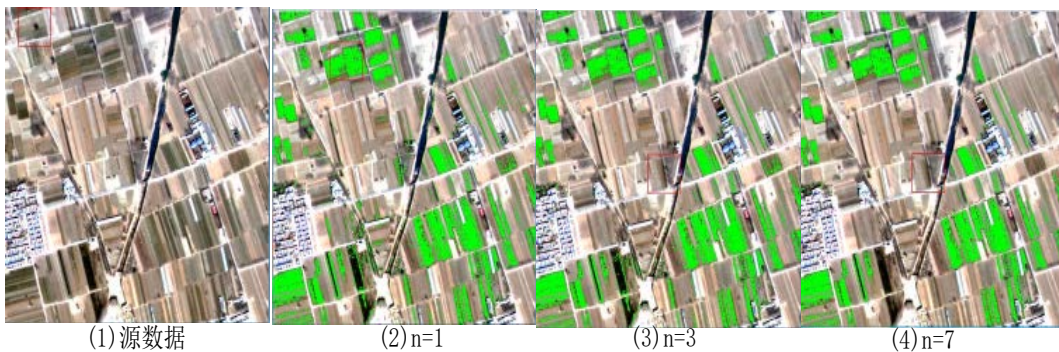


Fig1 classification result of remote sensing image

From figure 1 we can know that the second error is reduced as n vary from 1 to 7, but the first type of error is increased. at the same, if the first type of error is reduced, the second type of error

will be increased.

Our aim is to get the information of wheat in remote sensing image, so we hope that the second type of error as small as possible. But as the second type of error reduce, the first type of error will be increased, and the number of wheat pixels will be reduced. To solve this problem, the concept of accuracy and extraction rate is proposed.

### Accuracy and Extraction Rate

If the second type of error decreases, the number of wheat pixels will be reduced. Only part of wheat pixels will be got. For this problem we define extraction rate as follows:

$$\text{extraction rate} = \frac{s_{\text{classification}}}{s_{\text{real}}}$$

(1)  $s_{\text{classification}}$  is the number of pixels in result of wheat by classification.

(2)  $s_{\text{real}}$  is the number of pixels of wheat by visual interpretation.

At the same time, we hope that there is few other type of pixels in result of wheat pixels. If there are many other type of pixels in result of wheat, the workload of post classification will be greatly increased.

The concept of accuracy is proposed to describe condition above, accuracy is defined as follows:

$$\text{accuracy} = \frac{s_{\text{real}}^{\text{wheat}}}{s_{\text{classification}}}$$

(1)  $s_{\text{real}}^{\text{wheat}}$  is the number of wheat pixels in result of wheat classification by visual interpretation.

(2)  $s_{\text{classification}}$  is the number of pixels in result of wheat classification.

According to the change of the rate of  $\lambda(\alpha_2, w_1)$  and  $\lambda(\alpha_1, w_2)$ , a set of points can be got. Scatter plot can be drawn by this set of points and then we can get Fig2 as follows:

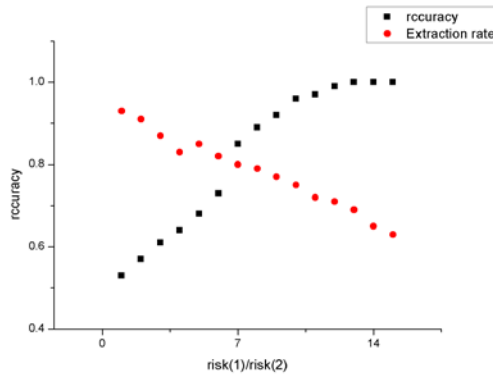


Fig 2 relationship between accuracy and extraction rate

For extraction rate of wheat, extraction rate will reduce if  $\lambda(w_2, \alpha_1) / \lambda(w_1, \alpha_2)$  reduce, so extraction rate and accuracy should be weighed in the interpretation of wheat. In this experiment, we can see that accuracy will be very close to 100% as  $\lambda(w_2, \alpha_1) / \lambda(w_1, \alpha_2)$  equal 13. If continue to increase  $\lambda(w_2, \alpha_1) / \lambda(w_1, \alpha_2)$ , addition of accuracy is very limited but extraction rate will reduce greatly. So continue to increase  $\lambda(w_2, \alpha_1) / \lambda(w_1, \alpha_2)$  is mean-less for interpretation of wheat.

x -coordinate is  $\lambda(w_2, \alpha_1) / \lambda(w_1, \alpha_2)$ , the black-line is accuracy, the red-line is extraction rate. we can know relationship between accuracy and extraction rate. By the following way we can get solution both satisfy accuracy and extraction rate. Proceed as follows:

(1) Compute  $\text{extraction rate} \geq T_1$  and get solution set  $A_1$

(2) Compute  $\text{accuracy} \geq T_2$  and get solution set  $A_2$

(3) The final solution set is  $A = A_1 \cap A_2$

## Conclusions

By studying we can get conclusions as follows:

(1) The accuracy can be increased by maximum likelihood classification Based on minimal risk. But extraction rate will be decrease as accuracy increase.

(2) By setting thresholds of extraction rate and accuracy and computing intersection of two solution sets we can get solution set both satisfy extraction rate and accuracy.

## Reference

- [1]. Chen Jing-bo, LIU Shun-xi, WANG Cheng-yi, et al. Research on Urban Water Body Extraction Using Knowledge-based Decision Tree. Remote sensing information. Vol. 28 (2013) No. 1, p. 29-33.
- [2]. WANG Wei-hong, HE Min, et al. Multi-scale segmentation in land-use information extraction based on object-oriented method. Science of surveying and mapping. Vol. 36 (2011) No. 4, p. 160-161.
- [3]. Sergios theodoridis, et al. Patter Recognition. China Machine Press, 2009, p. 13-86.