Fuzziness Measurement of Fuzzy Sets and Its Application in Cluster Validity Analysis

Xinbo Gao, Jie Li, Dacheng Tao, Xuelong Li

Abstract

To measure the fuzziness of fuzzy sets, this paper introduces a distance-based and a fuzzy entropy-based measurements. Then these measurements are generalized to measure the fuzziness of fuzzy partition, namely *partition fuzziness*. According to the relationship between the validity of fuzzy partition and its partition fuzziness, a family of cluster validity functions is proposed based on the modified *partition fuzziness*. The new cluster validity functions overcome the increasing tendency of the traditional partition fuzziness with the cluster number, which provides an effective analysis methodology for fuzzy cluster validity. The experimental results with different testing data sets illustrate the effectiveness, reliability, sensitivity and applicability of the proposed cluster validity function.

Keywords: Fuzziness measurement, partition fuzziness, cluster validity, partition entropy.

1. Introduction

In 1965, Zadeh first proposed the concept of fuzzy sets in [1]. In fuzzy sets theory, the relationship between any element *x* and a given fuzzy set $\tilde{A}$ is not belonging or not ($x \in \tilde{A}$ or $x \notin \tilde{A}$), but always belonging with some degree, which can be denoted by a fuzzy membership function $\mu_{\tilde{A}}(x)$. Obviously, in fuzzy sets the value of $\mu_{\tilde{A}}(x)$ is extended from binary $\{0,1\}$ to the unit interval $[0,1]$.

The rise of fuzzy sets settled the third crisis in the mathematical history [2], *i.e.*, the logic paradox, and broke the famous *law of contradiction* (for any a set A, if A^c is its complementary set, we have $A \cap A^c = \phi$) and the *law of the excluded middle* ($A \cup A^c = \Omega$, where Ω is the universal set.) in classical set theory. The fuzzy sets theory provides a new analysis methodology for the uncertainty information processing. Since a fuzzy set is defined by a specified membership function, it is usually of much more subjectivity for the membership function given by human being. For some fuzzy sets defined over a given universe, the important consideration is the order or shape rather than exact values for their membership function. So, with their own criteria, different users may define different membership functions for the same fuzzy set. Thus, the different membership structures will lead to different fuzziness. To measure the fuzziness of any fuzzy set, a reasonable measurement is demanded [3].

In classical set theory, the partition of a set is based on some crisp classification methods. Therefore, the homogenous elements are classified into the same category or subset with an identical membership degree, which loses the close or distant relationship among the elements in the same category. With the development of fuzzy sets theory, the fuzzy partition becomes more and more popular for its ability to describe the different membership degrees of elements to the same category. For a good partition, it should be compact within category and distinct between categories, whose fuzziness is certainly much small. From this viewpoint, it can be concluded that there is a very close relationship between the fuzziness measurement and the validity of fuzzy partition. As well known, the partition validity is an open problem in cluster analysis [4], which is usually used to determine the optimal cluster number. For this purpose, this paper aims to propose a new cluster validity function based on the fuzziness measurement and find an effective method to determine the optimal number of categories in unsupervised learning.

2. The Fuzziness Measurements of Fuzzy Sets

To construct a reasonable and effective measurement for fuzzy sets, some properties about the fuzziness measurement should be first specified according to the common senses and intuition of human beings.

For a given fuzzy set $\tilde{A}$ and any an element x_i , if we have $\mu_{\tilde{A}}(x_i) \in \{0,1\}$, then the membership of x_i to $\tilde{A}$ is distinct or crisp. That is to say, the fuzziness of

Corresponding Author: Xinbo Gao is with the School of Electronic Engineering, Xidian University, Xi'an 710071, China.
E-mail: xbgao@mail.xidian.edu.cn
Manuscript received 25 Aug. 2007; revised 17 Nov. 2007

x_i belonging to $\tilde{A}$ is 0. So, the first property of fuzziness measurement can be given as

(P1) $f_i(\mu_{\tilde{A}}) = f(\mu_{\tilde{A}}(x_i)) = 0$, iff $\mu_{\tilde{A}}(x_i) \in \{0,1\}$

Obviously, the closer the value of $\mu_{\tilde{A}}(x_i)$ approaches to 1 or 0, the less its fuzziness is. On the contrary, the further the value of $\mu_{\tilde{A}}(x_i)$ is apart from 1 or 0, the larger its fuzziness is. Hence, a reasonable supposition arises that the biggest fuzziness of membership must be occurred at the point of $\mu_{\tilde{A}}(x_i)=0.5$, because this point departs from 1 and 0 equally. Hereby, the second and third properties of fuzziness measurement are arrived.

(P2) $f_i(\mu_{\tilde{A}})$ gets its maximum, iff $\mu_{\tilde{A}}(x_i)=0.5$

(P3) In the case of $\mu_{\tilde{A}}(x_i) \geq 0.5$, if $\mu_{\tilde{A}}(x_i) < \mu^*$, then $f_i(\mu_{\tilde{A}}) \geq f(\mu^*)$; In the case of $\mu_{\tilde{A}}(x_i) < 0.5$, if $\mu_{\tilde{A}}(x_i) > \mu^*$, then $f_i(\mu_{\tilde{A}}) \geq f(\mu^*)$

As well known, for the complementary set of $\tilde{A}$, $\tilde{A}^c$ can be defined by its membership function as

$$\mu_{\tilde{A}^c}(x_i) = 1 - \mu_{\tilde{A}}(x_i)$$

According to the principle of duality, it can be concluded that $\tilde{A}$ and $\tilde{A}^c$ have equal fuzziness, *i.e.*,

(P4) $f_i(\mu_{\tilde{A}}) = f_i(\mu_{\tilde{A}^c})$, or $F(\tilde{A}) = F(\tilde{A}^c)$

For the finite set $\tilde{A}$, a function can be designed to meet the above four properties.

$$F(\tilde{A}) = \xi\left(\sum_{i=1}^{n} \alpha_i \cdot f_i(\mu_{\tilde{A}}(x_i))\right), \tag{1}$$

where, f_i is a real function with $f_i(0) = f_i(1) = 0$, $f_i(0.5)$ being the only maximum of f_i, and f_i being monotonic increase within $[0, 0.5]$ and monotonic decrease within $[0.5,1]$; $\alpha_i > 0$; $\xi(\cdot)$ is a nonnegative monotonic increasing function. For $\forall \mu \in [0,1]$, we have $f_i(\mu) = f_i(1-\mu)$.

A. *Distance-based Fuzziness measurement*

For (1), $\forall i$, let

$$f_i(\mu) = \begin{cases} \mu^p, & \mu \in [0,0.5] \\ (1-\mu)^p, & \mu \in [0.5,1] \end{cases},$$

and $\alpha_i = \frac{1}{n}$, $\xi(y) = y^{\frac{1}{p}}$, which is plotted in Figure 1. We have

$$F_d(\tilde{A}) = \left\{\frac{1}{n}\sum_{i=1}^{n}\left|\mu_{\tilde{A}}(x_i) - \mu_{\tilde{A}_{0.5}}(x_i)\right|^p\right\}^{1/p}. \tag{2}$$

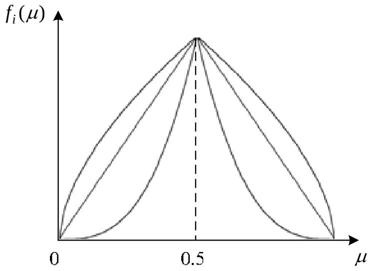


Figure 1. $f_i(\mu)$ *vs* μ with p = 0.5, 1, 2

It implies that $F_d(\tilde{A})$ stands for the Minkowsky distance between $\tilde{A}$ and its closest crisp set $\tilde{A}_{0.5}$ which is the 0.5-level set of $\tilde{A}$ and defined as

$$\mu_{\tilde{A}_{0.5}}(x_i) = \begin{cases} 0, & \mu_{\tilde{A}}(x_i) < 0.5 \\ 1, & \mu_{\tilde{A}}(x_i) \geq 0.5 \end{cases}.$$

When $p = 1$, (2) is degraded to the Hamming distance between $\tilde{A}$ and $\tilde{A}_{0.5}$. On the other hand, when $p = 2$, it corresponds to the Euclidean distance, and when $p \to \infty$, it becomes the Chebyshev distance.

B. *Fuzzy entropy of fuzzy set*

For (1), $\forall i$, if taking $\alpha_i = \frac{1}{N}$, $\xi(y) = y$, and

$$f_i(\mu) = -\mu \ln \mu - (1-\mu)\ln(1-\mu),$$

we have

$$F_e(\tilde{A}) = -\frac{1}{n}\sum_{i=1}^{N}\left\{\begin{aligned}&\mu_{\tilde{A}}(x_i)\ln \mu_{\tilde{A}}(x_i) + \\ &(1-\mu_{\tilde{A}}(x_i))\ln(1-\mu_{\tilde{A}}(x_i))\end{aligned}\right\} \tag{3}$$

Obviously, if $\mu_{\tilde{A}}(x_i) = 0.5$, for $\forall x_i$, the fuzzy entropy gets its maximum. The larger the fuzzy entropy, the more uncertainty the fuzzy set has.

3. Fuzzy Partition and Its Fuzziness Measurements

A. *Fuzzy partition*

Assume that a classical set $X = \{x_1, x_2, \cdots, x_n\}$ is partitioned into c subsets, $X_1, X_2, \cdots, X_c$, namely *c*-partition, we have

$$\bigcup_{i=1}^{c} X_i = X;\ \ X_i \cap X_j = \phi,\ i \neq j,\ \ i, j = 1,2,\cdots,c.$$

For each $x_k \in X$, its membership function to each partition $X_i\,(1 \leq i \leq c)$ can be denoted as

$$\mu_{X_i}(x_k) = \mu_{ik}^h = \begin{cases} 1, & x_k \in X_i \\ 0, & x_k \notin X_i \end{cases}.$$

Such partition is usually called as *crisp* or *hard* *c*-partition, whose partition space can be represented as

$$P_h = \left\{ \mu_{ik}^h \middle| \mu_{ik}^h \in \{0,1\}; \sum_{i=1}^{c} \mu_{ik}^h = 1, \forall k; 0 < \sum_{k=1}^{n} \mu_{ik}^h < n, \forall i \right\}.$$

With the introduction of fuzzy sets, one can easily generalize the crisp partition to its fuzzy version. According to the definition of fuzzy sets, if the value range of μ_{ik} is generalized from $\{0,1\}$ to $[0,1]$, denoted as μ_{ik}^f, we can obtain the fuzzy partition space of *X*.

$$P_f = \left\{ \mu_{ik}^f \middle| \mu_{ik}^f \in [0,1]; \sum_{i=1}^{c} \mu_{ik}^f = 1, \forall k; 0 < \sum_{k=1}^{n} \mu_{ik}^f < n, \forall i \right\}$$

For the obtained *c* fuzzy subsets $\tilde{X}_1, \tilde{X}_2, \cdots, \tilde{X}_c$ of set *X*, there is

$$\bigcup_{i=1}^{c} \mathrm{supp}(\tilde{X}_i) = X \ ,$$

where $\mathrm{supp}(\cdot)$ stands for the support set of fuzzy set.

It is obvious that the μ_{ik}^h distributes at the vertices of *c*-hypecub associating to the three base vectors, while the μ_{ik}^f distributes on the hype-plane determined by $\sum_{i=1}^{c} \mu_{ik}^f = 1$.

B. *Fuzziness measurements of fuzzy partition*

The fuzzy partition of a given set can also be represented by its fuzzy partition matrix $U = \left[\mu_{ik}^f\right]_{c\times n}$, and its fuzziness measurement can be defined like the case of fuzzy set.

(1) *Distance based fuzziness measurement*

For a given fuzzy partition *U*, $\forall x_k \in X$, its membership degrees to the *c* partitions are denoted as $\mu_{ik}^f = \mu_{\tilde{X}_i}(x_k)$, $i = 1,2,\cdots,c$, whose closest crisp sets can be obtained by

$$\mu_{ik}^{(h)} = \begin{cases} 1, & \mu_{ik}^f = \max_{j=1}^{c}\{\mu_{jk}^f\} \\ 0 & \text{otherwise} \end{cases}, \ i = 1,2,\cdots,c$$

Hereby, the fuzziness measurement of fuzzy partition can be defined as the distance between the fuzzy partition and its closest crisp sets.

$$F_d(U,c) = \left\{ \frac{1}{n} \sum_{k=1}^{n} \sum_{i=1}^{c} \left| \mu_{ik}^f - \mu_{ik}^{(h)} \right|^p \right\}^{1/p} \quad (4)$$

As mentioned above, in the cases of $p = 1,2,\infty$, (4) corresponds to the Hamming, Euclidean and Chebyshev distances between the fuzzy partition $\tilde{X}_i$ and its closest crisp partition $(\tilde{X}_i)_h$.

For the partition fuzziness $F_d(U,c)$ defined in (4), we have

- $0 \le F_d(U,c) \le 2 - 2/c$
- $F_d(U,c) = 0$, iff the fuzzy partition is degraded into crisp one, *i.e.*, $\mu_{ik}^f \in \{0,1\}, \forall i,k$
- $F_d(U,c) = 2 - 2/c$, iff $\mu_{ik}^f = 1/c, \forall i,k$

The more distinct the fuzzy partition is, the less the value of $F_d(U,c)$. On the contrary, the fuzzier the partition is, the larger the value of $F_d(U,c)$. If and only if $\mu_{ik}^f = 1/c$, *i.e.*, the most fuzzy partition, the $F_d(U,c)$ takes its maximum 2-2/*c*.

(2) *Partition entropy of fuzzy partition*

Following the definition of fuzzy entropy of fuzzy set, we can define the partition entropy of fuzzy partition as

$$F_e(U,c) = -\frac{1}{n} \sum_{k=1}^{n} \sum_{i=1}^{c} \mu_{ik}^f \ln \mu_{ik}^f \quad (5)$$

In the case of $\mu_{ik}^f = 0$, let $\mu_{ik}^f \ln \mu_{ik}^f = 0$. For the fuzzy partition entropy defined in (5), we have

- $0 \le F_e(U,c) \le \log_a c$
- $F_e(U,c) = 0$, iff $\mu_{ik}^f \in \{0,1\}, \forall i,k$
- $F_e(U,c) = \log_a c$, iff $\mu_{ik}^f = \dfrac{1}{c}, \forall i,k$

It is obvious that the fuzzy partition entropy takes a small value when the partition is more distinct, and it takes the maximum $\ln c$, in the case of $\mu_{ik}^f = 1/c$.

4. Cluster Validity Function based on Partition Fuzziness

There are many methods available for the fuzzy partition of data set, in which the most popular algorithms are those based on fuzzy clustering, such as fuzzy *c*-means (FCM) and its variations. To evaluate the fuzzy partition results of different methods, some cluster validity functions are proposed as assessment criterion. For the clustering problem, the cluster validity is usually referred to the determination of the optimal cluster number *c**.

In what follows, several typical cluster validity functions are first introduced briefly. Then a family of new cluster validity functions is proposed based on the modified *partition fuzziness* (PF).

A. *Three typical cluster validity functions*

The existing cluster validity functions can be roughly classified into two categories. One is the fuzzy partition based methods, such as separation degree [5], partition entropy [6], proportional coefficient [7] and possibilistic distribution [8], *etc*., all of which are simple and easy to implement for their low computational complexity. Another is methods based on the geometric structure of data set, such as partition coefficient [9], separation coefficient [10], Xie-Beni index [11] and graph theory based validity function [12], all of which are closely related to the data set structure, but complex to express with large computational load.

Let the fuzzy c-partition of data set X be denoted as $U=[\mu_{ik}]_{c\times n}$. Three popular cluster validity functions are presented as follows.

(1) *Partition fuzziness*

$$F_p(U^*,c^*)=\min_c\{\min_{U\in\Omega_c}\{F_p(U,c)\}\},$$

where Ω_c indicates the finite set of optimal partition matrices corresponding to the given cluster number c. When $F_p=F_d$, it corresponds to the distance-based partition fuzziness. While, when $F_p=F_e$, it is identical to the fuzzy entropy based partition fuzziness.

The partition fuzziness is sensitive to the fuzzy partition result of data set, which makes it suitable to measure the cluster validity. Unfortunately, with the increase of the cluster number c, the partition fuzziness shows an upward tendency for uniformly distributed data set as shown in Figure 2, which will influent the choice of the optimal cluster number c*.

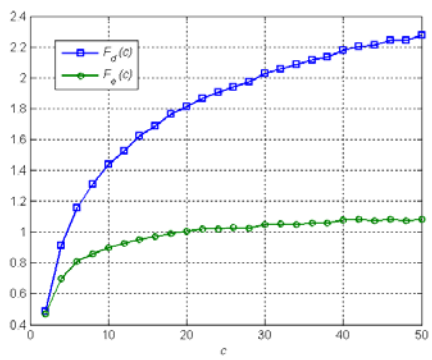


Figure 2. The upward tendency of partition fuzziness *vs*. cluster number c.

(2) *Xie-Beni index*

$$XB(c^*)=\min_c\left\{\frac{\sum_{i=1}^{c}\sum_{k=1}^{n}\mu_{ik}^m\cdot\|x_k-v_i\|^2}{n\cdot\min_{i,k}\|v_i-v_k\|^2}\right\},$$

where $v_i, i=1,2,\cdots,c$ stands for the cluster center of the i-th cluster, and m is fuzzy weighting exponent, $1<m<\infty$, which controls the fuzziness of the fuzzy partition. Xie-Beni index has been achieved wide applications for its high performance, because it fuses the information of fuzzy partition and geometric structure.

(3) *Fuzzy hyper volume*

$$FHV=\min_c\left\{\sum_{i=1}^{c}V_i\right\}=\min_c\left\{\sum_{i=1}^{c}|\Sigma_i|^{\frac{1}{2}}\right\}$$

where $\Sigma_i=\dfrac{\sum_{k=1}^{n}\mu_{ik}^m(x_k-v_i)(x_k-v_i)^T}{\sum_{k=1}^{n}\mu_{ik}^m}$ is the fuzzy covariance matrix of the i-th cluster. $V_i=|\Sigma_i|^{\frac{1}{2}}$ defines the fuzzy hyper volume of the i-th cluster. The less the value of *FHV*, the compacter the cluster is.

The Xie-Beni index and the fuzzy hyper volume belong to the second kind of cluster validity function, both of which are sensitive to the noisy data or outliers. Moreover, they are difficult to give a multi-resolution representation of data set.

B. *Cluster validity function based on the modified PF*

To overcome the upward tendency of the cluster validity based on partition fuzziness with the increasing of cluster number c, this section proposes a compensation measure and defines a modified PF as

$$F_p^m(U,c)=\frac{F_p(U,c)}{F_p^s(U,c)}.$$

As mentioned above, $F_p(U,c)$ has two kinds of forms, $F_d(U,c)$ and $F_e(U,c)$ as defined in (2) and (3) separately. $F_p^s(U,c)$ is used to represent the upward tendency of PF with the increasing of cluster number c, which can be defined as the following two formulas but not limited.

(1)
$$\begin{aligned}F_p^{s_1}(U,c)&=F_p(c_0)+\frac{F_p'(c_0)}{1!}(c-c_0)+\cdots\\&+\frac{F_p^{(l)}(c_0)}{l!}(c-c_0)^l\\&=\alpha_0+\alpha_1c+\alpha_2c^2+\cdots+\alpha_lc^l=\sum_{i=0}^{l}\alpha_ic^i\end{aligned}$$

That is to say, $F_p^{s_1}(U,c)$ is defined as the first l orders Taylor series of $F_p(U,c)$. For the sake of convenience of computation, the l orders polynomial fitting result of $F_p(U,c)$ is used to approximate its varying tendency.

(2) $F_p^{s_2}(U,c)=(F_{LP}(F_e(U,c)))^\gamma$

where, $F_{LP}(\cdot)$ denotes the low-pass filter, *i.e.*, smoothing function, and γ is the weighting exponent.

Here, we take the distance-based modified PF as an example to define the new cluster validity function.

$$F_p^{m_1}(U^*,c^*) = \min_c\left\{\min_{\Omega_c}\frac{F_d(U,c)}{F_d^{s_1}(U,c)}\right\}$$
$$= \min_c\left\{\min_{\Omega_c}\frac{\left\{\frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{c}\left|\mu_{ik}^f - \mu_{ik}^{(h)}\right|^p\right\}^{\frac{1}{p}}}{\sum_{i=1}^{l}\alpha_i c^i}\right\}$$

$$F_p^{m_2}(U^*,c^*) = \min_c\left\{\min_{\Omega_c}\frac{F_d(U,c)}{(F_{LP}(F_e(U,c)))^\gamma}\right\}$$
$$= \min_c\left\{\min_{\Omega_c}\frac{\left\{\frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{c}\left|\mu_{ik}^f - \mu_{ik}^{(h)}\right|^p\right\}^{\frac{1}{p}}}{\left(F_{LP}\left(-\frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{c}\mu_{ik}^f\cdot\ln\mu_{ik}^f\right)\right)^\gamma}\right\}$$

In the following experiments, the influence of parameters p,l,γ on the performance of cluster validity function will be discussed in details.

5. Experimental Results and Analysis

To test the performance of the proposed cluster validity functions based on the modified PF, four experiments are designed to verify its effectiveness, reliability, sensitivity and applicability separately. The experiments are performed on the synthesized data sets and real data sets from UCI [13]. The compared experimental results with several typical cluster validity functions mentioned in Section 4.A are used to illustrate the advantages of the proposed cluster validity functions.

A. *Experiment for effectiveness test*

Figure 3 (a) shows a synthesized test data set, which consists of 9 subsets, and can be classified into 3 groups at a coarse resolution level. The functions of distance-based PF with cluster number *c*, $F_d(c)$, are plotted at Figure 3 (b). It can be found that the first and second minima of the cluster validity functions are occurred at *c*=3 and *c*=9 respectively. With the increasing of *p*, the second minimum becomes larger and larger gradually. Whatever the value of *p* takes, $F_d(c)$ shows the upward tendency with the increasing of *c*, which obviously is not favorable for the choice of *c**.

Figure 3 (c) and (d) present the plot of in variation of the two functions of modified PF, $F_p^{m_1}(c)$ and $F_p^{m_2}(c)$ with the cluster number *c*, respectively, in which the order of polynomial fitting takes $l=2$ in Figure 3 (c), and the weighting exponent takes $\gamma = 2$. It is obvious that the upward tendency of PF is compensated very well, and it becomes easy to find the two minima at c=3 and $c=9$ from the function curves of $F_p^{m_1}(c)$ and $F_p^{m_2}(c)$, which corresponds to the optimal cluster numbers of data set at high and low resolution levels, and illustrate the effectiveness of the proposed cluster validity functions.

In addition, it can be found from Figure 3 (c) that there is little influence of parameter *p* on the variation of $F_p^{m_1}(c)$. So, we will take *p*=1 hereafter for the sake of convenience of computation. On the other hand, it is a big influence of the parameter *p* on the variation of $F_p^{m_2}(c)$. With the increasing of *p*, it becomes more and more unreliable to determine the optimal *c**.

To study the influence of parameters l and γ in $F_p^{m_1}(c)$ and $F_p^{m_2}(c)$ on the cluster validity analysis, some experiments are conducted with *l*=0, 1, 2, 5, 20 and $\gamma = 0.5, 0.7, 1.0, 1.2, 2.0$. The experimental results are presented in Figure 3 (e) and (f). When $l=0$, $F_p^{m_1}(c)$ is degraded to $F_d(c)$. For $l \geq 1$, there is little influence of the value of *l* on the variation of $F_p^{m_1}(c)$. For the influence of γ on the variation of $F_p^{m_2}(c)$, it can be found from Figure 3(f) that $\gamma = 2$ obtains the better compensation performance.

B. *Experiment for reliability test*

In this experiment, a complex-structured test data set is design as shown in Figure 4(a), which consists of 27 small subsets at a high resolution level, 9 big subsets at a middle resolution level, and 3 bigger subsets at a low resolution level. For such a data set, it is difficult for the existing cluster validity algorithms to detect all the optimal cluster numbers at three resolution levels.

Figure 4 (b) presents the detection results of four kinds of cluster validity algorithm, *i.e.*, Xie-Beni index, fuzzy hyper volume, partition entropy, and partition coefficient. Almost all the four kinds of algorithms can detect the optimal cluster number *c** = 27. However, *c** = 9 and 3 are not detected. Figure 4 (c) and (d) show the detection results of the proposed two kinds of cluster validity algorithms. All the three optimal cluster numbers *c** = 3, 9, 27 can be detected, which demonstrates the reliability of the proposed cluster validity algorithms.

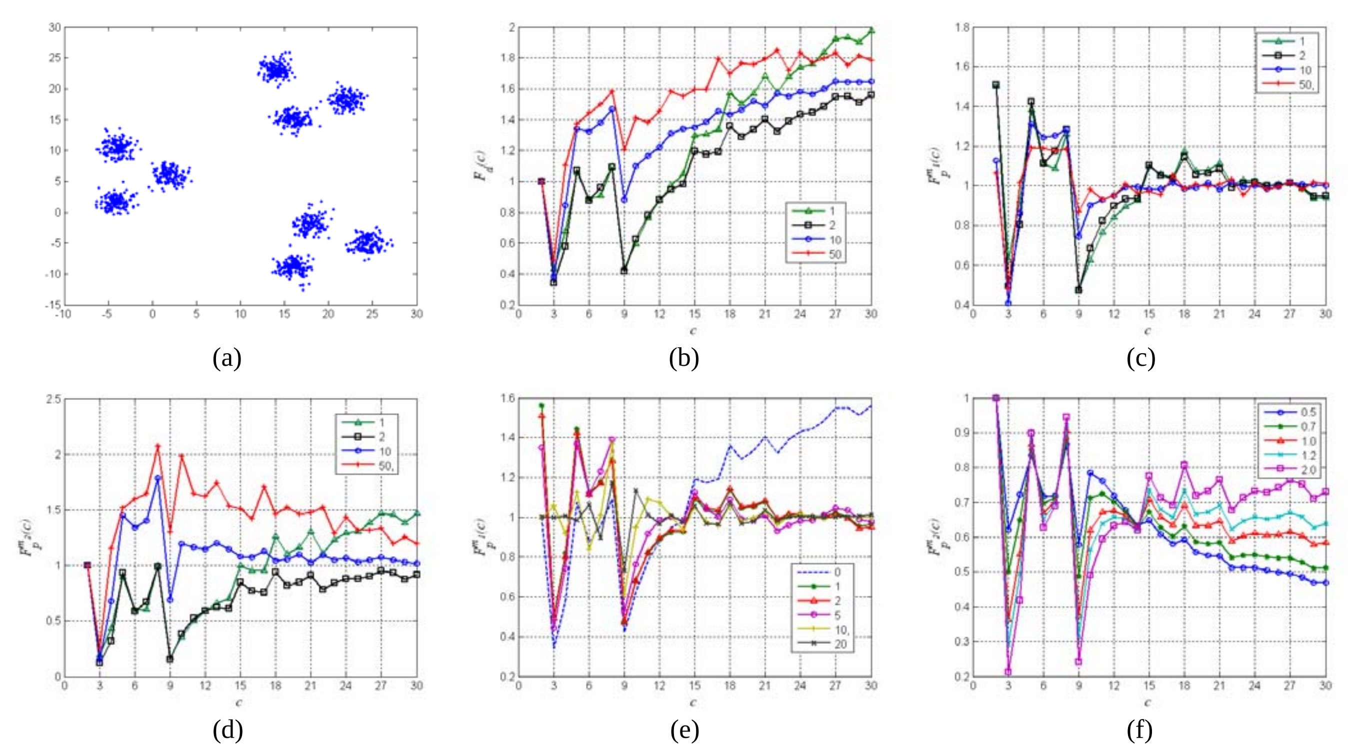


Figure 3. The test experimental results for the effectiveness of the proposed cluster validity functions: (a) Test data set; (b) Partition fuzziness ($p=1,2,10,50$); (c) $F_p^{m_1}(c)$ *vs* c $(l=2, p=1,2,10,50)$; (d) $F_p^{m_2}(c)$ *vs* c $(\gamma=2, p=1,2,10,50)$; (e) $F_p^{m_1}(c)$ *vs* c $(p=1, l=0,1,2,5,20)$; (f) $F_p^{m_2}(c)$ *vs* c $(p=1,\gamma=0.5,0.7,1.0,1.2,2.0)$.

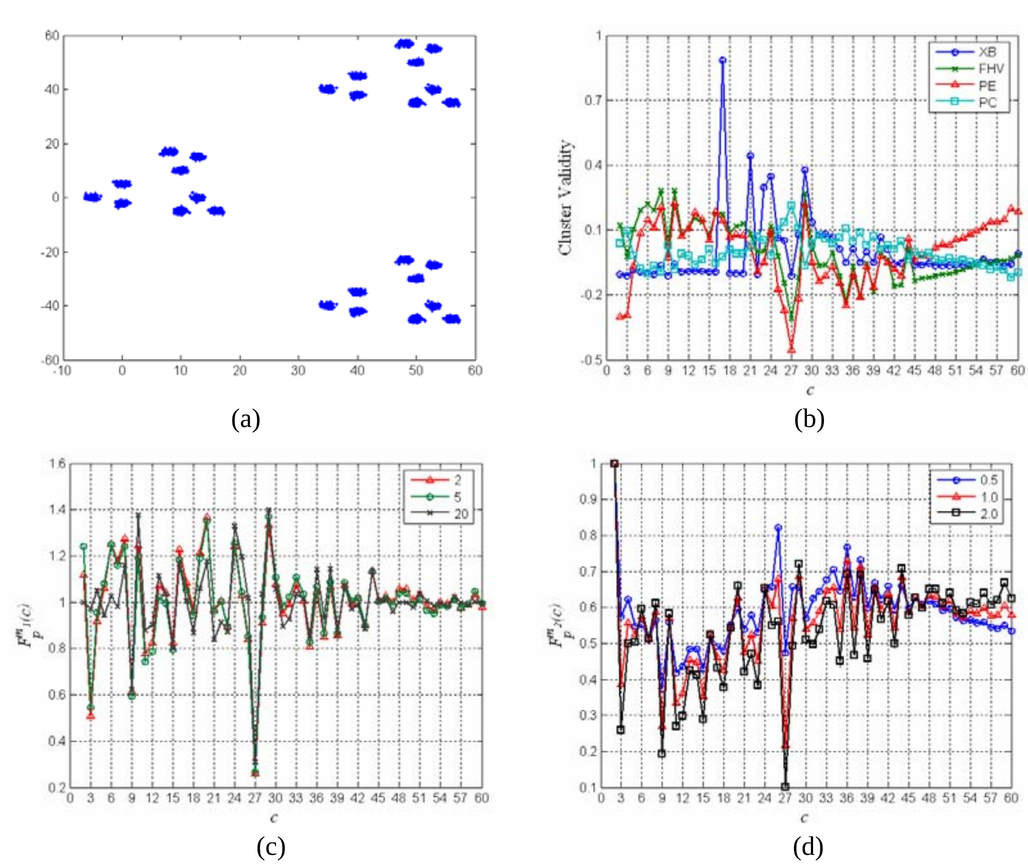


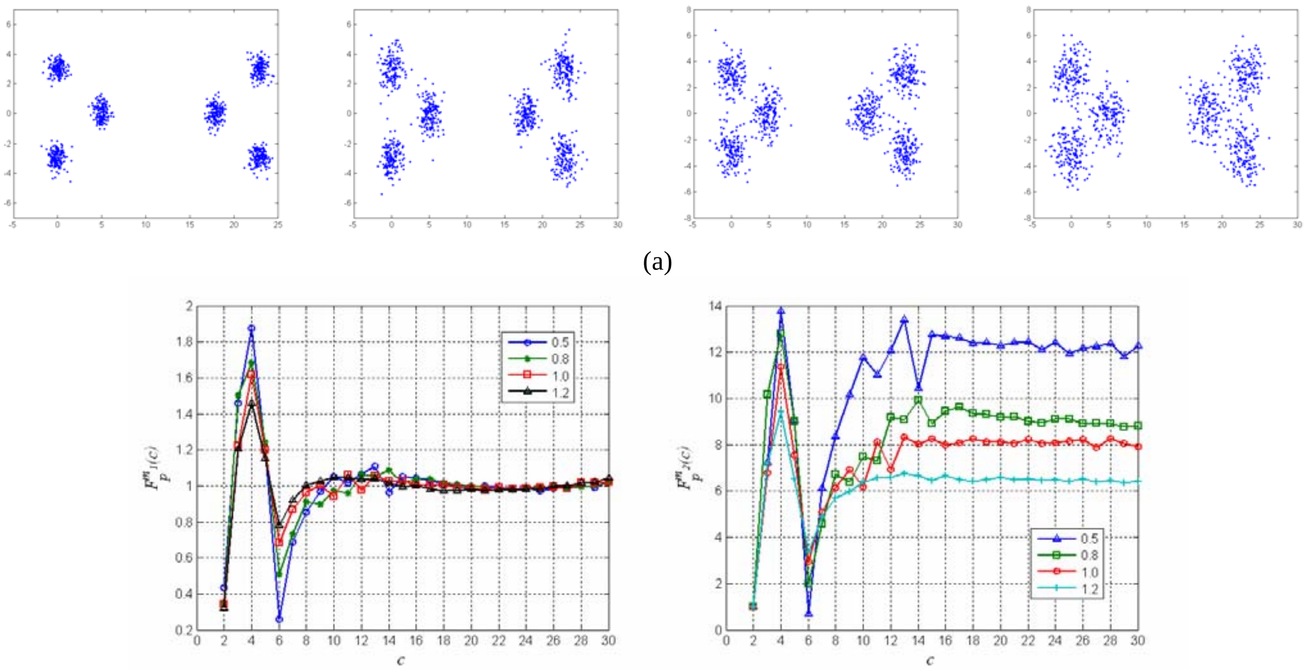
Figure 4. The test experimental results for the reliability of the proposed cluster validity functions: (a) Test data set; (b) The comparison of several cluster validity functions; (c) $F_p^{m_1}(c)$ *vs* c $(p=1, l=2,5,20)$; (d) $F_p^{m_2}(c)$ *vs* c $(p=1,\gamma=0.5,1.0,2.0)$.

C. *Experiment for sensitivity test*

To test the sensitivity of the proposed cluster validity algorithms, a synthesized data set is designed with two subsets, each of which is composed of 3 independent identical distributed Gaussian subsets with the same variance.

Firstly, the variance of each subset is changed from 0.5, 0.8, 1.0 to 1.2 as shown in Figure 5 (a). Figure 5 (b) and (c) presents the detection results of the proposed two kinds of cluster validity algorithms. With the increasing of the variance, the separability within each category becomes worse and worse. The proposed two algorithm can always detect the optimal cluster numbers c^* =6, 2. When the variance is increased from 0.5 to 1.2, the minimum of $F_p^{m_1}(c)$ and $F_p^{m_2}(c)$ is changed from $c=6$ to $c=2$, which is agree with the intuition of human beings and illustrate the sensitivity of the proposed cluster validity functions.

Secondly, we fix $\sigma^2=0.8$ and change the sample number within each small subset, n = 50, 100, 500, 1000, which is shown in Figure 6 (a). Figure 6 (b) and (c) show the functions of $F_p^{m_1}(c)$ and $F_p^{m_2}(c)$ with the cluster number c. It is obvious that $F_p^{m_1}(c)$ and $F_p^{m_2}(c)$ is independent to the sample number in each subset.



(a)

(b) (c)

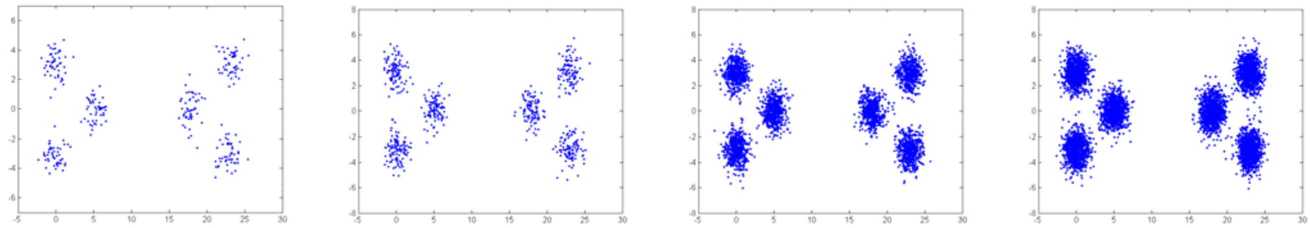
Figure 5. The first test experimental results for the sensitivity of the proposed cluster validity functions: (a) Test data sets (variance = 0.5, 0.8, 1.0, 1.2, respectively); (b) $F_p^{m_1}(c)$ *vs* c ($p=1, l=2$); (c) $F_p^{m_2}(c)$ *vs* c ($p=1, \gamma=2.0$).

Thirdly, the relationship between the modified partition fuzziness and the distribution of data set is studied with two sets of data, each of which consists of 5 subsets. Each subset of the test data set I is Gaussian distributed, and the subsets of the test data set II are uniformly distributed, which are shown in Figure 7 (a) and (b). Figure 7 (c) and (d) present the variation of $F_p^{m_1}(c)$ and $F_p^{m_2}(c)$ with the cluster number c. It can be found that both the two functions get the minimum at c = 5. For the two functions, the minimum of Gaussian distributed data set is appreciably smaller that that of uniformly distributed data set, which is agree with our intuition.

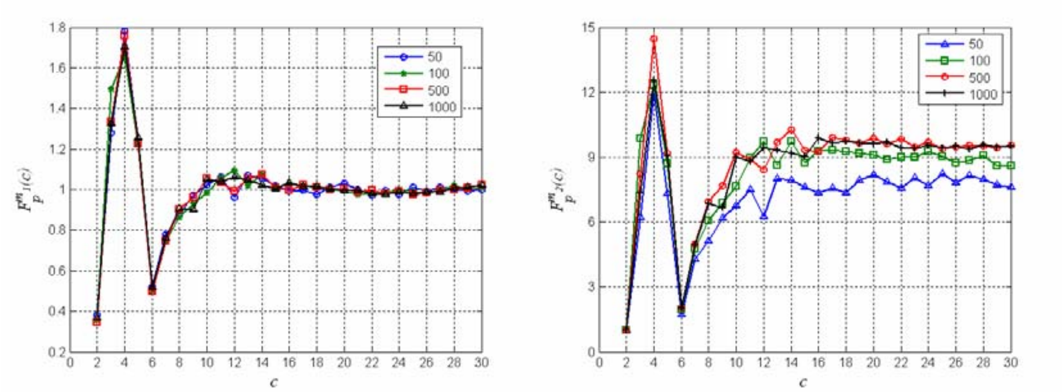
D. *Experiment for applicability test*

To consider the practical application effect of the proposed cluster validity functions, two sets of real data are selected as testbed from UCI database. One is *Pendigits* data set, which consists of 10 categories and 7494 samples with 16 dimensional features. Another one is *Animal* data set, which consists of 4 categories and 256 samples with 72 dimensional features.

Figure 8 (a) and Figure 9 (a) show the plots of variation of $F_p^{m_1}(c)$ with cluster number c on the *Pendigits* and *Animals* data sets respectively, from which we can get the optimal cluster numbers are $c^*=10$ for the first data set and $c^* = 4$ for the second data set. Figure 8 (b) and Figure 9 (b) show the plots of variation of $F_p^{m_2}(c)$ with c on the *Pen digits* and *Animals* data sets respectively. Although there are local minima at $c^* = 10$ in Figure 8 (b) and $c^* = 4$ in Figure 9 (b), it is not easier to determine the optimal cluster number for $F_p^{m_2}(c)$ than that for $F_p^{m_1}(c)$. Figure 8 (c) and Figure 9 (c) present the detection results of the four popular cluster validity functions. It is obvious that they cannot determine the optimal cluster number.

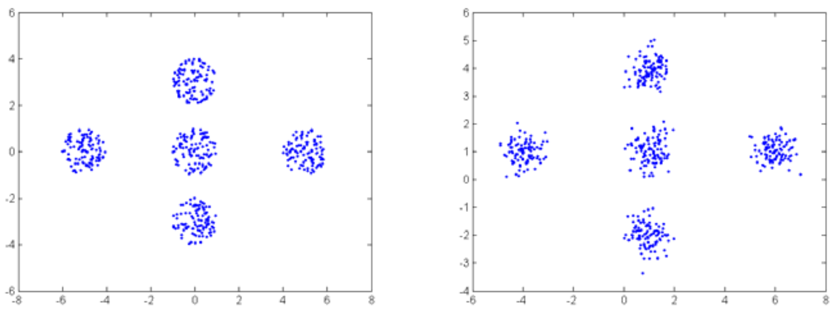


(a)

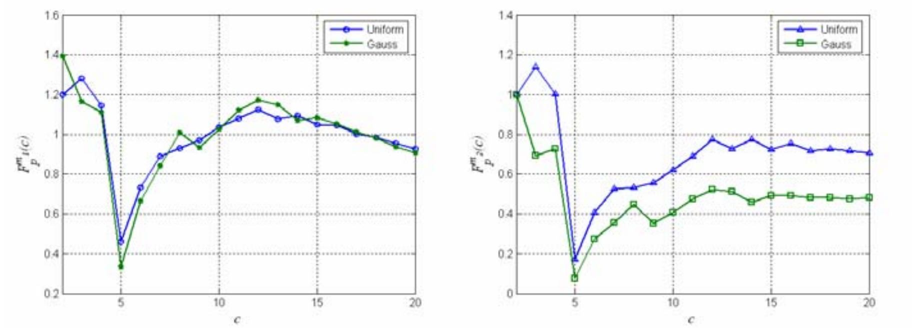


(b) (c)

Figure 6. The second test experimental results for the sensitivity of the proposed cluster validity functions: (a) Test data sets (Variance = 0.8, the number of sample = 50, 100, 500, 1000, respectively); (b) $F_p^{m_1}(c)$ *vs* c $(p = 1, l = 2)$; (c) $F_p^{m_2}(c)$ *vs* c $(p = 1, \gamma = 2.0)$.



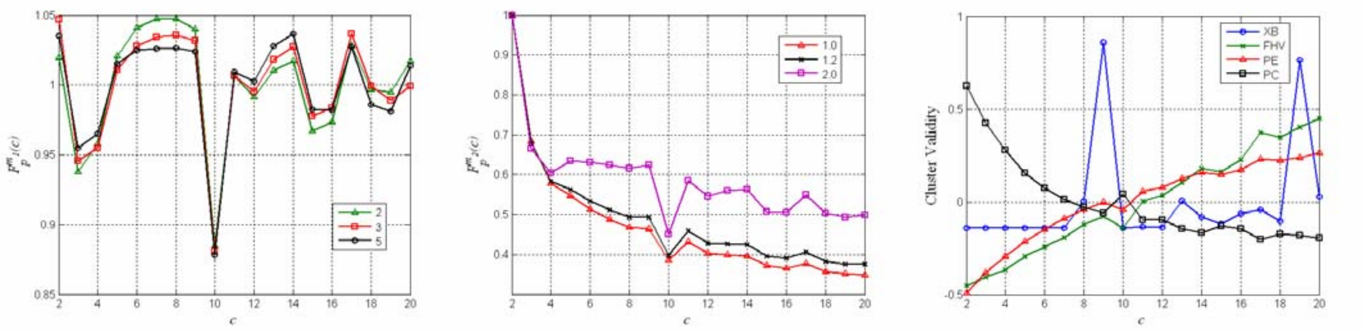
(a) (b)



(c) (d)

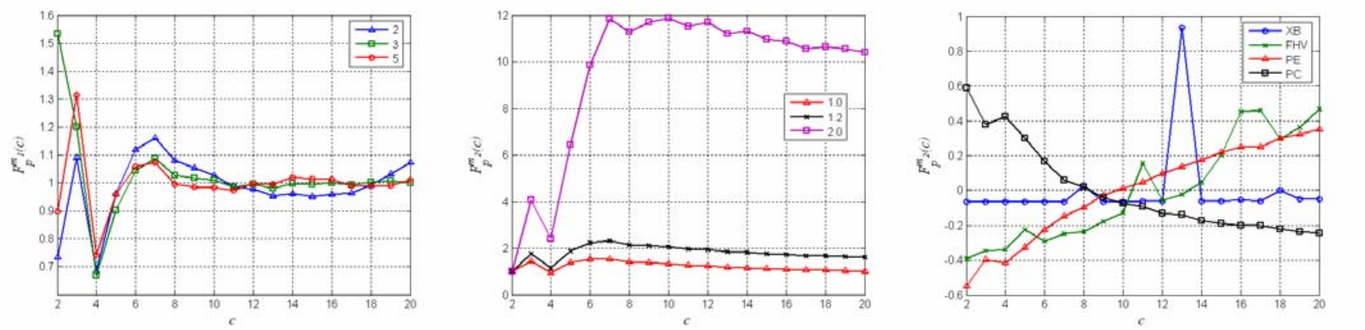
Figure 7. The third test results for the sensitivity of the proposed cluster validity functions. (a) Test data set A; (b) Test data set B;

(c) $F_p^{m_1}(c)\ vs\ c \quad (p=1, l=2)$; (d) $F_p^{m_2}(c)\ vs\ c \quad (p=1, \gamma=2.0)$.



(a) (b) (c)

Figure 8. The first test results for the sensitivity of the proposed cluster validity functions. (a) $F_p^{m_1}(c)\ vs\ c \quad (p=1, l=2,3,5)$; (b) $F_p^{m_2}(c)\ vs\ c \quad (p=1, \gamma=1.0,1.2,2.0)$; (c) Plots of 4 kinds of cluster validity functions



(a) (b) (c)

Figure 9. The second test results for the sensitivity of the proposed cluster validity functions. (a) $F_p^{m_1}(c)\ vs\ c \quad (p=1, l=2,3,5)$; (b) $F_p^{m_2}(c)\ vs\ c \quad (p=1, \gamma=1.0,1.2,2.0)$; (c) Plots of 4 kinds of cluster validity functions.

6. Conclusions

This paper firstly studies the fuzziness measurements of fuzzy sets, and proposes a distance-based fuzziness measurement $F_d(\mu)$ and a fuzzy entropy-based fuzziness measurement $F_e(\mu)$. Then, the fuzziness measurements of fuzzy sets are generalized to the fuzziness measurements of fuzzy partition. Furthermore, the fuzziness measurements of fuzzy partition are used to deal with the cluster validity problem for determining the optimal cluster number in unsupervised learning.

To compensate the upward tendency of partition fuzziness function with cluster number *c*, this paper presents two kinds of modified partition fuzziness functions $F_p^{m_1}(c)$ and $F_p^{m_2}(c)$. The first one is compensated by the polynomial fitting of $F_d(c)$, and the second one is compensated by the smoothing result of $F_e(c)$. The four sets of test experiments illustrate the effectiveness, reliability, sensitivity and applicability of the proposed cluster validity analysis method. Of course, one can design more modified partition fuzziness function according to the basic principle described in this paper.

Acknowledgment

This work was supported by the Program for New Century Excellent Talents in University (NCET-04-0948), the Program for Changjiang Scholars and Innovative Research Team in University of China (IRT0645), the National Natural Science Foundation of China (No. 60771068, No.60702061), and the Key Lab of ATR, Shenzhen University, China.

References

[1] L. A. Zadeh, "Fuzzy sets," *Information and control*, vol. 8, no. 3, pp. 338-353, 1965.

[2] Petr Hajek, Jeff Paris, John Shepherdson, "The liar paradox and fuzzy logic," *The Journal of Symbolic Logic*, vol. 65, no. 1, pp. 339-346, 2000.

[3] R. R. Yager, "On the measure of fuzziness and negation, Part I: membership in the unit interval," *Int. J. of General Systems*, vol. 5, pp. 221-229, 1979.

[4] N. R. Pal, J. C. Bezdek, "On cluster validity for

the fuzzy c-means model," *IEEE Trans. on Fuzzy System*, vol. 3, no. 3, pp. 370-379, 1995.

[5] L. A. Zadeh, "Similarity relations and fuzzy orderings," *Information Science*, vol. 3, no. 2, pp. 177-200, 1971.

[6] J. C. Bezdek, "Clustering validity with fuzzy sets," *J. Mathematical Biology*, vol. 22, no. 1, pp. 57-71 1974.

[7] G. J. Williams and Z. Huang, "A case study in knowledge acquisition for insurance risk assessment using a KDD methodology," In *Proceedings of the Pacific Rim Knowledge Acquisition Workshop*, Sydney, Australia, pp. 117-129, 1996.

[8] Jiulun Fan, Jihong Pei, Weixin Xie, "A cluster validity function based on possibilistic distribution," *Chinese Journal of Electronics*, vol. 26, no. 1, pp. 127-130, 1998.

[9] J. C. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well separated cluster," *J. Cybernet*, vol. 3, no. 1, pp. 32-57, 1974.

[10] R. Gunderson, "Applications of fuzzy ISODATA algorithms startracker printing systems," In *proceedings 7th Triennial World IFAC Congress*, Helsinki, Finland, pp. 1319-1323, 1978.

[11] X. L. Xie, G. Beni, "A validity measure for fuzzy clustering," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 13, no. 8, pp. 841-847, 1991.

[12] N. R. Pal and J. Biswas, "Clustering validation using graph theoretic concepts," *Pattern Recognition*, vol. 30, no. 6, pp. 847-857, 1997.

[13] http://mlearn.ics.uci.edu/MLRepository.html

Xibo Gao received the B.S. in electronic engineering, the M.S. and the Ph.D. degrees both in signal and information processing from Xidian University, Xi'an, China, in 1994, 1997 and 1999, respectively. From 1997 to 1998, he was with the Computer Games Research Institute of Shizuoka University, Japan, as a research fellow. From 2000 to 2001, he also worked at Multimedia Lab of the Chinese University of Hong Kong as a research Associate. Since 2003, he has been a Professor in the School of Electronic Engineering at Xidian University, Xi'an, China, and the Director of Video/Image Processing System Laboratory (VIPSL). His research interests include video/image processing, pattern recognition, and machine learning. Dr. Gao was selected as a member of the program for New Century Excellent Talents in University of China by the Ministry of Education (MOE) in 2004. He was authorized the title Pacemaker of Ten Excellent Young Teacher of Shaanxi Province in 2005. In 2006, he was awarded the Young Teacher Award of High School by the Fok Ying Tung Education Foundation.