Comparison Studies on Classification for Remote Sensing Image Based on Data Mining Method

Hang XIAO [1] , Xiubin ZHANG [1]
1: School of Electronic, Information and Electrical Engineering
Shanghai Jiaotong University
No. 1954, Huashan Road, Shanghai
P.R. China
xiaohang@sjtu.edu.cn, http://www.sjtu.edu.cn

Abstract: - Data mining methods have been widely applied on the area of remote sensing classification in recent years. In these methods, neural network, rough sets and support vector machine (SVM) have received more and more attentions. Although all of them have great advantages on dealing with imprecise and incomplete data, there exists essential difference among them. Until now, researches of these three methods have been introduced in lots of literatures but how to combine these theories with the application of remote sensing is an important tendency in the later research. However, all of them have their own advantage and disadvantage. To reveal their different characters on application of remote sensing classification, neural network, rough sets and support vector machine are applied to the area of remote sensing image classification respectively. Comparison result among these three methods will be helpful for the studies on emote sensing image classification. And also the paper provides us a new viewpoint on remote sensing image classification in the future work.

Key-Words: - remote sensing image classification, data mining, neural network, variable precision rough sets model, support vector machine, comparison studies

1 Introduction

Studies of remote sensing image classification have received more and more attention since remote sensing was proposed in 1960's as a synthetical technique which has been used in a wide range of applications such as geosicence, agriculture, and environment and so on. With continual increase of image resolution, more useful data and information can be obtained from remote sensing image. Remote sensing image classification is one of the most important technologies in the field of remote sensing image processing. How to increase precision of classification is a key issue in remote sensing image processing. To extract useful information from remote sensing image, different data mining methods are adopted on this research area to improve the precision .Currently, the major methods in remote sensing image classification are statistic classification method, structural classification method and fuzzy classification method. In these methods, statistic such as mean, variance, standard deviation and discrete degree are taken as criteria to distinguish different categories. All of them need numerous statistic calculation but in a low classification precision. Neural network has been used in remote sensing image classification in these years and gotten a satisfying result. There have plenty of literatures in the field [1][2][3]. However, there also have several weakness in neural network such as slow learning rate, difficult convergence, complex network structure and umbiguous meaning of network.

With the development of research, there appear many new theories in information processing and data mining since 1980's. In these data mining theories, rough sets theory receives a most attraction. However, in remote sensing image classification, few literatures on rough sets in the field are found in the research [4].

Large advantages have been shown in both rough sets theory and neural network on dealing with various imprecise, incomplete information. However, there exists essential difference between them. Rough sets simulate abstract logic mind of our human being while neural networks simulate intuition mind. Rough sets theory express logic rules based on indiscernibility relation and knowledge reduction while neural networks state relation between input and output by using nonlinear mapping. In general, neural networks can not reduce dimensions of inputs. More complex structures and training cost required in neural networks of a higher input dimensions. Rough sets theory can be used to decrease redundance among input information through finding their relations, but rough sets theory is very sensitive to noises. Therefore, the good results derived from sample data may not appear good when they are applied in the set of test data. That is to say, rough sets have a weak error tolerance and generalization performance. Whereas, neural networks have a better capability of anti-noise, self-organization, and generalization [5]. Fortunately, variable precision rough sets (VPRS) model proposed by Ziarko provide us a very useful tool to solve the problem [6]. Variable precision rough sets model allow for some degree of misclassification, which can avoid the high sensitivity of computational results which is necessary to increase the system redundance. This paper mainly introduces application of the variable precision rough sets as an example on remote sensing image classification. In the result, performance of system can be found to achieve much improvement after using the variable precision rough sets.

Besides rough sets theory, support vector machine (SVM) is also a very attractive method in data mining area. However, in remote sensing image classification, SVM is still in its beginning stage the same as rough set. There is even fewer researches on SVM in field of remote sensing image classification [7][8].

SVM is a new data mining and machine learning theory proposed by Vapnik et al. in mid 1990s. It is a universal method to solve multidimensional function. It has been applied some areas such as function simulation, pattern recognition and data classification and obtained a perfect result. There exist some defects in neural network such as determination of network structure, local minima problems, under learning and over learning. All of them restrict the application of neural network. SVM has advantages in solving the problems of non-linear, pattern selected, high dimension, small specimen, which is good complementary with neural network.

This paper is organized as follows. In section II, basic theories of three methods are briefly reviewed. In section III, we apply three methods into the classification of remote sensing image respectively and give the comparative results. Section IV is the analysis of the three methods in which advantage and disadvantage of three methods are analyzed and some suggestions for future research are also presented. The last section is conclusion.

2 Brief Review of Neural network, Rough Sets and Support Vector Machine

2.1 Neural network

Since neural network researches revived in 1980s, substantial progress has been achieved in application as well as in theory. Neural networks have been widely applied in pattern recognition, control optimization, predicting management and so on. In the field of data mining, neural networks have been combined with genetic algorithm, fuzzy sets [9]. Classification is a very important task in area of information processing and knowledge discovery. Classification of neural network is a supervised training algorithm. It has a high tolerance capability and self-organization performance. Lots of works have been done and large numbers of literatures have been introduced in the field of neural networks. Presently, most methods of neural network in remote sensing image classification use BP learning algorithm for supervised learning classification. BP network is a feedforward network which is in fact a nonlinear criterion function[10].

2.2 Variable Precision Rough Sets

Rough sets theory is a new mathematical tool in data mining area to deal with vagueness and uncertainty data, which can analyze and deal with various imprecise and incomplete information. However, traditional rough sets are very sensitive to even small misclassification errors which restrict its application greatly. Hence, it is necessary to increase the system redundance. Here, we mainly introduce the VPRS model. And VPRS is also taken as an example in the following application.

In conventional rough sets, universe U is known and conclusion is only suitable for objects belonging to U. It is very difficult to satisfy the constrains in practice. To solve the problem, a method must be found to generalize conclusions obtained from sample data to a more wide area. VPRS is proposed by W.Ziarko to solve the problem.

Let X , Y be non-empty sets in finite field. If there exist $x \in Y$ for all $x \in X$, we call that $X \subseteq Y$. It is obviously that no misclassification errors are allowed for in the condition. A new idea is presented in VPRS which give a new measurement method on inclusion relation as follows.

$$c(X,Y) = \begin{cases} 1 - card(X \cap Y)/card(X) & if \quad card(X) > 0 \\ 0 & if \quad card(X) = 0 \end{cases} \tag{1}$$

where $card(\bullet)$ denote cardinal number of sets. $c(X,Y)$ denotes degree of misclassification set X into Y . That is to say, there are $c(X,Y)$ *100% elements misclassified. Obviously, $X \subseteq Y$ when $c(X,Y) = 0$. Therefore, we can give an admissible misclassification error $\beta(0 \le \beta \le 0.5)$. According to the definition, there is:

$$Y \overset{\beta}{\supseteq} X \quad \textit{if and only if} \quad c(X,Y) \le \beta \tag{2}$$

then Ziarko proposed definitions as follows,

Suppose that U is universe, R is indiscernibility relation on U. $R^* = \{E_1, E_2, \cdots, E_n\}$ are partitions of equivalent classes on U.

β -lower approximation (β -positive region of set X),

$$\underline{R}_\beta X = \bigcup \{E \in R^* : c(E, X) \le \beta\} \quad (3)$$

β -upper approximation(β -negative region of set X),

$$\overline{R}_\beta X = \bigcup \{E \in R^* : c(E, X) < 1 - \beta\} \quad (4)$$

β -boundary region,

$$BNR_\beta X = \bigcup \{E \in R^* : \beta < c(E, X) < 1 - \beta\} \quad (5)$$

β -negative region,

$$NEGR_\beta X = \bigcup \{E \in R^* : c(E, X) \ge 1 - \beta\} \quad (6)$$

Ziarko gives a very important definition in VPRS, namely quality of classification.

$$\gamma(P, Q, \beta) = card(POS(P, Q, \beta)) / card(U) \quad (7)$$

in which $POS(P, Q, \beta)$ is a β -positive region on partition Q^*.

Attribute reduction and optimal set of attribute are the most important conception in rough sets model. VPRS provide us two important criteria [6],

1. $\gamma(P, Q, \beta) = \gamma(RED(P, Q, \beta), Q, \beta)$
2. No attribute can be eliminated from $RED(P, Q, \beta)$ without affecting the requirement1.

There have been many algorithms for attribute reduction. Optimal reduction can be derived from combined minimum cost criterion naturally if it is possible to assign a cost function to attributes. In the absence of attribute cost function, two basic approaches were presented by Ziarko in which optimal reduction can be determined according to the number of attributes and rules [6].

2.3 Support Vector Machine

Support vector machine is a new machine learning theory proposed in mid 1990s which is arised from the statistic learning theory founded by the Vapnik research groups in 1960s. It has been successfully applied to function simulation, pattern recognition and data classification[11].

2.3.1 Linearly Separable SVM

Suppose that there are n samples vectors $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$ belonging to two separate classes where $y_i \in \{+1, -1\}$. Our target is to build a criterion function to separate the two classes.

If there exists a hyperplane $\boldsymbol{w} \bullet \boldsymbol{x} + b = 0$ which makes,

$$\begin{cases} (\boldsymbol{w} \bullet \boldsymbol{x_i}) + b \ge 1, & y_i = 1 \\ (\boldsymbol{w} \bullet \boldsymbol{x_i}) + b \le 1, & y_i = -1 \end{cases} \text{equivalent to}$$

$$y_i[(\boldsymbol{w} \bullet \boldsymbol{x_i}) + b] - 1 \ge 0 \quad (8)$$

We can obtain the optimizing problem as follows,

$$\min \quad \frac{1}{2}\|\boldsymbol{w}\|^2 \quad (9)$$

Then Lagrange function can be defined as below,

$$L(\boldsymbol{w}, b, \alpha) = \frac{1}{2}(\boldsymbol{w} \bullet \boldsymbol{w}) - \sum_{i=1}^{n} \alpha_i \{y_i[(\boldsymbol{w} \bullet \boldsymbol{x}_i) + b] - 1\} \quad (10)$$

where α_i are the Lagrange multipliers. (10) can be transformed to its dual problem in order to minimize the equation. According to Kühn-Tucker condition, we can obtain the optimal classification function,

$$f(x) = \mathrm{sgn}\{(\boldsymbol{w}^* \bullet \boldsymbol{x}) + b^*\} = \mathrm{sgn}\{\sum_{i=1}^{n} \alpha_i^* y_i (\boldsymbol{x_i} \bullet \boldsymbol{x}) + b^*\} \quad (11)$$

sgn is the symbolic function.

2.3.2 Linearly Non-separable SVM

Separable criterion function is built on the Euclidean distance, that is $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{x}_i \bullet \boldsymbol{x}_j$. For non-linear problem, sample x can be mapped onto a higher dimensional space and use the linear classifier on it. That is, we make the transformation on $\boldsymbol{x}$, $\Phi: \boldsymbol{R} \to \boldsymbol{H}$.

$$\boldsymbol{x} \to \Phi(\boldsymbol{x}) = (\phi_1(\boldsymbol{x}), \phi_2(\boldsymbol{x}), \cdots) \quad (12)$$

then the criterion function is changed into:

$$f(x) = \mathrm{sgn}\{\boldsymbol{w}^* \bullet \Phi(\boldsymbol{x}) + b^*\} = \mathrm{sgn}\{\sum_{i=1}^{n} \alpha_i^* y_i \, \Phi(\boldsymbol{x_i}) \bullet \Phi(\boldsymbol{x}) + b^*\} \quad (13)$$

Therefore, (11) is only related to inner product of training samples: $(\boldsymbol{x}_i \bullet \boldsymbol{x}_j)$. Thus, only the inner product calculations are needed in the higher dimensional space which can be realized using the function in initial space. Introduction of kernel function make us solve the problem on the input space without calculating the non-linear mapped patterns Φ explicitly. Therefore, when we construct the classification function, we make the comparison on the input space and make the non-linear transformation to results. Thus enormous work will be finished on input space instead of on higher dimensional space.

Thus its classification function is obtained,

$$f(x) = \operatorname{sgn}\{\sum_{i=1}^{n} \alpha_i^* y_i K(\boldsymbol{x_i} \bullet \boldsymbol{x}) + b^*\} \qquad (14)$$

This is called support vector machine.

Complexity of constructing SVM relies on the number of support vector instead of dimension of eigenspace. Different algorithms are formed using different kernel functions. The kernel functions in common use are Polynomial function, Radial Basis Function (RBF), Multi-layer Perceptron and so on that are shown as follows,

(1) Polymonial kernel:

$$\boldsymbol{K}(\boldsymbol{x}, \boldsymbol{x}_i) = [(\boldsymbol{x}, \boldsymbol{x}_i) + 1]^d$$

(2) RBF kernel: $\boldsymbol{K}(\boldsymbol{x}, \boldsymbol{x}_i) = \boldsymbol{e}xp\left\{-\gamma \left|\boldsymbol{x} - \boldsymbol{x}_i\right|^2\right\}$

(3) Sigmoid kernel:

$$\boldsymbol{K}(\boldsymbol{x}, \boldsymbol{x}_i) = \tanh(\gamma\ \boldsymbol{x_i}^{\boldsymbol{T}} \boldsymbol{x}_j - \Theta)$$

(4) Spline kernel :

$$\boldsymbol{K}(\boldsymbol{x}, \boldsymbol{x}_i) = 1 + (\boldsymbol{x} \bullet \boldsymbol{x}_i) + \frac{1}{2}(\boldsymbol{x} \bullet \boldsymbol{x}_i)(\boldsymbol{x} \wedge \boldsymbol{x}_i) - \frac{1}{6}(\boldsymbol{x} \wedge \boldsymbol{x}_i)^3$$

where $(\boldsymbol{x} \wedge \boldsymbol{x}_i) = \min(\boldsymbol{x}, \boldsymbol{x}_i)$

(5) Fourier kernel:

$$\boldsymbol{K}(\boldsymbol{x}, \boldsymbol{x}_i) = \frac{\sin \frac{2N+1}{2}(\boldsymbol{x} - \boldsymbol{x}_i)}{\sin \frac{\boldsymbol{x} - \boldsymbol{x}_i}{2}}$$

3 Empirical Comparison Studies of Neural Network, VPRS and SVM on Remote Sensing Image Classification

Neural network, rough set and support vector machine are briefly introduced in above sections. From the introduction, we can find all of them have their own features on data classifications.

To reveal their different performance, here, these methods are applied on classification for the remote sensing image respectively. Also, empirical comparisons are extracted from the experiment results to describe the difference among different neural network, and VPRS with different parameters and SVM with different kernel functions.

Experimental data of remote sensing image are obtained from the Ladsat image of someplace in China. The data include 1, 2, 3,4,5,7 six bands which correspond to blue, green, red, near infrared, infrared and far infrared band. Their relative images can be found from Fig.1 (a) to Fig.1 (f). Illustration of bands is shown in table 1. The image is 7351×6501×6. Objects in the image are classified by

Fig.1(a)

Fig.1(b)

Fig.1(c)

Fig.1(d)

Fig.1(e)

Fig.1(f)

Fig.1(a)~(f) are obtained from band 1,2,3,4,5 and 7 respectively

Table 1. Bands of remote sensing image

Bands	Blue	Green	Red	Near Infrared	Infrared	Far Infrared
wavelength(μm)	0.4787	0.5610	0.6614	0.8340	1.6500	2.2080

artificial method into five categories: river, house, forest, grass and road. Gray scale is used as features for classification [12]. To avoid noise disturbance, 30×30 image blocks are extracted from different band images in the same place. Calculate the mean gray value μ of the block of each band respectively. Therefore, for each samples, a mean vector gray can be obtained which denoted as $(\mu_1, \mu_2, \mu_3, \mu_4, \mu_5, \mu_7)$. 120 samples are selected from each class in which 80 is for training and the other 40 for testing. There are total 600 data selected for the experiment.

Samples are trained and tested by using neural network, rough sets and SVM respectively. For neural networks, we use the ordinary BP network (BPNN) and improvement BP network trained by Levenberg-Marquart algorithm (BPLM). For VPRS, we select the 0, 0.3 and 0.45 respectively as misclassification error. For support vector machine, RBF, spline function, bspline, anovaspline and sigmoid are choosed as the kernel function of support vector machine. Because classification of remote sensing image in the paper is a multi-classes problem while SVM mainly solve two-class problem, the multilevel classification strategy in the paper [7] is adopted to solve this problem. In paper [7], multi-classes problem can be transferred to many sub two-class problems by using multilevel method.

The experiment is worked on the Pentium IV 1.5G CPU, 512M memory PC. The results are shown in table 2 and table 3.

From talbe2 and table3, time consumed is the longest in BPLM, although precision is the highest in it. BPNN is not convergent in the experiment. The results show defects of neural network in remote sensing image classification. Higher dimensions of input information will lead to complex network

Table 2. Result of classification bv different methods

Method		River	House	Forest	Grass	Road
Neural Network	**BPNN**	**not convergent**				
	BPLM	79.1%	88.3%	88.4%	64.3%	88.2%
Rough Logic	**β=0**	55.1%	50.0%	70.7%	1.2%	60.0%
	β=0.3	55.5%	61.3%	71.3%	17.1%	69.2%
	β=0.45	60.0%	66.1%	75.3%	26.1%	77.7%
SVM	**RBF**	95.1%	60.3%	65.5%	95.0%	51.7%
	spline	95.7%	80.7%	70.0%	95.3%	71.8%
	bspline	94.7%	65.5%	65.8%	68.8%	75.1%
	anovaspline	96.1%	80.3%	55.9%	95.1%	55.4%
	sigmoid	60.1%	20.2%	35.4%	65.3%	39.3%

Table 3. Precison of classification

Method		Training Time(s)	Precision	
			Training Sample	**Test Sample**
Neural Network	**BPNN**	**not convergence**		
	BPLM	3577	98.4%	82.1%
Rough Logic	**β=0**	150	99.6%	47.4%
	β=0.3	151	98.8%	54.8%
	β=0.45	154	98.8%	61.0%
SVM	**RBF**	261	98.7%	73.6%
	spline	32	99.8%	82.7%
	bspline	70	99.1%	74.0%
	anovaspline	37	99.2%	76.6%
	sigmoid	110	54.5%	44.2%

Note: for the rough logic, training time is time of constructing the rough set rules

structure and long time training. In VPRS, short time are consumed on constructing the rough set rules. Good results are obtained in the training samples. However, the precision of classification to test data are not ideal which also account for the limitation of rough sets. We find that although performance of system is improved to some extent through changing β from 0 to 0.45, the capacity of recognition can not be improved essentially because of the defect in rough sets.

From the synthetic performance of the three methods, we find the SVM get the best results. In SVM, except for the sigmoid kernel, all the kernel functions obtain a better result. It spent shorter time in SVM than rough sets. The precision is high and it is even higher than neural networks when using spline kernel. All of them show a bright future of SVM on remote sensing image classification.

4 Analysis

In the paper, we apply neural network, variable precision rough sets and support vector machine to remote sensing image classification respectively. Advantages and disadvantages of the three methods are analyzed and summarized in detailed as belows.

Precision of classification is high using neural network. However, the training time is long and it probably exists local minima.

For multi-input system like remote sensing image, input vectors can be lessen and network structures be simplified by using rough sets. Whereas, neural networks structures will become very complex when increasing the input dimensions. The meaning of rough sets network is very clear once the rough logics are determined, while the number of hidden layer and neurons in neural networks are determined by experience.

In VPRS, some necessary redundance are remained in the process of information reduction which increase the anti-noise performance of system and decrease the loss of the useful information. Furthermore, it provides us a variable parameter for us according to our requirement. Although VPRS obtain more improvement than traditional rough sets in remote sensing image classification, it has also some disadvantages. First, the misclassification error β is not continuous, that is to say, the precision of VPRS is not continuous. Second, selection of optimal attributes still relies on the specific constraints. In the paper, we only simply select them according to [6]. In the specific condition, it does not mean that β is larger and structure of network is more complex. This is related to the number of attribute and logic selected. We can not try to improve the redundance through simply increasing β . Furthermore, besides β in remote sensing image classification, the numbers of equivalent classes also restrict the system performance. The number is more, the precision is higher, but the complexity also become higher and training time become longer. The problem also needs to be discussed in the future research.

In the experiment, we find that high precision and short training time can be obtained using SVM. Comparing with neural network, SVM is more suitable for processing the complex and high dimensional data. However, there are still many problems to be solved in SVM. The performance of SVM largely depends on the kernel. Selection of kernel function limits the application of SVM greatly. From the experiment in section III, the result is not ideal when Sigmoid kernel are used and training time is longer when use RBF kernel although other results are very good. Now, the research of kernel is still at its beginning stage [13]. The classification using SVM still focus on the two classes. How to process the problem of multiclasses classification needs to be studied in the future.

5 Conclusion

Neural network, rough sets and support vector machine are three attractive and effective data mining methods in recent years on dealing with various imprecise and incomeplete data. However, there exists essential difference among them. Except for neural network, until now rough sets and support vector machine are seldom used in the research of remote sensing image classification. In the paper, detail analysis are described which provides us a new viewpoint on remote sensing image classification.

From the experiment, we find network, rough sets and SVM all have their advantages and disadvantages in remote sensing image classification. Therefore, how to combine the three theories and apply them to remote sensing image classification better is an important tendency in the later research. This primary study can be found in the literature [14].

References:

[1] Heermann P D., Khazenie N, *Classification of multispectral remote sensing data using a back propagation neural network,* IEEE Trans on Geoscience and Remote Sensing, Vol.30, No.1, 1992, pp. 81-88.

[2] Atkinson P M, Tatnall A R L, *Neural networks in remote sensing*, Int J Remote Sensing,Vol.8, No.4, 1997, pp. 699-709.

[3] Solares. Cristina, Sanz. Ana Maria, *Bayesian network classifiers. An application to remote sensing image classification*, WSEAS Transactions on Systems, Vol. 4, No. 4, April, 2005, pp. 343-348.

[4] Pal S K, Mitra P, *Multispectral image segmentation using the rough set initialized EM algorithm*, IEEE Transactions on Geoscience and Remote Sensing, Vol.40, No.11, 2002, pp. 2495 -2501.

[5] Mak B, Munakata T, *Rule extraction from expert heuristics: A comparative study of rough sets with neural networks and ID3*, European Journal of Operational Research, Vol.136, No.1, 2002, pp. 212-229.

[6] Ziarko W, *Variable Precision Rough Set Model*, Journal of Computer and System Sciences, 1993(46), pp. 39-59.

[7] Azimi-Sadjadi M R., Zekavat S A, *Cloud classification using support vector machines*, IEEE 2000 International Proceedings on Geoscience and Remote Sensing Symposium, 2000(2), pp. 669-671.

[8] Melgani F, Bruzzone L, *Support vector machines for classification of hyperspectral remote-sensing images*, IEEE International on Geoscience and Remote Sensing Symposium, 2002(1), pp. 506 -508.

[9] Wang W Y, Li Y H, *Evolutionary Learning of BMF Fuzzy-Neural Networks Using a Reduced-Form Genetic Algorithm,* IEEE Transactions on Systems, Man and Cybernetics - Part B: Cybernetics, 2003, pp. 1-11.

[10] Stastny J, Skorpil V, *Neural networks learning methods comparison,* WSEAS Transactions on Circuits and Systems, Vol. 4, No. 4, April, 2005, pp. 325-330.

[11] Pajares G, *Support vector machines for shade identification in urban areas,* WSEAS Transactions on Information Science and Applications, Vol. 2, No. 1, January, 2005, pp. 38-41.

[12] Fisher P F, *Visualization of the Re Liability in classified Remotely Sensed Images,* Photogrammetric Engineering and Remote Sensing, 1994(7), pp. 905-910.

[13] Burges C J C, *Building Locally Invariant Kernels,* In: Proceedings of the 1997 NIPS Workshop on Support Vector Machines, 1998.

[14] Wu Zhaocong, *Research on remote sensing image classification using neural network based on rough sets,* 2001 International Conference on Info-Tech and Info-Net - ICII 2001, 2001(1), pp.279 -284.