Algorithm design and the application for cluster validity based on geometric probability

Jian-Wei Li[1*], Xiao-Wen Li[2]

[1]*College of physics and information engineering, Fuzhou University, Fuzhou 350108*

[2]*College of mathematics and computer science, Longyan University, Longyan 364012*

Received 20 May 2014, www.cmnt.lv

Abstract

Determining optimum cluster number is a key research topic included in cluster validity, a fundamental problem unsolved in cluster analysis. In order to determine the optimum cluster number, this article proposes a new cluster validity function for two-dimension datasets theoretically based on geometric probability. The function makes use of the corresponding relationship between a two-dimension dataset and the related two-dimension discrete point set to measure the cluster structure of the dataset according to the distributive feature of the point set in the characteristic space. It is designed from the perspective of intuition and thus easily understood. Through TM remote sensing image classification examples, compare with the supervision and unsupervised classification in ERDAS and the cluster analysis method based on geometric probability in two-dimensional square, which is proposed in literature 2. Results show that the proposed method can significantly improve the classification accuracy.

Keywords: Cluster validity, Geometric probability, optimum cluster number

1 Introduction

Clustering is an important method of multivariate statistical analysis; it is also an irreplaceable analysis tool in data mining and is widely applied in various research projects [1]. Complete cluster analysis includes: (1) clustering trend analysis; (2) clustering structure extraction; (3) clustering results evaluation. The task of clustering trend analysis is to determine whether a given data set has clustering structure. Clustering structure extraction is a narrow cluster analysis, is the process of applying some algorithm to get the clustering results. Clustering results evaluation is to determine the rationality of clustering results by some standards, whether the clustering results (including levels of classes, the number and the boundaries of classes on each level) are consistent with the clustering structure in the data set. It is also known as cluster validity. Cluster validity function is the basic method to measure the rationality of clustering results [2].

Many clustering structure extraction algorithms have been implemented [3-5], the adapting capabilities of same algorithm for different data sets and different algorithms for same data sets are different, in theory how to evaluate the adapting capabilities of a clustering algorithm and in application how to choose a applicative clustering algorithm for a particular data set is a problem must be solved. For this reason cluster validity is highly valued by scholars, much research focus on it. Currently many literatures preassign a range of optimum cluster number, then use exhaustive method to select the optimum cluster number. It is only fit for the data set that has less data. The method is lack of feasibility in the spatial cluster, which involves huge numbers of the objects (such as remote sensing images' automatic classification). This article proposes and implements a cluster validity function for huge data volume theoretically based on geometric probability [1].

2 Clustering validity classification based on geometric probability

2.1 ANALYTIC EXPRESSION OF THE FUNCTION

Assuming the minimum value and the maximum value of the first term's attribute is x_{min} and x_{max}, respectively; the minimum value and the maximum value of the second term's attribute is y_{min} and y_{max}, respectively, and let:

$$a = Min[(x_{\max} - x_{\min}),(y_{\max} - y_{\min})], \tag{1}$$

$$b = Max[(x_{\max} - x_{\min}),(y_{\max} - y_{\min})]. \tag{2}$$

Then S is mapped to a discrete point set (still named S) in a rectangular region of a*b. Based on the structural information indicator function and the structural information extraction algorithm about clustering point mode in square region, which is mentioned in reference [13], we design a function to measure the structural information of the point set's geometric distribution by geometric probability.

Do paired connection of n points in this discrete point set S and get a set of line segments: $L(s_i;r_i;\theta_i;i=1,2,3,...)$, where s_i, r_i, θ_i denote midpoint coordinate, length and angle between vertical axis of the i^th line segment in L , respectively, and

* ***Corresponding author*** e-mail: lwticq@163.com

$$0 < r_i \le \sqrt{a^2+b^2}, 0 \le \theta_i \le \pi, i = 1,2,3,\dots . \tag{3}$$

For L, construct H Function:

$$H(\theta,\Delta\theta,r,\Delta r) = \frac{N}{T} \times \frac{1}{P(\theta,\Delta\theta,r,\Delta r)} . \tag{4}$$

Here, T denotes the total number of line segments in L, and

$$T = C_n^2 = n(n-1)/2 . \tag{5}$$

N denotes the number of the line segments in L, which satisfies:

$$\left[\theta - \frac{\Delta\theta}{2}, \theta + \frac{\Delta\theta}{2}\right] \cap \left[r - \frac{\Delta r}{2}, r + \frac{\Delta r}{2}\right]. \tag{6}$$

Function $P(\theta,\Delta\theta,r,\Delta r)$ means the probability of the line segments satisfy

$$\left[\theta - \frac{\Delta\theta}{2}, \theta + \frac{\Delta\theta}{2}\right] \cap \left[r - \frac{\Delta r}{2}, r + \frac{\Delta r}{2}\right] \tag{7}$$

in random in rectangle region whose lengths of sides are $a,b(a \le b)$.

Function $H(\theta,\Delta\theta,r,\Delta r)$ means the ratio of the frequency of the number of the line segments in Set L is in the interval

$$\left[\theta - \frac{\Delta\theta}{2}, \theta + \frac{\Delta\theta}{2}\right] \cap \left[r - \frac{\Delta r}{2}, r + \frac{\Delta r}{2}\right] \tag{8}$$

and the probability of line segment is in the same interval in completely random.

So we expect

$$E\left[H(\theta,\Delta\theta,r,\Delta r)\right] = 1. \tag{9}$$

If $r = \frac{\sqrt{a^2+b^2}}{2}, \Delta r = \sqrt{a^2+b^2}$, then

$$P(\theta,\Delta\theta,r,\Delta r) = P\left(\theta,\Delta\theta,\frac{\sqrt{a^2+b^2}}{2},\sqrt{a^2+b^2}\right) = U(\theta,\Delta\theta) . \tag{10}$$

So

$$H(\theta,\Delta\theta,r,\Delta r) = \frac{N}{T} \times \frac{1}{U(\theta,\Delta\theta)} = H(\theta,\Delta\theta) \tag{11}$$

and we expect

$$E\left[H(\theta,\Delta\theta)\right] = 1 . \tag{12}$$

Because the structural information of points set S is stored in Set L, while the points in S gather to multi clusters, the line segments in L must be clustering. It means in completely random the value in some interval is greater than the value. The Figure of $H(\theta,\Delta\theta)$ function will get some peak value in the corresponding interval, the difference between the peak value and 1 indicates the level of clustering, and the number of the peak value indicates the number of clustering directions. A clear corresponding relation exists between the geometric distribution of the points set and the feature of the data set, also exists between Figure of $H(\theta,\Delta\theta)$ function and the structure feature of the point set, so this function is a cluster validity function, which can judge the cluster structure and the validity of clustering results in the data set. The difficulty in process is to derive the analytical expression of $P(\theta,\Delta\theta,r,\Delta r)$ function.

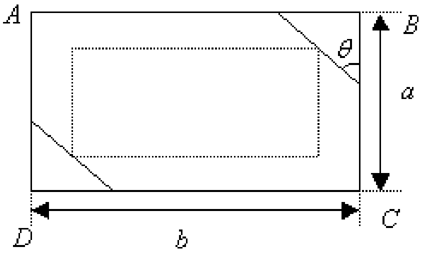

FIGURE 1 The measurement of line segments in the specific intervals and within a rectangle

As shown in Fig.1, if the side lengths of Rectangle $ABCD$ are $a,b(a \le b)$, $P(\theta,\Delta\theta,r,\Delta r)$ indicates the probability of the line segments got from a given rectangle region are in the interval:

$$\left[\theta - \frac{\Delta\theta}{2}, \theta + \frac{\Delta\theta}{2}\right] \cap \left[r - \frac{\Delta r}{2}, r + \frac{\Delta r}{2}\right], \tag{13}$$

$$\begin{aligned} &\int_0^a\int_0^\pi A(r,\theta)drd\theta + \int_a^b \int_{\arccos\frac{a}{r}}^{\pi-\arccos\frac{a}{r}} A(r,\theta)drd\theta + \int_b^{\sqrt{a^2+b^2}} \int_{\arccos\frac{a}{r}}^{\arcsin\frac{b}{r}} A(r,\theta)drd\theta \\ &= \frac{1}{3}a^3 + \frac{1}{3}b^3 + ab^2\ln\frac{\sqrt{a^2+b^2}+a}{b} + a^2b\ln\frac{\sqrt{a^2+b^2}+b}{a} - \frac{1}{3}(a^2+b^2)\sqrt{a^2+b^2} \end{aligned} , \tag{14}$$

$$\begin{aligned} &\int_{r_1}^{r_2}\int_{\theta_1}^{\theta_2} (b - r\sin\theta)(a - r\cos\theta)drd\theta \\ &= ab(\theta_2-\theta_1)(r_2-r_1) + \frac{1}{6}(r_2^3-r_1^3)(\sin^2\theta_2 - \sin^2\theta_1) - \frac{1}{2}(r_2^2-r_1^2)(b\sin\theta_2 - b\sin\theta_1 - a\cos\theta_2 + a\cos\theta_1) . \\ &= M(a,b,r_1,r_2,\theta_1,\theta_2) \end{aligned} \tag{15}$$

We measure the total of the line segment by measuring the area of the region where the points of line segments are included in. For a line segment whose length is $r(0<r\le\sqrt{a^2+b^2})$ and angle between Side CB in positive direction is $\theta(0\le\theta\le\pi)$, the probability of the line segment is in the Rectangle $ABCD$ is proportional to the area of region where the points of line segments are in(the rectangle with dashed lines in Fig. 4):

$$A(r,\theta)=(a-r|\cos\theta|)(b-r|\sin\theta|). \tag{16}$$

Let

$$r_1=r-\frac{\Delta r}{2},r_2=r+\frac{\Delta r}{2};\theta_1=\theta-\frac{\Delta\theta}{2},\theta_2=\theta+\frac{\Delta\theta}{2}, \tag{17}$$

Then

$$P(\theta,\Delta\theta,r,\Delta r)=P(\frac{\theta_1+\theta_2}{2},\theta_2-\theta_1,\frac{r_1+r_2}{2},r_2-r_1)=P, \tag{18}$$

$$P=\frac{\int_{r_1}^{r_2}\int_{\theta_{\min}}^{\theta_{\max}}A(r,\theta)drd\theta}{\int_0^a\int_0^\pi A(r,\theta)drd\theta+\int_a^b\int_{\arccos\frac{a}{r}}^{\pi-\arccos\frac{a}{r}}A(r,\theta)drd\theta+2\int_b^{\sqrt{a^2+b^2}}\int_{\arccos\frac{a}{r}}^{\arcsin\frac{a}{r}}A(r,\theta)drd\theta}. \tag{19}$$

Here

$$\theta_{\min}=f(r_1,r_2)\ge\theta_1,\theta_{\max}=g(r_1,r_2)\le\theta_2. \tag{20}$$

2.2 NUMERATOR VALUE OF THE FUNCTION

(1) $0\le r_1<r_2\le a$

If r_1,r_2 are in the interval $[0,a]$, $f(r_1,r_2)=\theta_1,g(r_1,r_2)=\theta_2$,

$$\begin{aligned}&\int_{r_1}^{r_2}\int_{\theta_{\min}}^{\theta_{\max}}A(r,\theta)drd\theta=\int_{r_1}^{r_2}\int_{\theta_1}^{\theta_2}A(r,\theta)drd\theta\\&=\int_{r_1}^{r_2}\int_{\theta_1}^{\theta_2}(a-r|\cos\theta|)(b-r|\sin\theta|)drd\theta\end{aligned}. \tag{21}$$

Note

TABLE 1 The value of numerator in P Function

Angle	the numerator value
$0<\theta_1<\theta_2\le\frac{\pi}{2}$	$M(a,b,r_1,r_2,\theta_1,\theta_2).$
$0<\theta_1<\frac{\pi}{2}<\theta_2\le\pi$	$M(a,b,r_1,r_2,\theta_1,\frac{\pi}{2})+M(a,b,r_1,r_2,\pi-\theta_2,\frac{\pi}{2}).$
$\frac{\pi}{2}\le\theta_1<\theta_2\le\pi$	$M(a,b,r_1,r_2,\pi-\theta_2,\pi-\theta_1)$

(2) $a\le r_1<r_2\le\sqrt{a^2+b^2}$

Firstly calculate three integral expression, let $0<a\le l_1<l_2\le\sqrt{a^2+b^2},0\le\theta_1<\frac{\pi}{2},0<\theta_2\le\frac{\pi}{2}$, then

1. $$\begin{aligned}&\int_{l_1}^{l_2}\int_{\arccos\frac{a}{r}}^{\arcsin\frac{b}{r}}A(r,\theta)drd\theta\\&=\int_{l_1}^{l_2}\int_{\arccos\frac{a}{r}}^{\arcsin\frac{b}{r}}(a-r\cos\theta)(b-r\sin\theta)drd\theta\end{aligned}, \tag{22}$$

$$\begin{aligned}&=abl_2(\arcsin\frac{b}{l_2}-\arccos\frac{a}{l_2})-abl_1(\arcsin\frac{b}{l_1}-\arccos\frac{a}{l_1})+\frac{1}{2}ab^2\ln\frac{l_2+\sqrt{l_2^2-b^2}}{l_1+\sqrt{l_1^2-b^2}}+\frac{1}{2}a^2b\ln\frac{l_2+\sqrt{l_2^2-a^2}}{l_1+\sqrt{l_1^2-a^2}}+\frac{1}{2}b(l_2\sqrt{l_2^2-a^2}-l_1\sqrt{l_1^2-a^2})\\&-\frac{1}{2}(a^2+b^2)(l_2-l_1)-\frac{1}{6}(l_2^3-l_1^3)+\frac{1}{2}a(l_2\sqrt{l_2^2-b^2}-l_1\sqrt{l_1^2-b^2})\\&=L(a,b,l_1,l_2)\end{aligned}, \tag{23}$$

2. $$\begin{aligned}&\int_{l_1}^{l_2}\int_{\arccos\frac{a}{r}}^{\theta_2}A(r,\theta)drd\theta=\int_{l_1}^{l_2}dr\int_{\arccos\frac{a}{r}}^{\theta_2}d\left[ab\theta-r(b\sin\theta-a\cos\theta)+\frac{1}{2}r^2\sin^2\theta\right]\\&=(ab\theta_2-\frac{1}{2}a^2)(l_2-l_1)-\frac{1}{2}(b\sin\theta_2-a\cos\theta_2)(l_2^2-l_1^2)+\frac{1}{2}a^2b\ln\frac{l_2+\sqrt{l_2^2-a^2}}{l_1+\sqrt{l_1^2-a^2}}+\frac{1}{2}b(l_2\sqrt{l_2^2-a^2}-l_1\sqrt{l_1^2-a^2}),\\&-\frac{1}{6}\cos^2\theta_2(l_2^3-l_1^3)-ab(l_2\arccos\frac{a}{l_2}-l_1\arccos\frac{a}{l_1})\\&=C(a,b,l_1,l_2,\theta_2)\end{aligned} \tag{24}$$

3. $$\begin{aligned}&\int_{l_1}^{l_2}\int_{\theta_1}^{\arcsin\frac{b}{r}}A(r,\theta)drd\theta\\&=\int_{l_1}^{l_2}dr\int_{\theta_1}^{\arcsin\frac{b}{r}}d\left[ab\theta-r(b\sin\theta-a\cos\theta)+\frac{1}{2}r^2\sin^2\theta\right]\\&=ab(l_2\arcsin\frac{b}{l_2}-l_1\arcsin\frac{b}{l_1})+\frac{1}{2}ab^2\ln\frac{l_2+\sqrt{l_2^2-b^2}}{l_1+\sqrt{l_1^2-b^2}}+\frac{1}{2}a(l_2\sqrt{l_2^2-b^2}-l_1\sqrt{l_1^2-b^2})\cdot\\&-(ab\theta_1+\frac{1}{2}b^2)(l_2-l_1)-\frac{1}{6}\sin^2\theta_1(l_2^3-l_1^3)-\frac{1}{2}(a\cos\theta_1-b\sin\theta_1)(l_2^2-l_1^2)\\&=S(a,b,l_1,l_2,\theta_1)\end{aligned} \tag{25}$$

When r is in the interval $\left[a,\sqrt{a^2+b^2}\right]$, $[f(r_1,r_2),g(r_1,r_2)]$ equals to

$$[\theta_1,\theta_2]\cap\left(\arccos\frac{a}{r},\pi-\arccos\frac{a}{r}\right). \tag{26}$$

Now, calculate the following integral expressions: $\int_a^l\int_{\theta_1}^{\theta_2} A(r,\theta)drd\theta$. So $R(l,\theta_1,\theta_2)=\int_a^l\int_{\theta_1}^{\theta_2} A(r,\theta)drd\theta$.

TABLE 2 The value of $R(l,\theta_1,\theta_2)$

Angle		Value of l_1, l_2	Range of l	Value of $R(l,\theta_1,\theta_2)$
$0<\theta_1<\theta_2<\beta$		$l_1=a/\cos\theta_1$, $l_2=a/\cos\theta_2$	$a<l\le l_1$	$M(a,b,a,l,\theta_1,\theta_2)$
			$l_1<l<l_2$	$M(a,b,a,l_1,\theta_1,\theta_2)+C(a,b,l_1,l,\theta_2)$
			$l\ge l_2$	$M(a,b,a,l_1,\theta_1,\theta_2)+C(a,b,l_1,l_2,\theta_2)$
$0<\theta_1<\alpha<\beta\le\theta_2\le\frac{\pi}{2}$		$l_1=a/\cos\theta_1$, $l_2=b/\sin\theta_2$	$a<l\le l_1$	$M(a,b,a,l,\theta_1,\theta_2)$
			$l_1<l<l_2$	$M(a,b,a,l_1,\theta_1,\theta_2)+C(a,b,l_1,l_2,\theta_2)$
			$l\ge l_2$	$M(a,b,a,l_1,\theta_1,\theta_2)+C(a,b,l_1,l_2,\theta_2)+L(a,b,l_2,l)$
$\alpha\le\theta_1\le\beta<\theta_2\le\frac{\pi}{2}$	$a\sin\theta_2-b\cos\theta_1>0$	$l_1=b/\sin\theta_2$, $l_2=a/\cos\theta_1$	$a<l\le l_1$	$M(a,b,a,l,\theta_1,\theta_2)$
			$l_1<l<l_2$	$M(a,b,a,l_1,\theta_1,\theta_2)+S(a,b,l_1,l,\theta_1)$
			$l\ge l_2$	$M(a,b,a,l_1,\theta_1,\theta_2)+S(a,b,l_1,l_2,\theta_1)+L(a,b,l_2,l)$
	$a\sin\theta_2=b\cos\theta_1$	$l_1=a/\cos\theta_1$, $l_2=b/\sin\theta_2$	$a<l\le l_1$	$M(a,b,a,l,\theta_1,\theta_2)$
			$l\ge l_1=l_2$	$M(a,b,a,l_1,\theta_1,\theta_2)+L(a,b,l_2,l)$
	$a\sin\theta_2-b\cos\theta_1<0$	$l_1=a/\cos\theta_1$, $l_2=b/\sin\theta_2$	$a<l\le l_1$	$M(a,b,a,l,\theta_1,\theta_2)$
			$l_1<l<l_2$	$M(a,b,a,l_1,\theta_1,\theta_2)+C(a,b,l_1,l,\theta_2)$
			$l\ge l_2$	$M(a,b,a,l_1,\theta_1,\theta_2)+C(a,b,l_1,l_2,\theta_2)+L(a,b,l_2,l)$
$\beta<\theta_1<\theta_2\le\frac{\pi}{2}$		$l_1=b/\sin\theta_2$, $l_2=b/\sin\theta_1$	$a<l\le l_1$	$M(a,b,a,l,\theta_1,\theta_2)$
			$l_1<l<l_2$	$M(a,b,a,l_1,\theta_1,\theta_2)+S(a,b,l_1,l,\theta_1)$
			$l\ge l_2$	$M(a,b,a,l_1,\theta_1,\theta_2)+S(a,b,l_1,l_2,\theta_1)$
Angle				**Value of** $R(l,\theta_1,\theta_2)$
$0<\theta_1<\frac{\pi}{2}<\theta_2\le\pi$				$R(l,\theta_1,\frac{\pi}{2})+R(l,\pi-\theta_2,\frac{\pi}{2})$
$\frac{\pi}{2}\le\theta_1<\theta_2\le\pi$				$R(l,\pi-\theta_2,\pi-\theta_1)$

3) $0\le r_1\le a\le r_2\le\sqrt{a^2+b^2}$

The numerator value is:

$$\int_{r_1}^{r_2}\int_{\theta_1}^{\theta_2} A(r,\theta)drd\theta=\int_{r_1}^{a}\int_{\theta_1}^{\theta_2} A(r,\theta)drd\theta+\int_{a}^{r_2}\int_{\theta_1}^{\theta_2} A(r,\theta)drd\theta=\int_{r_1}^{a}\int_{\theta_1}^{\theta_2} A(r,\theta)drd\theta+R(r_2,\theta_1,\theta_2). \tag{27}$$

The value of $R(l,\theta_1,\theta_2)$ is shown as Table 2, and $\alpha=\arccos\frac{a}{b},\beta=\arccos\frac{a}{\sqrt{a^2+b^2}}$.

When $a\le r_1<r_2\le\sqrt{a^2+b^2}$, the value of numerator is $R(r_2,\theta_1,\theta_2)-R(r_1,\theta_1,\theta_2)$.

3.2 CLUSTERING RESULTS AND COMPARATIVE ANALYSIS

Figure 2 is a synthetic false colour remote sensing images based on the TM data of Band 5, 4 and 3. After the clustering with our proposed algorithm, there are seven categories, including residential areas, shadows, close planting is, dilute vegetation, water, Cho and dry land (Shown as Figure 3).

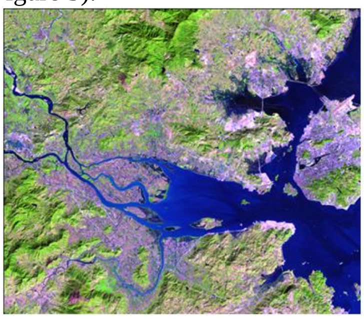

FIGURE 2 Original Image

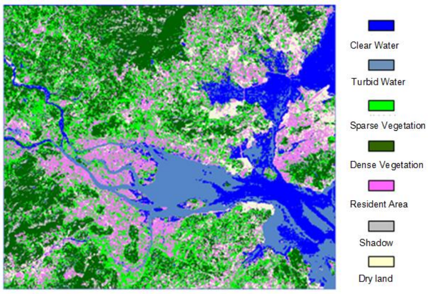

FIGURE 3 Classification Results

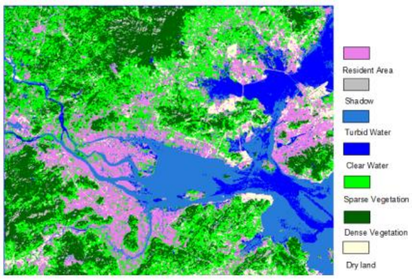

FIGURE 4 clustering results of literature

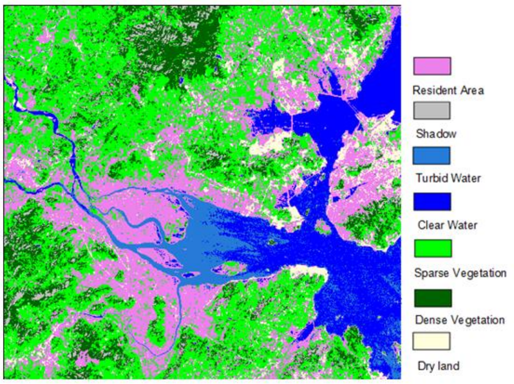

FIGURE 5 results of supervised classification

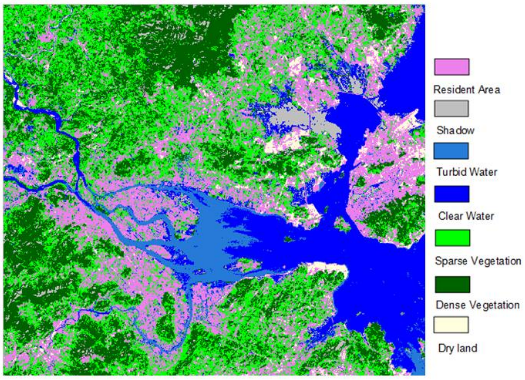

FIGURE 6 Results of Unsupervised Classification

According to the clustering results (Shown as Figure 4) of literature [2] and ERDAS, we conduct supervised classification and unsupervised classification to the same image separately. We pick up 250 pixels in the classification results to accuracy assessment, and compare with the clustering results of literature [2]. Upon examination, specific classification results of four clustering methods are shown as table 1, table 2, table 3 and table 4. The clustering methods include ERDAS supervised classification, ERDAS unsupervised classification, clustering method of literature [2] and clustering method of this paper. Bold figures in the table indicate the number of correct classification pixels for each class. Then calculate classification accuracy and the specific quantitative indicators using the figures in the four tables, results are shown in paper [1].

The paper [1] shows that the classification accuracy of the cluster validity method based on geometric probability is 85.20%, while the classification accuracy of the method, supervised classification and unsupervised classification, which are used in literature [2] are 81.60%, 62.80% and 58.00%, so this cluster validity classification method based on geometry probability is superior to the methods in literature [2].

4 Conclusions

According to the basic idea and clustering steps of geometric probability-based classification method of cluster validity, we select 1498 × 1281 pixels of Xiamen Jiulong River TM images in 2002 Winter, evaluate and compare the clustering results with the supervised classification and unsupervised classification results by ERDAS of the same image. Experimental results show that the classification method of cluster validity based on geometric probability is superior to the literature [2], and it is also superior to the methods of supervised classification and unsupervised classification.

Acknowledgment

This project is sponsored by the National Science Foundation of China (Grant No.31100415) and Education Foundation of Fujian Province (Grant No.Jb12210).

References

[1] Li X W, Mao Z Y, Li J W 2008 A Cluster Validity Function Based on Geometric Probability *Journal of Image and Graphics* **13**(12) 2351-6
[2] Huang L W, Mao Z Y, Li W Z, Wang X Q, Wu Sheng 2007 The Cluster Analysis Approaches Based on Geometric Probability and Its Application in the Classification of Remotely Sensed Images *Journal of Image and Graphics* **4**
[3] Gao X B 2004 *Fuzzy Cluster Analysis and Its Application* Xi'an Electronic Science and Technology University Press
[4] He Q 1998 Research of Fuzzy Clustering Theory and Applications *Fuzzy Systems and Mathematics* **12**(2) 89-94
[5] Gao X B, Xie W X 1999 Development and Application of Fuzzy Clustering Theory *Chinese Science Bulletin* **44**(21)
[6] Jain A K, Flynn P J 1996 *Image segmentation using clustering* In: Ahuja N, Bowyer K, eds. Advances in Image Understanding: A Festchrift for Azriel Rosenfeld. Piscataway: IEEE Press 65 − 83
[7] Cades I, Smyth P, Mannila H 2001 Probabilistic modeling of transactional data with applications to profiling, visualization and prediction, sigmod *In: Proc. of the 7th AC M SIGKDD. San Francisco: ACM Press, 2001. 37− 46* http://www.sigkdd.org/kdd2001/
[8] Jain A K, Murty M N, Flynn P J 1999 Data clustering: A review *ACM Computing Surveys* **31**(3) 264 − 323
[9] Gelbard R, Goldman O, Spiegler I 2007 Investigating diversity of clustering methods: An empirical comparison *Data & Knowledge Engineering* **63**(1) 155−66

Authors

Li Jianwei, born on October 27, 1979, Longyan, China

Current position, grades: lecturer, Master of Geography information system and Faculty of Fuzhou university
University studies: physics and information engineering in Fuzhou University
Scientific interest: Image Processing and Virtual reality
Publications: 20 Papers

Li Xiaowen, born on November 5, 1977, Longyan, China

Current position, grades: lecturer, Master of Geography information system and Faculty of Longyan university
University studies: physics and information engineering in Fuzhou University
Scientific interest: Image Processing and Virtual reality
Publications: 20 Papers