An Empirical Research of Multi-Classifier Fusion Methods and Diversity Measure in Remote Sensing Classification

Hongchao Ma
SRSAIE, Wuhan Univ. Wuhan Hubei, 430079, China

Wei Zhou
SRSAIE, Wuhan Univ. Wuhan Hubei, 430079, China
zhouwei_prc@yahoo.com.cn

Xinyi Dong
LIESMARS, Wuhan Univ. Wuhan Hubei, 430079, China

Honggen Xu
SRSAIE, Wuhan Univ. Wuhan Hubei, 430079, China

Abstract

In this paper, Multi-Classifier System (MCS) is applied to the automatic classification of remote sensing images, and some effective multi-classifier fusion methods with relatively high accuracy are proposed based on substantive experiments. The classification accuracy of MCS has been remarkably improved compared to single classifier with an average increment of 5%. In addition, a diversity measure named EPD is presented, and the paper proves that its ability in predicting the performance of classifiers combining can be used to assist the construction of multiple classifier systems.

1. Introduction

In the technique of pattern recognition for processing remote sensing data, it usually plays an important part for the classifier to classify the image with multi-dimensional characteristic subtracted by the progress of characteristic selection. General algorithms are usually based on one certain classifier. For most cases, one single classifier could hardly perform well in all the sample characteristics. It has been observed that different classifiers result in separate misclassified groups, although one of them might perform better than all the others in a special category, which indicates that different classifiers potentially offer complementary information for the decision-making of classification. Consequently, this complementary information could be utilized to improve the result of classification-- combining the results of different classifier through some pattern to find objects, which means building multiple classifier systems (MCS).

E-FORENSICS 2008, January 21-23, Adelaide, Australia

DOI 10.4108/wkdd.2008.2731

2. Algorithms about linear combination of multiple classifiers

In the field of remote sensing image classification, suggest that M categories are involved in the pattern space: the pattern space $D = C_1 \cup C_2 \cup \cdots C_M$, here $C_i, \forall i \in \wedge = \{1,2,\cdots,M\}$ represent a particular category. Assume that there are K independent classifiers $e_k(k = 1,2,\cdots,K)$ working on pattern space D , corresponding to the input x , each with a classifier output j_k , namely $j_k = e_k(x)$. Algorithms about the linear combination of multiple classifiers, mainly refers to determining the destination of sample x, that is $E(x) = j, j \in \wedge \cup \{M+1\}$, by linear assembling of the outputs of different sub-classifiers' $j_k(k = 1,2,\cdots,K)$. Formula described as follows:

$$E(x) = a_1 * j_1 + a_2 * j_2 + \cdots + a_K * j_K \tag{1}$$

Here, $a_1, a_2, \cdots, a_K$ are weight coefficients.

2.1. The abstract level combination method

Confusion matrix is used as a prior knowledge in this method. Training is implemented for each sub-classifier, which helps to collect the statistics of the performance of identification:

$$CM_k = \begin{bmatrix} n_{11}^{(k)} & n_{12}^{(k)} & \cdots & n_{1M}^{(k)} & n_{1(M+1)}^{(k)} \\ n_{21}^{(k)} & n_{22}^{(k)} & \cdots & n_{2M}^{(k)} & n_{2(M+1)}^{(k)} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ n_{M1}^{(k)} & n_{M2}^{(k)} & \cdots & n_{MM}^{(k)} & n_{M(M+1)}^{(k)} \end{bmatrix}, \quad k = 1,2,\cdots,K \tag{2}$$

Here, M is the number of classes; K is the number of sub-classifiers. $n_{ij}^{(k)}$ is the number of sample pixels which belongs to C_i yet classified as C_j by e_k .

While identified as $j(C_j)$, the probability that the pixel actually belongs to C_i could be described with conditional probability as:

$$p_{ij}^{(k)} = p(D \in C_i | e_k(D) = j) = \frac{n_{ij}^{(k)}}{n_j^{(k)}} = \frac{n_{ij}^{(k)}}{\sum_{i=1}^{M} n_{ij}^{(k)}} \tag{3}$$

Here, $j = 1,2,\cdots,M$.

K confusing matrixes could be obtained for K different classifiers. If K classifiers were selected to perform classification based on the characteristics extracted from target signals, K results would be produced, which means $e_k(D) = j_k (k = 1,2,\cdots, K;\ j_k \in \wedge)$. Combining (2) and (3) would lead to the matrix describing the recognition performance of these classifiers:

$$PM_k = \begin{bmatrix} P_{11}^{(k)} & P_{12}^{(k)} & \cdots & P_{1M}^{(k)} & P_{1(M+1)}^{(k)} \\ P_{21}^{(k)} & P_{22}^{(k)} & \cdots & P_{2M}^{(k)} & P_{2(M+1)}^{(k)} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ P_{M1}^{(k)} & P_{M2}^{(k)} & \cdots & P_{MM}^{(k)} & P_{M(M+1)}^{(k)} \end{bmatrix} \tag{4}$$

(1) Additive combination based on voting method. PM_k is used as prior knowledge for combination of classifiers. That is, vote is defined as $P(x \in C_i \ / e_k(x) = j)$ and the probability of $x \in C_i$ (the number of votes for C_i) is defined as:

$$P_E(x \in C_i) = \sum_{k=1}^{K} P(x \in C_i \ / e_k(x) = j_k) \quad i = 1,2,\cdots,M; \tag{5}$$

Here, M is the number of classes, K is the number of sub-classifiers. Voting rules which use the identification performance as the prior knowledge is defined as follows:

$$E(x) = \begin{cases} j, & if P_E(x \in C_j) = \max_{i \in \wedge} P_E(x \in C_i) \geq T; \\ M+1, & else \end{cases} \tag{6}$$

(2) Multiplicative combination based on Bayesian method. The credibility of the result $x \in C_i$ produced by sub-classifier e_k could be calculated with the following formula:

$$bel(x \in C_i / e_k(x)) = P(x \in C_i / e_k(x) = j_k), \quad \begin{matrix} i = 1,\cdots,M \\ k = 1,\cdots,K \end{matrix} \tag{7}$$

Here, M is the number of classes; K is the number of sub-classifiers.

Suppose the credibility of the classification result: $x \in C_i$ could be written as $bel(i) = \eta \prod_{k=1}^{k} P(x \in C_i / e_k(x) = j_k)$.

Here η is a constant, which meets this condition $\sum_{i=1}^{M} bel(i) = 1$ (since $x \in C_i, i = 1,2,\cdots M$).

Finally, based on the analysis mentioned above, sample x cloud be classified based on the following rules (refusal situation is considered):

$$E(x) = \begin{cases} j, & \text{if} \quad bel(j) = \max_{i \in \wedge} bel(i) \geq \alpha \ ; \\ M+1, & \text{else} \end{cases} \tag{8}$$

Here, $0 < \alpha \leq 1$, which serves as the threshold.

2.2. Measurement level-weight combination method based on output

While classifying remote sensing images, there are generally two forms of the measurement level output of sub-classifier, namely probability vector and distance vector. Therefore, it is required to convert the spectral vector of distance classifier into "probability", which meets the three axiom of probability theory, and comparable with probability vector. This conversion can be realized through the following formula:

$$P_k(X \in C_i \ / X) = \frac{1 \ / d_k(C_i \ / X)}{\sum_{i=1}^{M} 1 / d_k(C_i \ / X)} \tag{9}$$

Here, M is the number of classes, K is the number of sub-classifiers, $d_k(C_i / X)$ is the spectral vector distance between pattern X of the kth classifier and class C_i .

While output vector of each sub-classifier have been converted to consistent posterior probability vector, the output of classifiers can be directly used to calculate the corresponding weight $p(D \in C_l / e_j(x_j) = y_j)$ of classifier e_j while $y_j = e_j(x_j)$ as follow:

$$p(D \in C_l / e_j(x_j) = y_j) = \frac{\exp(\ \|E_j^l - y_j\|^2)}{\sum_{i=1}^{M} \exp(\ \|E_j^i - y_j\|^2)} \tag{10}$$

Here: y_j is the output of classifier e_j; E_j^i is the desired output of e_j with C_i as the input.

When weight coefficients has been calculated, both additive and multiplicative combination could be used, which share the same principles of additive and multiplicative methods in abstract level. It is only required to replace the weight coefficient in abstract level with that in measurement level.

3. Diversity Measure

With increasingly diversity of the assembling techniques, it is crucial to consider how to measure the relationship between sub-classifiers of the multiple classifier system, as well as how to predict their ability of integration.

In our experiments, prior test should be carried out to each base classifier, in order to obtain the statistics of every sub-classifier's classification result about a given remote sensing image, and then the parameters in table1 are calculated. For a multi-classifier system constituted by L sub-classifiers, the EPD method could be implemented as follow:

(1) For every two base classifiers $\mathrm{D_i}$ and $\mathrm{D_j}$, calculate the entropy based on the situation that these two sub-classifiers give correct or false results contemporarily:

$$E_0(i,j) = -p_{i,j}^{00}\log_2 p_{i,j}^{00} - p_{i,j}^{11}\log_2 p_{i,j}^{11} \qquad (11)$$

(2) For every two base classifiers $\mathrm{D_i}$ and $\mathrm{D_j}$, calculate the entropy based on the situation that one of them gives correct results while the other gives false results:

$$E_1(i,j) = -p_{i,j}^{01}\log_2 p_{i,j}^{01} - p_{i,j}^{10}\log_2 p_{i,j}^{10} \qquad (12)$$

(3) Get the final result of EPD based on calculating the average value of all the above parameters:

$$EPD = \frac{2}{L(L-1)}\sum_{i=1}^{L-1}\sum_{j=i+1}^{L}(E_0(i,j)+E_1(i,j)) \qquad (13)$$

Table 1. The right/wrong relationship of sub-classifiers in pair-wise diversity measure

	$\mathrm{D_i}$ is right	$\mathrm{D_j}$ is wrong
$\mathrm{D_i}$ is wrong	$p^{11}(i,j)$	$p^{10}(i,j)$
$\mathrm{D_j}$ is right	$p^{01}(i,j)$	$p^{00}(i,j)$
$p^{11}(i,j)+p^{01}(i,j)+p^{10}(i,j)+p^{00}(i,j)=1$		

4. Experimentation

4.1. Data description

The data used in this experiment is a TM image of Wuhan City in 2002 as figure 1.

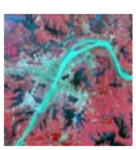

Figure 1. TM image of Wuhan region in 2002

4.2. Experimental process

In this study, the design of this experiment and its main procedure are listed as follow:

(1) Selection of sub-classifiers. In this study, five traditional classifiers which are commonly used in the field of processing remote sensing data includes Maximum Likelihood Classifier (MaxLC), Minimum Distance Classifier (MinDC), Mahalanobis Distance Classifier (MahDC), Spectral Angle Classifier (SpeAC) and ISODATA Unsupervised Classifier (IsoUC).

(2) Diversity measurement. EPD entropy method is applied to measure the diversity of selected sub-classifiers, and average measurement is calculated. If the measurement meets the settled range of threshold, step (3) will be conducted; otherwise, the sub-classifiers should be re-selected.

(3) Combination of multiple classifiers. Combine the output of each sub-classifier in accordance with the rules of the selected combination.

(4) Classification accuracy assessment. In this study, evaluation method based on confusion matrix is selected with the evaluation index for classification accuracy as: producers accuracy, users accuracy, overall accuracy and kappa coefficient.

4.3. Linear combination of multiple classifiers based on the same training samples

4.3.1. The comparison between the result of single classifier and multi-classifier system. Overall classification accuracy and kappa coefficient before and after combination of sub-classifiers are shown in table 2. With ranked precision, it is obvious that accuracies of MCS selected in this study are better than single classifiers, with an increase of 5% on average. It is proved that linear combination of multiple classifiers is good at making use of the complementarities between single sub-classifiers, and result in better classification performance of the combination system.

4.3.2. The comparison of classification results produced by different MCS. Comparing the results, we can get the conclusion that the combining method in the measurement level perform better than the abstract level, which shows an average improvement of 0.7%. In addition, due to make full use of the classification information of each sub-classifier, the additive combination method based on output vectors shows a better stability than other methods, and its classification result is closer to the field survey.

4.3.3. The relationship between a MCS's performance and its diversity value. In this study, different sub-classifiers were selected to perform abstract level classification with a combination of three sub-classifiers, which helps to analyze the effect on classification performance of MCS placed by the diversity of sub-classifiers, and show how the diversity could instruct the selection of sub-classifiers. Accuracy assessment result is listed in Table 3. It could be observed in the table that, classification performance of the combination system deteriorates with an average difference too small or too large. So in the progress of classification with combination of multiple classifiers, sub-classifiers selection should be supervised by setting a suitable range of threshold of the average difference, which is [1.000, 1.205] in this study.

Table 2. The accuracy report of sub-classifiers and MCS combined by three sub-classifiers

No.		Classify Method	Overall Accuracy	Kappa	Rank
1		MaxLC	84.32%	0.827	7
2		MinDC	73.08%	0.683	10
3		MahDC	84.33%	0.828	6
4		SpeAC	75.73%	0.689	9
5		IsoUC	65.53%	0.585	11
MCS 1～3	AL	VM	84.31%	0.827	8
		ACVM	84.82%	0.834	5
		MCBM	87.49%	0.868	1
	ML	BAM	87.29%	0.865	2
		ACOV	85.76%	0.845	3
		MCOV	85.64%	0.844	4

Table 3. The accuracy report and diversity value of different MCSs combined by three sub-classifiers in abstract level

No.	Classify method	Overall Accuracy	Kappa	Rank	Diversity Value
1,	VM	84.31%	0.827	4	
2,	ACVM	84.82%	0.834	3	1.066
3	MCBM	87.49%	0.868	1	
1,	VM	81.04%	0.783	9	
2,	ACVM	84.98%	0.836	2	1.205
4	MCBM	81.19%	0.786	7	
1,	VM	84.31%	0.827	4	
3,	ACVM	81.31%	0.787	6	0.880
4	MCBM	81.10%	0.785	8	
2,	VM	81.04%	0.783	9	
3,	ACVM	80.46%	0.776	11	1.207
4	MCBM	80.10%	0.771	12	

Here, “AL” means the abstract level, “ML” means the measurement level, “VM” means simple voting method, “ACVM” means additive combination based on voting method, “MCBM” means multiplicative combination based on Bayesian method, “BAM” means Bayesian Average Method, “ACOV” means additive combination based on output vectors, “MCOV” means multiplicative combination based on output vectors.

5. Conclusions

Conclusions based on the result are listed as follows:

1) Classification performance could be effectively improved with the utilization of MCS;

2) Additive voting method based on output vector shows significant advantages over other methods;

3) Measurement of diversity can be used to guide the selection of sub-classifiers. Classification accuracy will decrease when the average difference becomes too large or too small. The classification performance of MCS will be significantly improved if difference falls in a certain range of threshold, with a relatively large average difference, and the classification accuracy of single classifiers selected to be combined should also be larger than settled threshold.

Acknowledgements

This paper was granted by the National Basic Research Program of China (973 Program) (No. 2003CB415205).

References

[1] Ching Y. Suen, and Louisa Lam, “Multiple Classifier Combination Methodologies for Different Output Levels [A]”. *Kittler, Roli, Proceedings of First International Workshop on Multiple Classifier Systems*，Springer，2000.

[2] Terry Windeatt, “Diversity measures for multiple classifier system analysis and design”, *Information Fusion 6*, 2005, Page(s):21－36.

[3] Burzzone L, Cossu R, and Prieto D J, “Combining Parametric and Nonparametric Classifiers for an Unsupervised Updating of Land Cover Maps ［A］”. *Kittler, Roli. Proceedings of First International Workshop on Multiple Classifier Systems ［C］*, Springer，2000.

[4] Xu L, Krzyzak A, and Suen C Y, “Methods of Combining Multiple Classifiers and Their Applications to Handwriting Recognition ［ J ］”. *IEEE Transaction on Systems, Man, and Cybernetics*, 1992, Page(s):418-435.

[5] Lam, L, and Suen, C.Y, “A theoretical analysis of the application of majority voting to pattern recognition”, *IEEE Transaction on Systems, Man, and Cybernetics*, Volume 2, 9-13 Oct. 1994, Page(s):418-420.