Fuzzy C-Means Clustering Validity Function Based on Multiple Clustering Performance Evaluation Components

Guan Wang[1] · Jie-Sheng Wang[1] · Hong-Yu Wang[1]

Abstract Clustering is the process of grouping a set of physical or abstract objects into multiple similar objects. Fuzzy C-means (FCM) clustering is one of the most widely used clustering methods, whose main research goal is to find the optimal clustering number of data sets, which is related to whether the data can be effectively divided. The study of clustering validity function is the process of evaluating the clustering quality and determining the optimal clustering number. Based on the idea of components, six cluster performance evaluation components are proposed to define compactness, variation, similarity, overlap and separation of data sets, respectively. Then a new validity function based on FCM clustering algorithm is synthesized by these six components. Finally, the proposed validity function and eight typical validity functions are compared on five artificial data sets and eight UCI data sets. The simulation results show that the proposed clustering validity function can evaluate the clustering results more effectively and determine the optimal clustering number of different data sets.

Keywords Fuzzy C-means clustering algorithm · Clustering validity function · Clustering performance evaluation components

1 Introduction

As an important research content in the field of data mining, clustering is an unsupervised pattern recognition method, which is to cluster data sets under the guidance of no prior knowledge, so that the data in the same category is as similar as possible, and the greater the difference between different categories, the better [1]. Clustering is mainly divided into two directions: hard clustering and fuzzy clustering. Hard clustering, for example, k-means clustering algorithm [2] performs clustering according to the idea of "non-0 equals 1", which requires that each sample must be clearly divided into different sub-categories, only belonging and not belonging two situations. But in fact, most data are uncertain, and a data sample will belong to multiple categories to varying degrees [3]. This hard clustering method ignores the existence of overlapping data samples between the two classes, which is often not logical in application. As Ruspini [4] introduced the concept of fuzzy division on the basis of hard clustering, fuzzy clustering came into being. It can allocate each element in the data set proportionally according to the size of different membership degrees, so that the element belongs to multiple classes. Among them, fuzzy C-means (FCM) clustering algorithm is the most commonly used clustering method in fuzzy clustering, which is more consistent with the actual sample situation, convenient operation, wide application range and other characteristics [5]. In recent years, FCM algorithm and fuzzy logic have been continuously improved, and their application fields are more and more extensive. For example, in 2014, M. A. Sanchez et al. [6] proposed a new fuzzy grain-size gravity clustering algorithm for multi-variable data. Ari et al. [7] proposed generalized Possibilistic Fuzzy C-means with Novel Cluster Validity Indices for Clustering noisy data in 2017.

✉ Jie-Sheng Wang
  wang_jiesheng@126.com

1 School of Electronic and Information Engineering,
   University of Science & Technology Liaoning,
   Anshan 114044, China

In the same year, Elid Rubio et al. [8] extended the fuzzy possibility C-Means (FPCM) algorithm of Type-2 fuzzy logic technology and improved the efficiency of FPCM algorithm. In 2018, F. V. Farahani et al. [9] proposed an intelligent approach based on spatially focused FCM and ensemble learning for the diagnosis of pulmonary nodules. To solve the extended target tracking problem in PHD filter, Bo Yan et al. [10] proposed an improved segmentation algorithm based on FCM algorithm in 2019. In 2020, Haibo Liang et al. [11] proposed the improved SVM-FCM algorithm for rock image segmentation. However, FCM algorithm needs to pass clustering validity verification to judge the optimal partition result of data samples, so finding an appropriate clustering validity function is an important research direction in the field of FCM clustering algorithm [12].

At present, many scholars have proposed various cluster validity functions. However, due to the large differences in the structure and attributes of the data sets, it is impossible for any cluster validity functions to correctly divide all data sets, and no one validity function is always better than others. It is precisely because of this status quo that new clustering validity functions continue to appear. The current research directions of fuzzy clustering validity functions are mainly concentrated in the following two categories:

(1) Fuzzy clustering validity functions based only on membership degree. In 1974, Bezdek et al. firstly proposed the validity function for fuzzy clustering, namely the partition coefficient (V_{PC}) [13]. Subsequently, Bezdek et al. proposed the partition entropy (V_{PE}) based on V_{PC} and Shannon's theorem in Shannon's information theory [14]. Dave et al. proposed an improved partition coefficient (V_{MPC}) in 1996 [15]. In 1999, Fan et al. proposed an improved fuzzy entropy clustering validity function (V_{MPE}) [16], which can suppress the monotonic change of V_{PC} and V_{PE} as the number of clusters increases. Gai-yun Gong et al. redefines the fuzzy partition matrix based on data information in 2004 and proposed a clustering validity function (V_{PF}) based on the partition ambiguity [17]. Zalik et al. proposed a clustering validity function (V_{CO}) that adopts the membership to define overlap and compactness in 2010 [1]. In 2019, Yong-li Liu et al. added the degree of separation module in V_{CO} and proposed a novel clustering validity function [18], which can improve the stability of V_{CO}. Jiashun et al. proposed a clustering validity function (V_{CS}) that can effectively suppress noise data in 2013 [19]. Joopudi et al. used the largest membership degree and the second largest membership degree to measure the

overlap of data in the same year, and proposed the clustering validity function (V_{GD}) [20]. In 2014, Zhang et al. proposed a membership validity function based on two-part modularization [21].

(2) Fuzzy clustering validity functions based on membership degree and geometric structure of data sets. In 1991, Xie and Beni et al. proposed V_{XB} [22] as the first clustering validity function that considered the geometric structure of the data set. In 1996, Bensaid et al. also proposed the clustering validity function (V_{SC}) using the ratio of the degree of compactness within a class and the degree of separation between classes [23]. In 1998, Know et al. introduced a penalty term on the basis of V_{XB} and proposed the clustering effectiveness function (V_K) [24]. In 2005, Kuo-Lung Wu et al. proposed a PCAES clustering validity function (V_{PCAES}) based on exponential operation [25]. In 2019, Zhu et al. proposed a clustering validity function (V_{ZLF}) in the form of a ratio [26]. Ouchicha et al. proposed a standardized superimposed clustering validity function (V_{ECS}) in 2020 [27]. Liu et al. proposed an IMI clustering validity function (V_{IMI}) in 2021 [28]. In 2021, Hong-Yu Wang et al. proposed a clustering validity function (V_{HY}) based on intra-class compactness and inter-class separation [29]. In the same year, Wang et al. proposed a hybrid weighted combination method (HWCVF) [30].

Any clustering validity function is composed of several sub-parts representing different geometric meanings, namely components. Although the excellent performance of clustering validity function emerge in endlessly, but there is no scholar from the view of the components of a validity function for further study. Therefore, this paper puts forward six components (the compactness, similarity, variation degree, the degree of separation between data sets and the overlap degree of clustering to realize the performance evaluation. At the same time, the theoretical basis of these six components is explained in details. Finally, these six components constitute a new validity function of fuzzy C-means clustering. This function can suppress the influence of noise data and divide high-dimensional data and overlapping data accurately. Simulation results show that the proposed validity function can obtain correct clustering results on both artificial and UCI data sets.

In the second section, this article will explain the basic theory and algorithm process of the FCM algorithm, and finally introduce the advantages and disadvantages of traditional cluster effectiveness functions.

2 FCM Clustering Algorithm and Cluster Validity Functions

2.1 FCM Clustering Algorithm

FCM clustering algorithm is based on the fuzzy C-partition of the objective function. Through iteration, the objective function is minimized and optimized to obtain c fuzzy subsets [31]. Suppose there are n samples in the data set, and each sample is p-dimensional, $X = \{x_1, x_2, \cdots, x_n\}$, $X \in R^p$. $V = \{v_1, v_2, \cdots, v_n\}$ is the collection of cluster centers in the data set X, which is a matrix of $c \times p$. The objective function of FCM clustering algorithm was defined as:

$$J_m(U, V) = \sum_{i=1}^{c} \sum_{j=1}^{n} (u_{ij})^m \|x_j - v_i\|^2 \tag{1}$$

where, $J_m(U, V)$ is the square error criterion function, and the minimum value of $J_m(U, V)$ is called the minimum error balance point. In Eq. (1), $m \in (1, \infty)$ represents the fuzzy weighted index to control the degree of fuzzy membership of each group of data. When m infinity approaches 1, FCM clustering algorithm will increasingly tend to the hard clustering algorithm. When $m = 1$, FCM clustering algorithm is equivalent to the K-means clustering algorithm. On the contrary, when m tends to infinity, all data objects x_j and the cluster centers v_i will overlap, and the membership degree of the data objects x_j belonging to each cluster will be the same, whose value is $1/C$. The parameter c is the number of divisions of the fuzzy subset. $\|x_j - v_i\|$ is the Euclidean distance between the cluster center v_i and the data point x_j. $u_{ij}(0 \leq u_{ij} \leq 1)$ is the membership degree of the data point x_j to the cluster center v_i, where $u_{ij} \in U$ and U is a $c \times n$ membership matrix that satisfies $\sum_{i=1}^{c} u_{ij} = 1$ and $0 \leq \sum_{j=1}^{n} u_{ij} \leq n$, where $1 \leq j \leq n$, $1 \leq i \leq c$.

FCM clustering algorithm minimizes the objective function by constantly and iteratively updating the cluster centers and the membership matrix to find the optimal solution for c. The steps of FCM clustering algorithm are described as follows.

Step 1: Given the number of clusters c and the fuzzy index m.

Step 2: Initialize the cluster center V and the membership matrix U, and then obtain U_0 and V_0.

Step 3: Update the fuzzy cluster center V by Eq. (2)

$$v^{(t+1)} = \frac{\sum_{j=1}^{n} (u_{ij}^{(t)})^m x_i}{\sum_{j=1}^{n} (u_{ij}^{(t)})^m} \ among\ them\ 1 \leq i \leq c \tag{2}$$

Step 4: Update the fuzzy partition matrix $U = (u_{ij}^{(t+1)})_{c \times n}$ by Eq. (3).

$$u_{ij}^{(t+1)} = \left[\sum_{k=1}^{c} \left(\frac{\|x_j - v_i\|}{\|x_j - v_k\|} \right)^{1/(m-1)} \right]^{-1} \tag{3}$$

where, $1 \leq i \leq c$, $1 \leq j \leq n$.

Step 5: Calculate the error $e = \|v_{t+1} - v_t\|$. If $e \leq \varepsilon$ (ε is a threshold from 0.001 to 0.01), the loop ends and the final clustering result is obtained, otherwise go to Step 2.

2.2 Clustering Validity Functions

2.2.1 Partition Coefficient (V_{PC}) and Partition Entropy (V_{PE})

Bezdek first proposed the partition coefficient V_{PC} and partition entropy V_{PE} for fuzzy clustering as shown in Eq. (4) and Eq. (5), respectively.

$$V_{PC} = \frac{1}{n} \sum_{i=1}^{c} \sum_{j=1}^{n} (u_{ij})^2 \tag{4}$$

$$V_{PE} = -\frac{1}{n} \sum_{i=1}^{c} \sum_{j=1}^{n} [u_{ij} \log_a(u_{ij})] \tag{5}$$

The principle of partition coefficient and partition entropy is simple, and the amount of calculation is small. But they only consider the similarity within the data set, and lack the connection with the geometric structure of the data set. As the number of clusters increases, V_{PC} and V_{PE} will show a monotonic trend, which makes it impossible to divide the data set correctly. V_{PC} and V_{PE} take the minimum and maximum c, respectively, as the optimal number of clusters.

2.2.2 Improved Partition Coefficient (V_{MPC})

The improved partition coefficient V_{MPC} corrects the existing monotonic reduction problem of V_{PC}, but still lacks a direct connection with the geometric structure of the data set. c with the maximum value of V_{MPC} is the best clustering result, and V_{MPC} can be calculated by:

$$V_{MPC} = 1 - \frac{c}{c-1}(1 - V_{PC}) \tag{6}$$

2.2.3 Xie-Beni Clustering Validity Function (V_{XB})

The calculation of V_{XB} can be realized by:

$$V_{XB} = \frac{\sum_{i=1}^{c} \sum_{j=1}^{n} (u_{ij})^m \lVert x_j - v_i \rVert^2}{n \min_{i \neq j} \lVert v_i - v_j \rVert^2} \tag{7}$$

The numerator of V_{XB} represents the tightness, and the smaller the numerator, the tighter the data within the class. The denominator of V_{XB} is used to define the degree of separation. The larger the denominator, the more separated the data between classes. Therefore, when V_{XB} takes the minimum value, c is the optimal number of clusters. However, this clustering validity function has two shortcomings. (1) When $c \to n$, V_{XB} will tend to zero. (2) When $m \to \infty$, V_{XB} will become infinite.

2.2.4 Knows Clustering Validity Function (V_K)

Know et al. introduced a penalty term $\frac{1}{c} \sum_{i=1}^{c} \lVert v_i - \overline{v} \rVert^2$ on the basis of V_{XB}, which successfully suppressed the monotonically decreasing trend of V_{XB} when $c \to n$. But when $m \to \infty$, V_{XB} will become infinite. V_K is also valid to take the minimum value, and calculated by Eq. (8).

$$V_K = \frac{\sum_{i=1}^{c} \sum_{j=1}^{n} (u_{ij})^2 \lVert x_j - v_i \rVert^2 + \frac{1}{c} \sum_{i=1}^{c} \lVert v_i - \overline{v} \rVert^2}{\min_{i \neq j} \lVert v_i - v_j \rVert^2} \tag{8}$$

2.2.5 PCAES Clustering Validity Function (V_{PCAES})

V_{PCAES} is a clustering validity function with an exponential operation method shown in Eq. (9). The first part of V_{PCAES} is the ratio of the sum of the squares of the degree of membership and the minimum degree of membership. The second part of V_{PCAES} represents the relative distance between the cluster centers, and the adoption of exponential operation can suppress the monotonic trend of $(-\min_{k \neq i} \lVert v_i - v_j \rVert^2)/\beta_T$.

$$V_{PCAES} = \sum_{i=1}^{c} \sum_{j=1}^{n} \frac{u_{ij}^2}{u_M} - \sum_{i=1}^{c} \exp \left(\frac{-\min_{k \neq i} \lVert v_i - v_k \rVert^2}{\beta_T} \right) \tag{9}$$

where, $u_M = \min_{1 \leq i \leq c} \sum_{j=1}^{n} u_{ij}^2$, $\beta_T = \frac{1}{c} \sum_{i=1}^{c} \lVert v_i - \overline{v} \rVert^2$, $\overline{v} = \frac{1}{c} \sum_{i=1}^{c} v_i$.
The maximum value of V_{PCAES} is valid.

2.2.6 Tang Clustering Validity Function (V_T)[32]

V_T is a new fuzzy clustering validity function proposed by Tang et al. which is based on V_K [23]. The second term in

its numerator is a penalty term, which represents the distance between any two different cluster centers. It can also suppress the problem of monotonically decreasing of V_{XB} when $c \to n$. But unlike V_K, the penalty term $1/c$ in the denominator of V_T can solve this problem that V_{XB} will become infinite when $m \to \infty$. This validity function can be defined as:

$$V_T = \frac{\sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m \lVert x_j - v_i \rVert^2 + \frac{1}{c(c+1)} \sum_{i=1}^{c} \sum_{k=1; k \neq i}^{c} \lVert v_i - v_k \rVert^2}{\min_{i \neq k} \lVert v_i - v_k \rVert^2 + \frac{1}{c}} \tag{10}$$

2.2.7 P clustering Validity Function (V_P)[33]

Chen and Links proposed a clustering validity function V_P based only on membership degree in 2004, as shown in Eq. (11). The first half of V_P represents the sum of the maximum values of the data in each category. The larger the value, the better the tightness of the data within the category. The second part of V_P represents the degree of similarity between classes, and is used to judge the membership degree of the intersection between v_i and v_j. The smaller the value, the more separation between classes. When taking the maximum value of V_P, c is the optimal number of clusters.

$$V_P = \frac{1}{n} \sum_{j=1}^{n} \max_i (u_{ij}) - \frac{1}{k} \sum_{i=1}^{c-1} \sum_{j=i+1}^{c} \left[\frac{1}{n} \sum_{k=1}^{n} \min(u_{ik}, u_{ik}) \right] \tag{11}$$

2.2.8 WL Clustering Validity Function (V_{WL})[34]

V_{WL} is a cluster validity function proposed by Chih-Hung Wu et al., as shown in Eq. (12). The numerator of V_{WL} adopts the sum of the average Euclidean distances from the data points to all cluster centers to define the compactness. The first term of the denominator represents the compactness between clusters, and the second term is the median distance between all cluster centers. For a uniformly distributed data set, it has a better classification performance.

$$V_{WL} = \frac{\sum_{i=1}^{c} \left(\frac{\sum_{j=1}^{n} u_{ij}^2 \lVert x_j - v_i \rVert^2}{\sum_{j=1}^{n} u_{ij}} \right)}{\min_{i \neq k} \lVert v_i - v_k \rVert^2 + median \lVert v_i - v_k \rVert^2} \tag{12}$$

2.2.9 FM Clustering Validity Function (V_{FM})[35]

Taking the partition entropy (V_{PE}) and the fuzzy partition factor (α_f) into consideration, the compactness and separation of clustering are defined, as shown in Eq. (13).

However, this clustering validity function does not perform well on noisy data sets.

$$V_{FM} = \alpha_f \times \left\{ -\frac{1}{n} \sum_{i=1}^{c} \sum_{j=1}^{n} \left[u_{ij} \log_a(u_{ij}) \right] \right\} \quad (13)$$

where, $\alpha_f = \frac{\sum_{i=1}^{c} \sum_{j=1}^{n} (u_{ij} - \frac{1}{c})^2 \|x_j - v_i\|^2}{n \min_{i \neq k} \|v_i - v_k\|^2}$. V_{FM} is valid with the minimum value.

2.2.10 PBMF Clustering Validity Function (V_{PBMF})[36]

In 2005, Pakhiar et al. proposed a cluster validity function in the form of a product, as shown in Eq. (14). The first item of the product is the reciprocal of the number of clusters, and the second item of the product is used to measure the intra-class compactness, which represents the ratio of the intra-class distance when the entire data set is divided into one category and classified into c. The last item of V_{PBMF} indicates the maximum separation. When V_{PBMF} takes the maximum value, c is the optimal number of clusters.

$$V_{PBMF} = \left(\frac{1}{c} \times \frac{\sum_{j=1}^{n} \|x_j - v_i\|}{\sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m \|x_j - v_i\|^2} \times \max_{i,j=1}^{c} \|v_i - v_j\| \right)^2 \quad (14)$$

2.2.11 Clustering Validity Function Proposed by Zhu (V_{ZLF})

V_{ZLF} is composed of the ratio of compactness and separation, which is shown in Eq. (15).

$$V_{ZLF} = \frac{comp}{sep} = \frac{\sum_{j=1}^{n} \frac{1 - \max_i u_{ij}}{\sum_{i=1}^{c} \|x_j - v_i\|}}{\sum_{k=1}^{c} \sum_{i=1; i \neq k}^{c} \|v_i - v_k\| \Big/ \frac{c(c-1)}{2}} \quad (15)$$

where, $\max_i u_{ij}$ is the maximum membership degree of all samples in a certain class. If the data in the class is more compact, $\max_i u_{ij} \to 1$ and $(1 - \max_i u_{ij}) \to 0$. The smaller the value, the better the clustering performance. $\|x_j - v_i\|$ in the numerator is the compactness within the class. The denominator of V_{ZLF} is the average of the sum of the distances between cluster centers, which defines the degree of separation between clusters. $\|v_i - v_k\|$ is the distance between cluster centers, and $c(c-1)/2$ is the number of distances between cluster centers. When V_{ZLF} takes the minimum value, c is the optimal number of clusters.

2.2.12 ECS Clustering Validity Function (V_{ECS})

The V_{ECS} is proposed by Ouchicha in 2020, which is defined in Eq. (16).

$$V_{ECS} = \frac{SC - SC_{\min}}{SC_{\max} - SC_{\min}} + \frac{PEC - PEC_{\min}}{PEC_{\max} - PEC_{\min}} \quad (16)$$

where, the SC was defined as $\left(\frac{1}{n} \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij} \|x_j - v_i\|^2\right) / \frac{1}{c} \sum_{i=1}^{c} \|v_i - \bar{v}\|^2$, and the PEC was defined as $\left(\frac{1}{n(1-m)} \sum_{i=1}^{c} \sum_{j=1}^{n} \log(u_{ij}^m + (1 - u_{ij})^m)\right) / V_{PC}$. V_{ECS} was the normalized of SC and PEC. The minimum value of V_{ECS} corresponds to the optimal number of clusters.

In the next section, six multiple clustering performance evaluation components are introduced in details so as to form a new fuzzy cluster effectiveness function V_{WG}.

3 FCM Clustering Validity Function Based on Multiple Clustering Performance Evaluation Components

3.1 Clustering Performance Evaluation Components

Based on the FCM clustering algorithm, fuzzy membership and the geometric structure of the data set, this paper defines six clustering performance evaluation components, including compactness, variation, overlap, similarity, and separation.

3.1.1 Compactness (comp)

Equation (17) is the criterion function of the FCM clustering algorithm, which represents the sum of the distance between the cluster center v_i and the data sample x_j. The smaller the value, the higher the similarity of the data within the class, and the closer the data within the class.

$$comp = \sum_{i=1}^{c} \sum_{j=1}^{n} (u_{ij})^2 \|x_j - v_i\|^2 \quad (17)$$

3.1.2 Variation (var) [37]

Equation (18) adopts the exponential function to suppress the characteristics of noise data interference, which is used to measure the compactness of the data within the class.

Similarly, the smaller the value, the more compact the data in the class.

$$\text{var} = \frac{1}{n} \sum_{i=1}^{c} \sum_{j=1}^{n} \left(\exp -\frac{\|x_j - v_i\|^2}{\varepsilon} \right) \tag{18}$$

where, $\varepsilon = \frac{1}{n} \sum_{j=1}^{n} \|x_j - \overline{x}\|^2$ ($\overline{x} = \frac{1}{n} \sum_{j=1}^{n} x_j$ is the mean value of the data samples) represents the average sum of the distance between the data within the class and the mean value of the data samples. Equation (18) is used to measure the degree of variation of the data within the class. The smaller the calculated value, the more stable the data within the class.

3.1.3 Overlap (overlap)

Equation (19) is based on the degree of membership to express the degree of overlap between data sets. If the data point x_k in the overlapping part belongs to the i-th cluster, then the membership degree u_{ik} will be close to 1, and if the data point x_k belongs to the j-th cluster, the degree of membership u_{jk} will be close to 0. If the data in the overlapping part are separated better, the difference of $|u_{ik} - u_{jk}|$ will be closer to 1, and the value of $1 - |u_{ik} - u_{jk}|$ will be smaller. Therefore, a small value for this item indicates a better classification effect.

$$overlap = \min_{i \neq j} \left(\frac{1}{n} \sum_{j=1}^{n} \left(1 - |u_{ik} - u_{jk}| \right) \right) \tag{19}$$

3.1.4 Similarity (sim)

Equation (20) represents the sum of the squares of the membership degrees of the data point x_j belonging to the cluster center v_i. The larger the value of this term, the more similar the data within the cluster.

$$sim = \min_{1 \leq i \leq c} \sum_{j=1}^{n} (u_{ij})^2 \tag{20}$$

3.1.5 Separation (sep)

Equation (21) and Eq. (22) are two clustering performance evaluation components used to define the degree of overlap between clusters. $sep_1 = \min_{i \neq k} \|v_i - v_k\|^2$ indicates the minimum distance between any two clusters. The larger the value, the better the separation between classes. $sep_2 = \frac{1}{c} \sum_{i=1}^{c} \|v_i - \overline{v}\|^2$ ($\overline{v} = \frac{1}{c} \sum_{i=1}^{c} v_i$ is the average of all cluster centers) represents the sum of the average distances

from all cluster centers to the balance point of the cluster centers. The larger the value, the more separated the classes are from each other.

$$sep_1 = \min_{i \neq k} \|v_i - v_k\|^2 \tag{21}$$

$$sep_2 = \frac{1}{c} \sum_{i=1}^{c} \|v_i - \overline{v}\|^2 \tag{22}$$

3.2 FCM Clustering Validity Function Based on Multiple Clustering Performance Evaluation Components

Based on the six clustering performance evaluation components (compactness, variability, overlap, similarity, and separation), a ratio type of clustering validity function was constituted, which is defined as follows:

$$V_{WG} = \frac{comp + \text{var} + overlap}{sim + sep_1 + sep_2}$$

$$= \frac{\sum_{i=1}^{c} \sum_{j=1}^{n} (u_{ij})^2 \|x_j - v_i\|^2 + \frac{1}{n} \sum_{i=1}^{c} \sum_{j=1}^{n} \exp(-\frac{\|x_j - v_i\|^2}{\varepsilon}) + \min_{i \neq j}(\frac{1}{n} \sum_{j=1}^{n} (1 - |u_{ik} - u_{jk}|))}{\min_{1 \leq i \leq c} \sum_{j=1}^{n} (u_{ij})^2 + \min_{i \neq k} \|v_i - v_k\|^2 + \frac{1}{c} \sum_{i=1}^{c} \|v_i - \overline{v}\|^2} \tag{23}$$

where, $\varepsilon = \frac{1}{n} \sum_{j=1}^{n} \|x_j - \overline{x}\|^2$, $\overline{x} = \frac{1}{n} \sum_{j=1}^{n} x_j$, $\overline{v} = \frac{1}{c} \sum_{i=1}^{c} v_i$.

Then V_{WG} is applied to the FCM clustering algorithm, and the algorithm flowchart for obtaining the optimal number of clusters is shown in Fig. 1. The flow of the FCM clustering algorithm based on V_{WG} is described as follows.

Step 1: Given the maximum number of clusters $c_{\max}(c_{\max} \leq \sqrt{n})$, the maximum number of iterations $I_{\max}$, the fuzzy index m $(1.5 \leq m \leq 2.5)$ and the termination threshold ε.

Step 2: Initialize the membership matrix U, the cluster center V and the number of clusters c.

Step 3: Update the fuzzy partition matrix $U^{(t+1)}$ and cluster center $V^{(t+1)}$ and judge whether $e = \|v_{t+1} - v_t\|$ is less than ε. If $e < \varepsilon$, go to Step 4. Otherwise, if $e \geq \varepsilon$, then go to Step 2.

Step 4: Let $c = c + 1$, use FCM clustering algorithm to calculate the minimum value of V_{WG}, and obtain the optimal solution of c. If $c < c_{\max}$, repeat Step 2. If $c \geq c_{\max}$, go to Step 5.

Step 5: Select the number of clusters $\min\{V_{WG}(U, V, c_o)\}$ corresponding to c_o as the optimal number of clusters, and finally output the value of V_{WG}.

Next, the proposed validity function and eight typical validity functions are compared with five artificial data sets and eight UCI data sets to verify the effectiveness of the proposed V_{WG}.
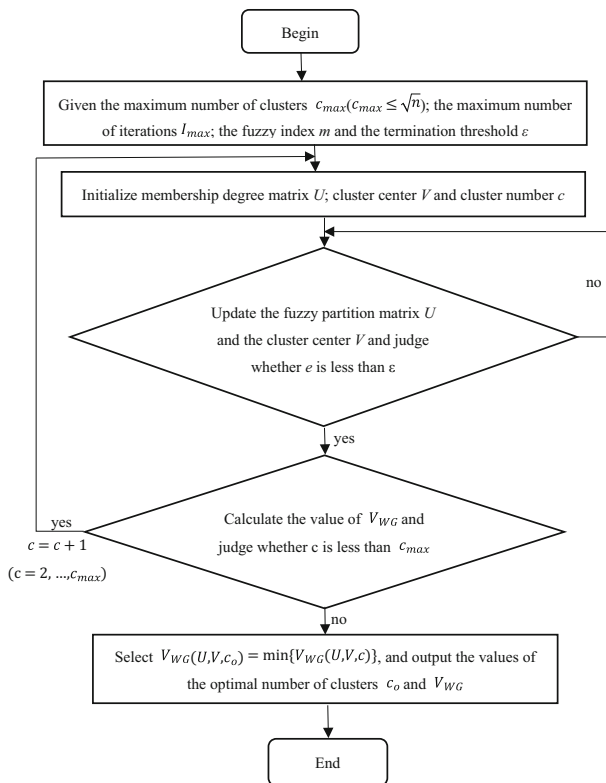
Fig. 1 Flowchart of FCM clustering algorithm based on V_{WG}

4 Simulation Experiments and Results Analysis

4.1 Testing Artificial Data Sets

In order to verify the validity of the proposed validity function (V_{WG}), this paper selects 8 typical cluster validity functions (V_{MPC}, V_{XB}, V_T, V_P, V_{PCAES}, V_{FM}, V_{WL} and V_{ZLF}) for comparison experiments. According to the prior knowledge [38], the fuzzy index $1.5 \leq m \leq 2.5$ and the number of clusters $2 \leq c \leq \sqrt{n}$ can be determined. This paper chooses $m = 2$, $2 \leq c \leq 14$ to conduct simulation experiments, and then judge whether the classification of each clustering validity function is accurate when using different data sets. Four artificial data sets selected to carry out the simulation experiments are listed in Table 1. Among them, Data_2_3 (noise) in the artificial data set is a Gaussian 2-dimensional 3-category data set with 100 noise data added, Data_2_3 (overlap) is a Gaussian 2-dimensional 3-category data set with overlapping data between classes, Data_3_3 is a 3-dimensional 3-type data set that obeys Gaussian distribution, Data_3_6 is a 3-dimensional

6-type data set that obeys a uniform distribution. Data_2_5 is a 2-dimensional 5-type data set that obeys a uniform distribution. The samples distributions of these four artificial data sets are shown in Fig. 2a–e.

4.2 Simulation Experiments and Results Analysis on Artificial Data Sets

The experimental results on the artificial data sets (Data_2_3 with noise, Data_2_3 with overlap, Data_3_3, Data_3_6 and Data_2_5) are shown in Fig. 3a–i to Fig. 7a–i. It can be seen from Fig. 3 that when classifying the noise data set Data_2_3(noise), except for V_{PCAES} and V_{FM}, all other clustering validity functions can find the optimal number of clusters. Figure 4 shows that for the overlapping data set Data_2_3(overlap), only V_{WG} can get the optimal number of clusters $c = 3$. Figure 5 shows that when the dimension of the data set is increased to 3, V_{XB}, V_P, V_{ZLF}, V_{WG} can still determine the optimal number of clusters $c = 3$. As can be seen in Fig. 6, only V_{WG} can be classified into 6 categories. Finally, as shown in Fig. 7, only V_{MPC}, V_P and V_{WG} can find the optimal clustering number c = 5. The simulation results on five artificial data sets show that only V_{WG} can obtain the optimal number of clusters for the five sets of artificial data sets. Among them, V_{XB}, V_{WL}, and V_{MPC} can distinguish the optimal number of clusters for 2 sets of data sets, V_P can divide 3 groups of data sets, V_T and V_{ZLF} can only divide 1 group of data sets, but for 5 sets of artificial data sets, V_{PCAES} and V_{FM} cannot get the correct number of clusters. The above experiment shows that V_{WG} is better than these typical clustering validity functions. This is because when the data set is affected by noisy data, overlapping data and high-dimensional data, V_{WG} is less disturbed.

In order to better observe the changing trend of each clustering performance evaluation component of V_{WG} on the artificial data sets, six clustering performance evaluation components of V_{WG} are normalized and placed in the

Table 1 Artificial data sets

Data sets	Data numbers	Attributes	Classes
Data_2_3(noise)	400	2	3
Data_2_3(overlap)	300	2	3
Data_3_3	500	3	3
Data_3_6	500	3	6
Data_2_5	110	2	5

(a) Data_2_3(noise)

(b) Data_2_3(overlap)

(c) Data_3_3

(d) Data_3_6

(e) Data_2_5

Fig. 2 Artificial data sets

(a) V_{MPC} (c=3)

(b) V_{XB} (c=3)

(c) V_T (c=3)

(d) V_P (c=3)

(e) V_{PCAES} (c=13)

(f) V_{FM} (c=2)

(g) V_{WL} (c=3)

(h) V_{ZLF} (c=3)

(i) V_{WG} (c=3)
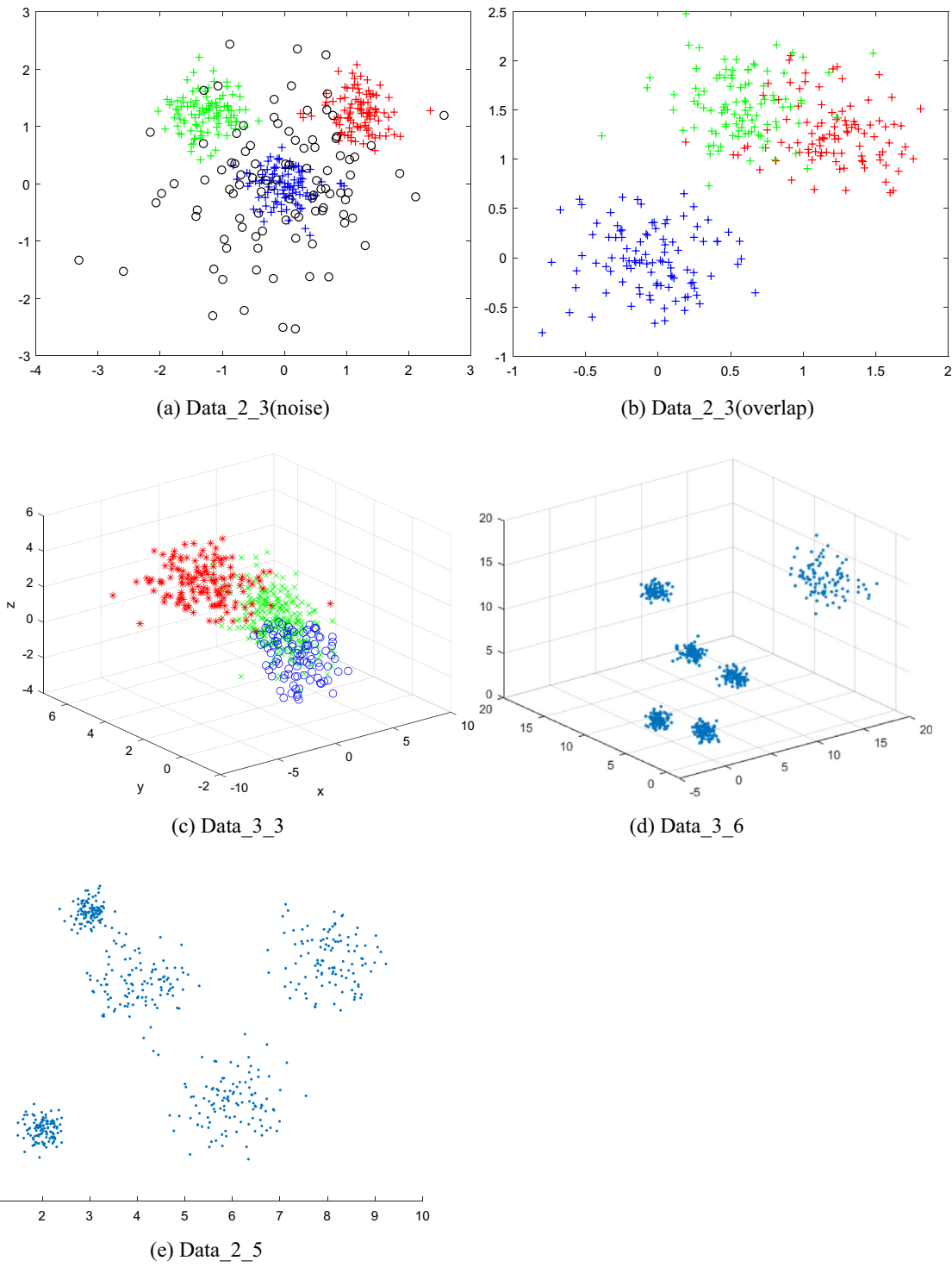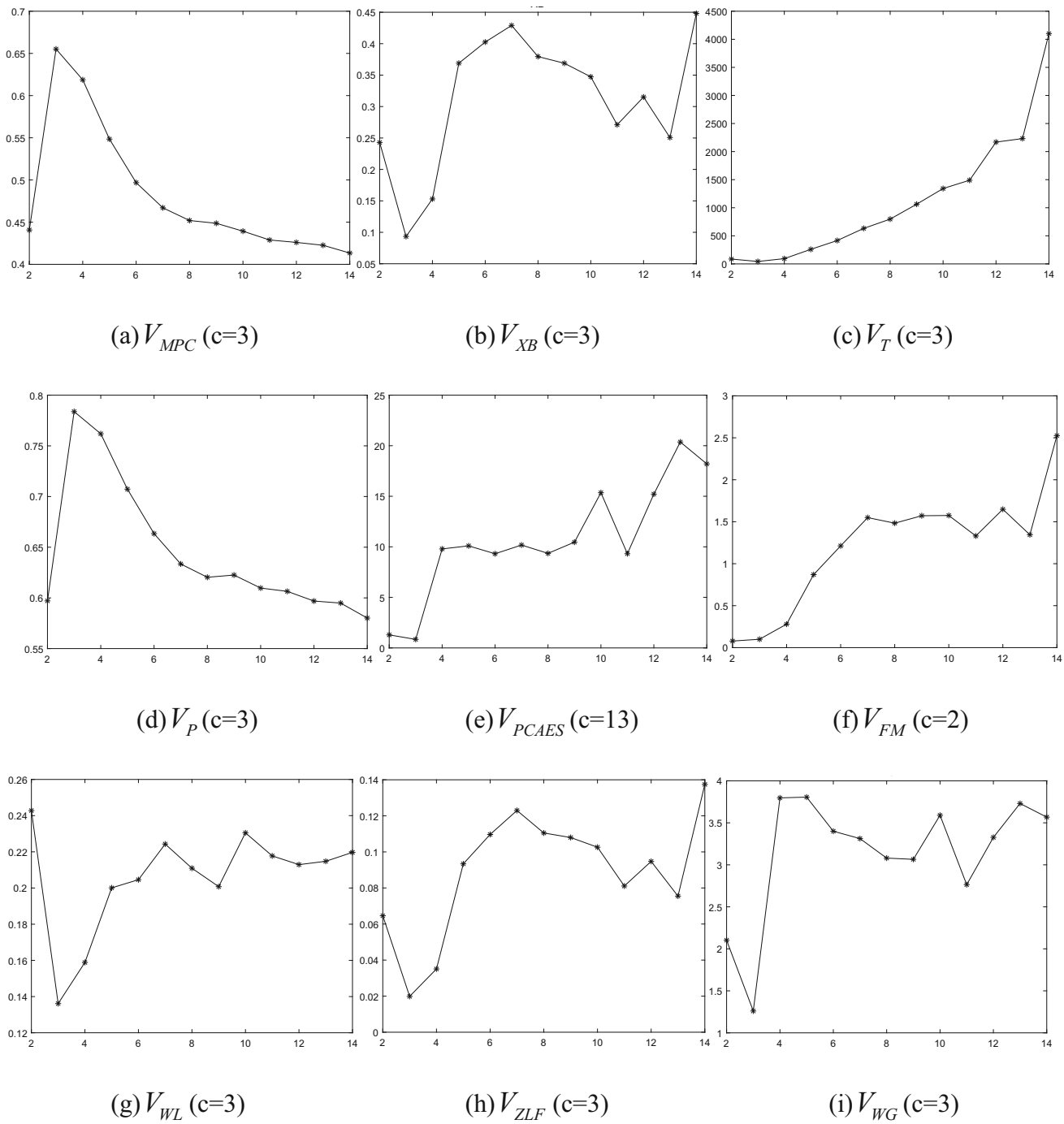
Fig. 3 Change trend of clustering validity functions under Data_2_3 (noise) data set

(a) V_{MPC} (c=2)

(b) V_{XB} (c=2)

(c) V_T (c=2)

(d) V_P (c=2)

(e) V_{PCAES} (c=14)

(f) V_{FM} (c=2)

(g) V_{WL} (c=2)

(h) V_{ZLF} (c=2)

(i) V_{WG} (c=3)
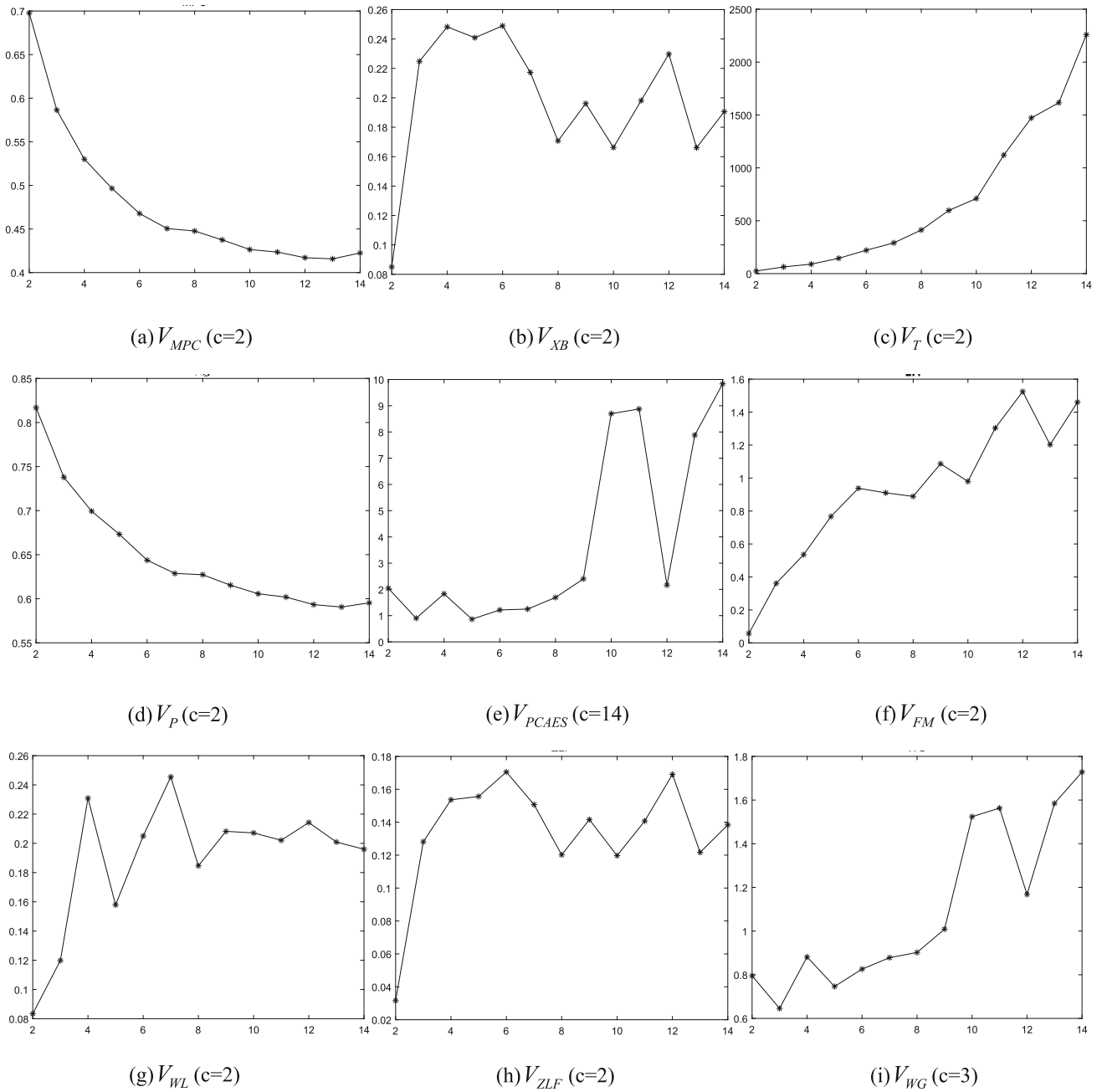
Fig. 4 Change trend of clustering validity functions under Data_2_3 (overlap) data set

(a) V_{MPC} (c=2)   (b) V_{XB} (c=3)   (c) V_T (c=2)

(d) V_P (c=3)   (e) V_{PCAES} (c=10)   (f) V_{FM} (c=2)

(g) V_{WL} (c=3)   (h) V_{ZLF} (c=2)   (i) V_{WG} (c=3)
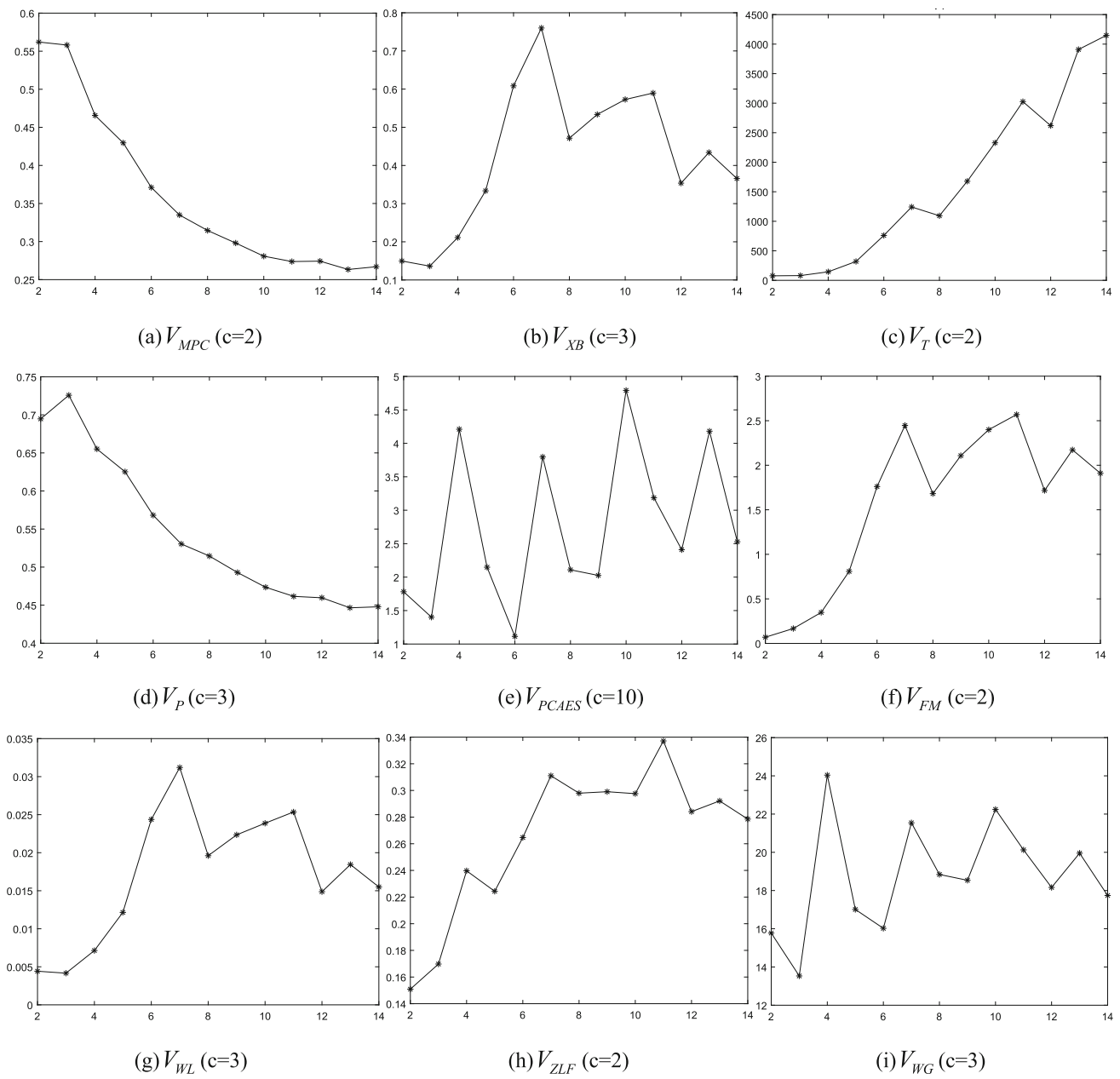
Fig. 5 Change trend of clustering validity functions under Data_3_3 data set
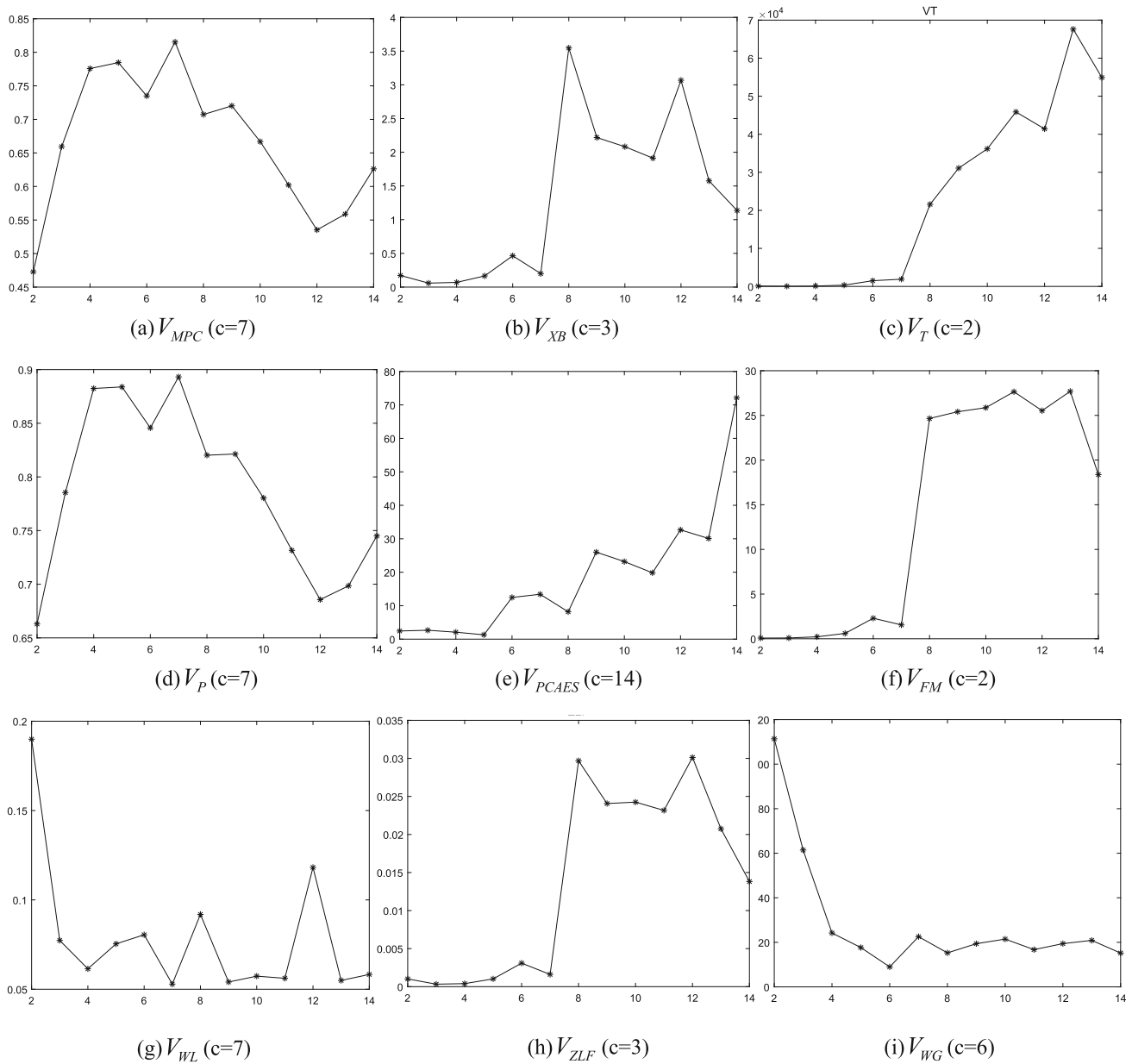
Fig. 6 Change trend of clustering validity functions under Data_3_6 data set

same coordinate system, which is shown in Fig. 8a–e. Next, The function values of V_{MPC}, V_{XB}, V_T, V_P, V_{PCAES}, V_{FM}, V_{WL}, V_{ZLF} and V_{WG} on the artificial data sets are

placed in the normalized coordinate system, which are shown in Fig. 9a–e. In this way, the clustering effect of V_{WG} and other validity functions can be compared more
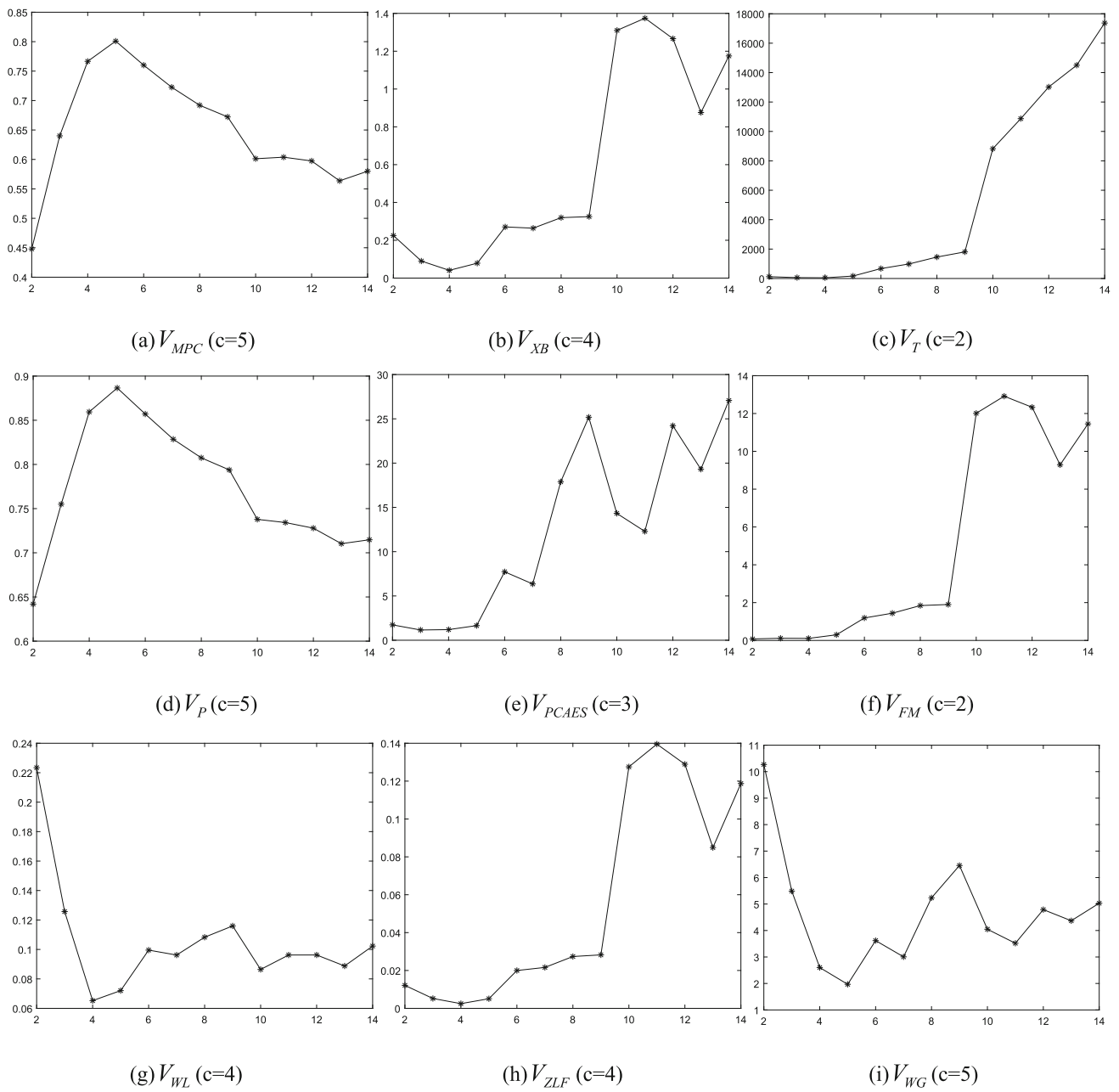
Fig. 7 Change trend of clustering validity functions under Data_2_5 data set
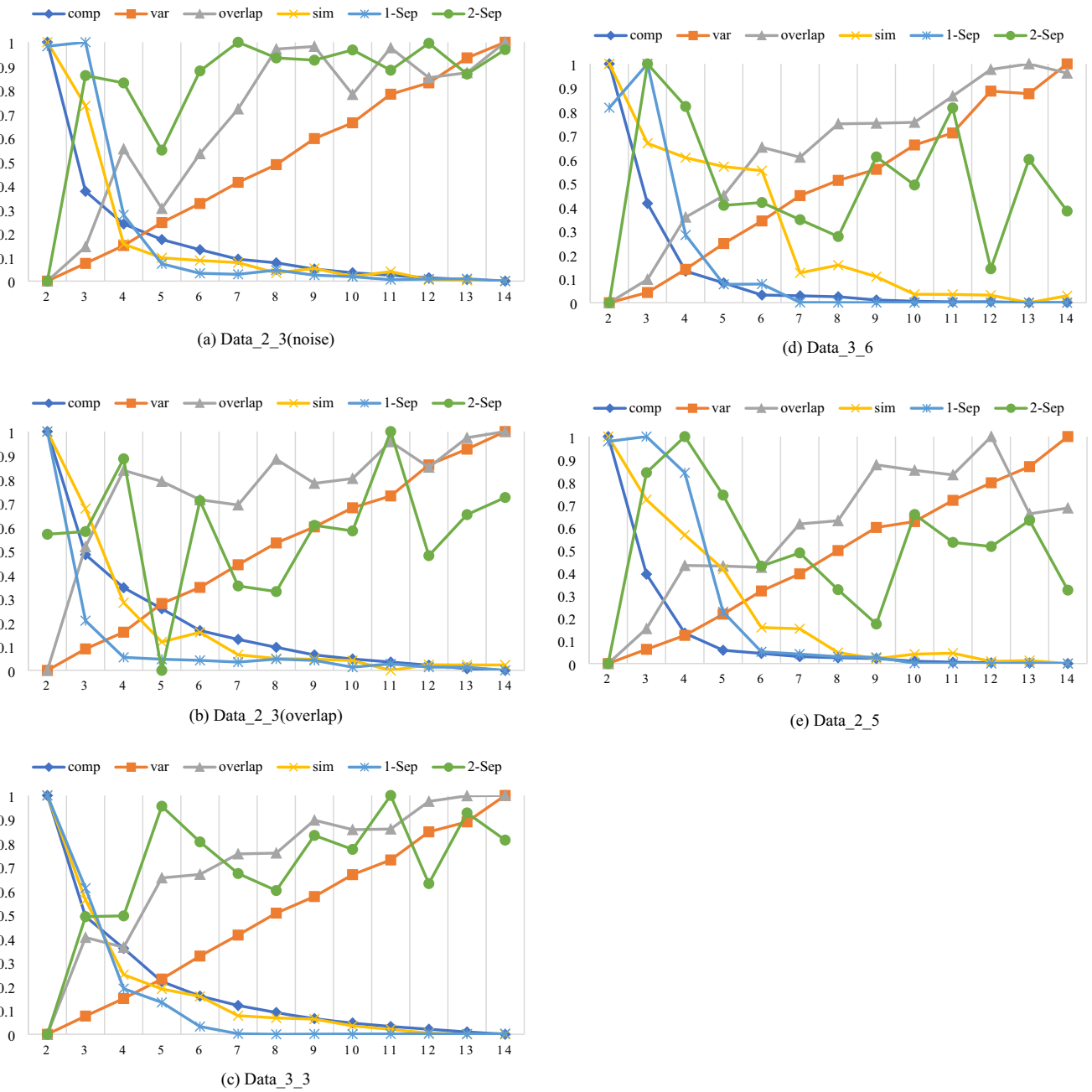
(a) V_{MPC} (c=5)

(b) V_{XB} (c=4)

(c) V_T (c=2)

(d) V_P (c=5)

(e) V_{PCAES} (c=3)

(f) V_{FM} (c=2)

(g) V_{WL} (c=4)

(h) V_{ZLF} (c=4)

(i) V_{WG} (c=5)

(a) Data_2_3(noise)

(d) Data_3_6

(b) Data_2_3(overlap)

(e) Data_2_5

(c) Data_3_3

Fig. 8 Changing trend of each component in V_{WG} under artificial data sets
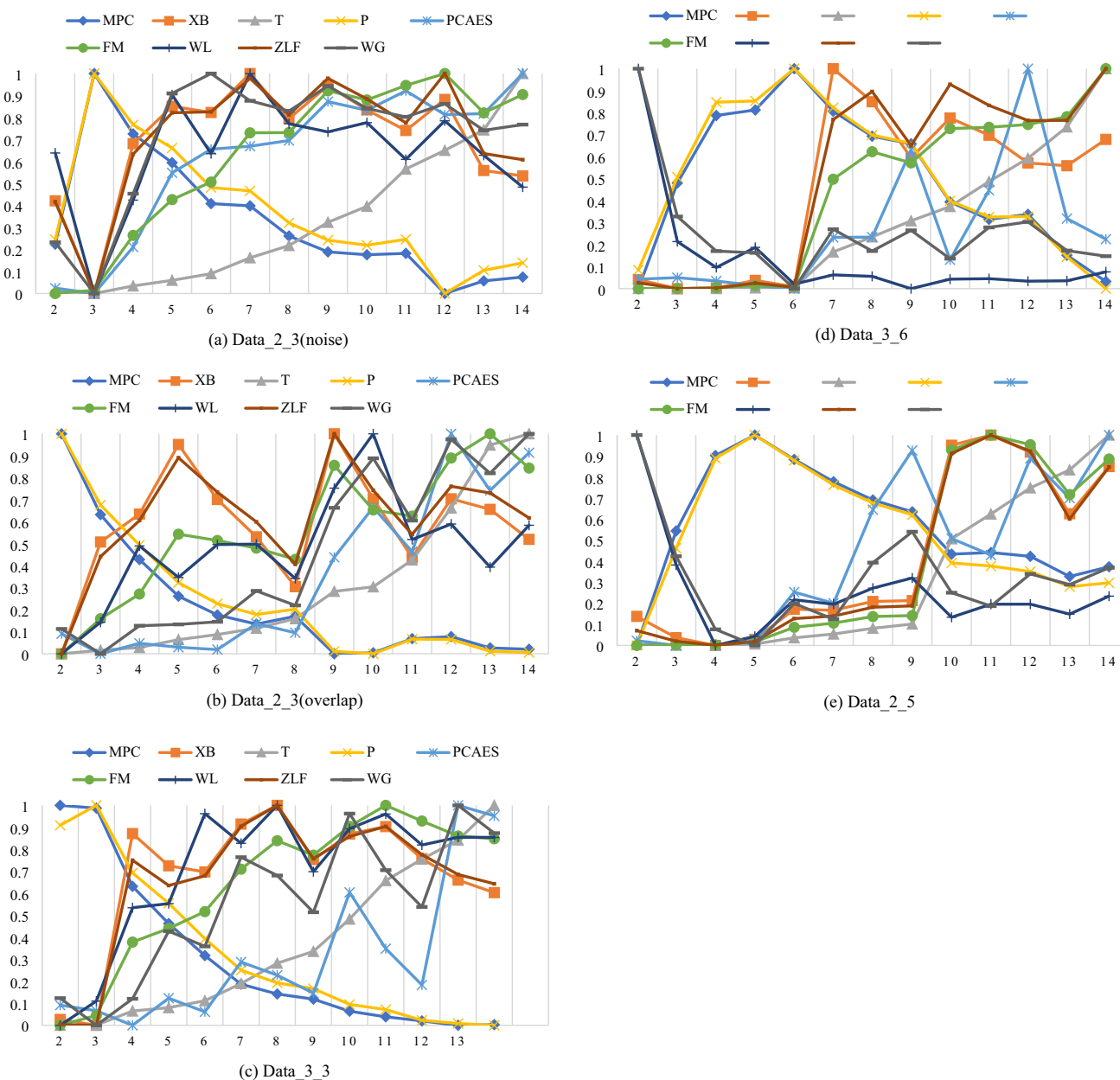
Fig. 9 Changing trend of normalized clustering validity functions under artificial data sets

Table 2 Optimal number of clusters under different validity functions for artificial data sets

Data	Optimal c	V_{MPC}	V_{XB}	V_T	V_P	V_{PCAES}	V_{FM}	V_{WL}	V_{ZLF}	V_{WG}
Data_2_3(noise)	3	**3**	**3**	**3**	**3**	13	2	**3**	**3**	**3**
Data_2_3(overlap)	3	2	2	2	2	14	2	2	2	**3**
Data_3_3	3	2	**3**	2	**3**	10	2	**3**	2	**3**
Data_3_6	6	7	3	2	7	14	2	7	3	**6**
Data_2_5	5	**5**	4	2	**5**	3	2	4	4	**5**

The significance of bold indicate that clustering validity functions achieve the optimal cluster number

Table 3 UCI data sets
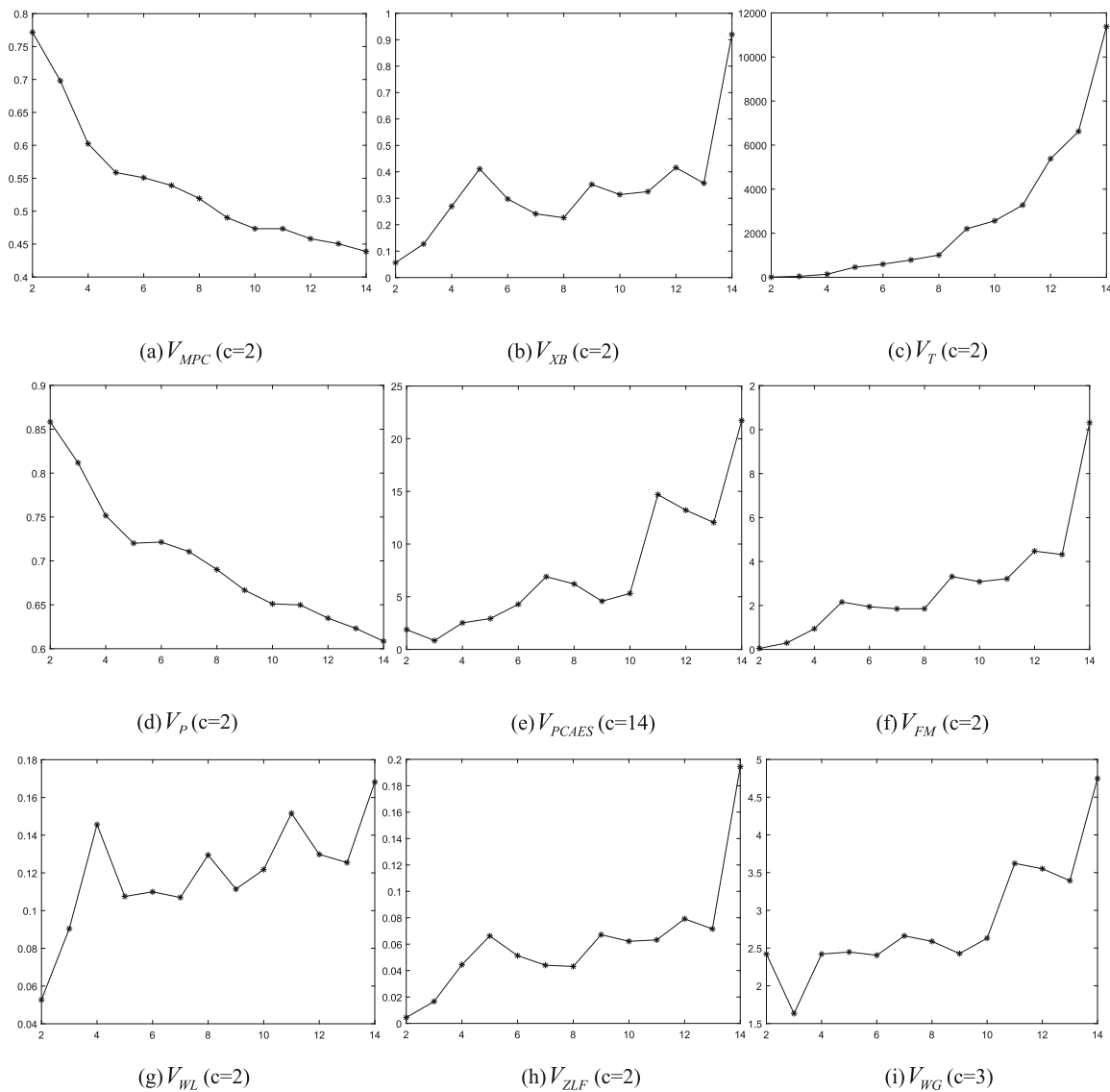
| Data sets	Data numbers	Attributes	Classes
Iris	150	4	3
Seeds	210	7	3
Phoneme	300	5	2
Haberman	306	3	2
HTUR2	17,898	9	2
Hfcr	299	13	4
Segment	2310	19	7
Heart	270	13	2

intuitively. Finally, for different artificial data sets, the optimal number of clusters of each clustering validity are listed in Table 2.

4.3 Simulation Experiments and Results Analysis on UCI Data Sets

Due to the simple structure of the artificial data sets and the small number of samples, in order to verify the effectiveness of V_{WG} clustering on the complex data sets, the following simulation experiments will be carried out using UCI data sets. The real data sets selected in the experiment are Iris, Seeds, Phoneme, Haberman, HTUR2, Hfcr,

(a) V_{MPC} (c=2)　　　　(b) V_{XB} (c=2)　　　　(c) V_T (c=2)

(d) V_P (c=2)　　　　(e) V_{PCAES} (c=14)　　　　(f) V_{FM} (c=2)

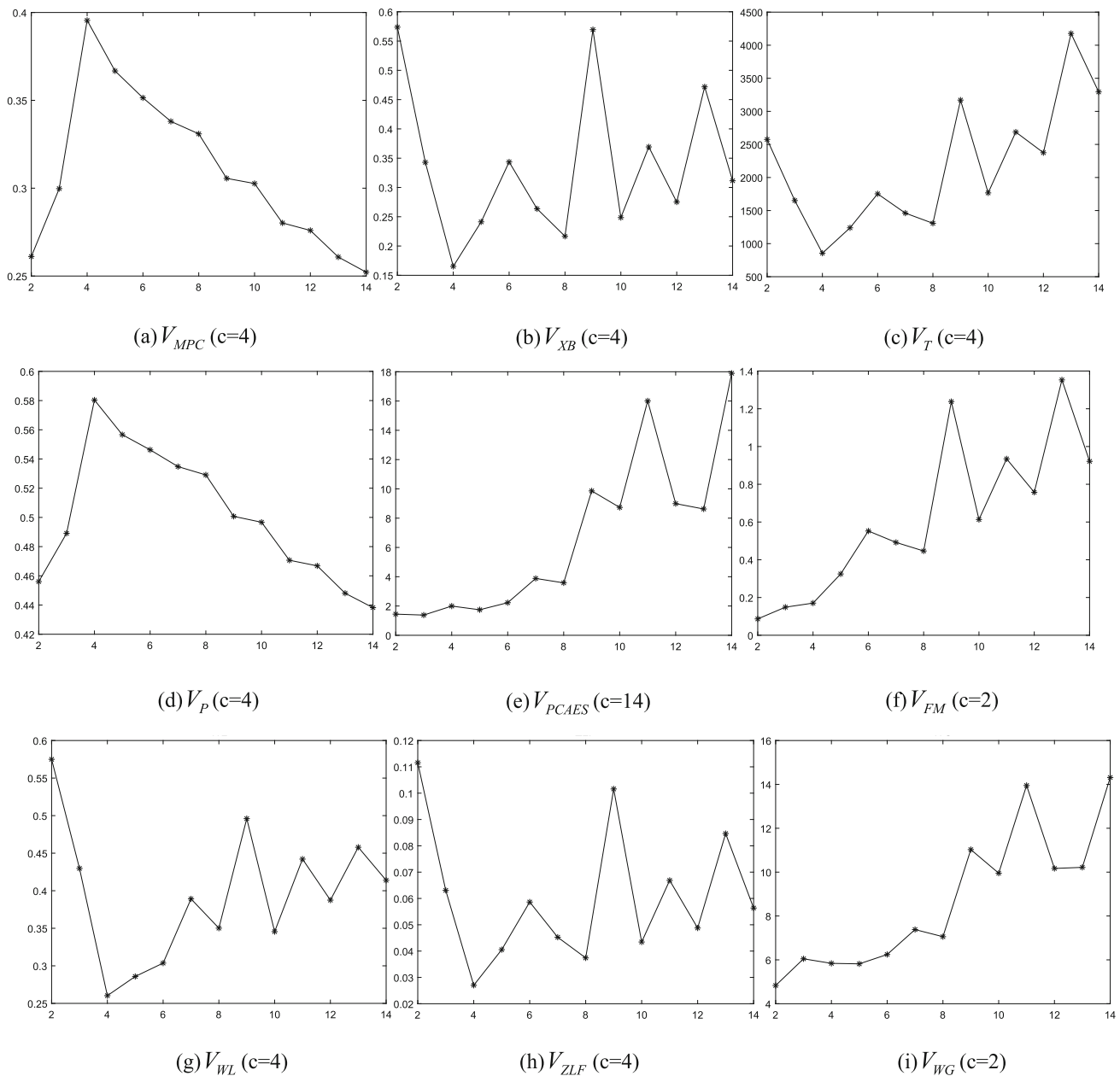(g) V_{WL} (c=2)　　　　(h) V_{ZLF} (c=2)　　　　(i) V_{WG} (c=3)

Fig. 10 Change trend of clustering validity functions under Iris data set
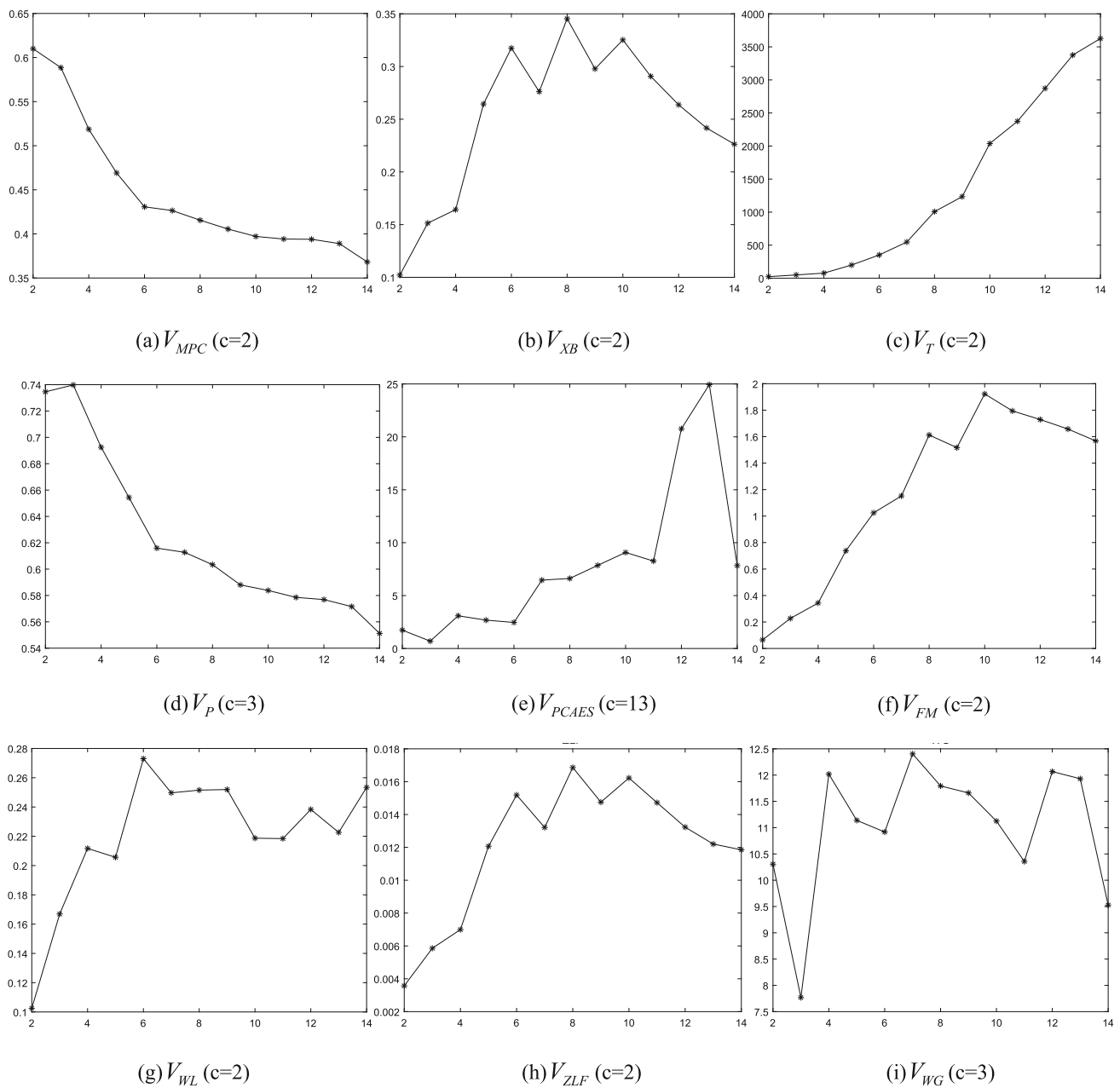
Fig. 11 Change trend of clustering validity functions under Seeds data set

Segment and Heart in the UCI database. The data volume, category and attributes of each UCI data set are listed in Table 3. The experimental results on UCI data sets are shown in Fig. 10–Fig. 17a–i. It can be seen from Fig. 10 that only V_{WG} can get the optimal number of clusters $c = 3$. Figure 11 shows that V_P and V_{WG} can determine the best

Fig. 12 Change trend of clustering validity functions under Phoneme data set

clustering result $c = 3$. It can be seen from the experimental results in Fig. 10 and Fig. 11 that when processing data sets with higher dimension such as Iris and Seeds, the optimal number of clusters can be obtained by V_{WG}. This
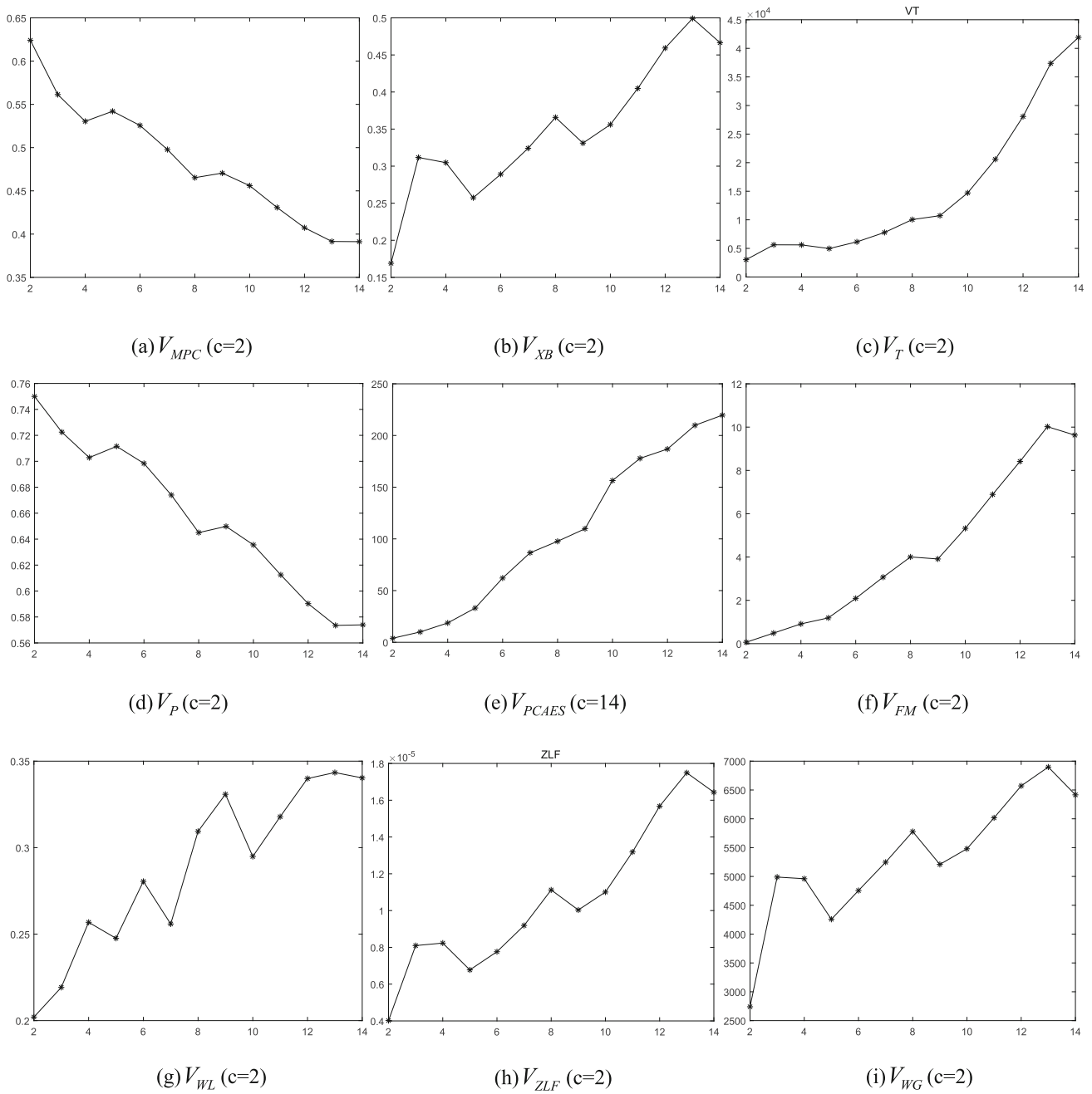
shows that the number of clusters is relatively high in the data set. In this case, the classification ability of V_{WG} is better than other typical clustering validity functions. It can be seen from Fig. 12 that only V_{FM} and V_{WG} can divide two
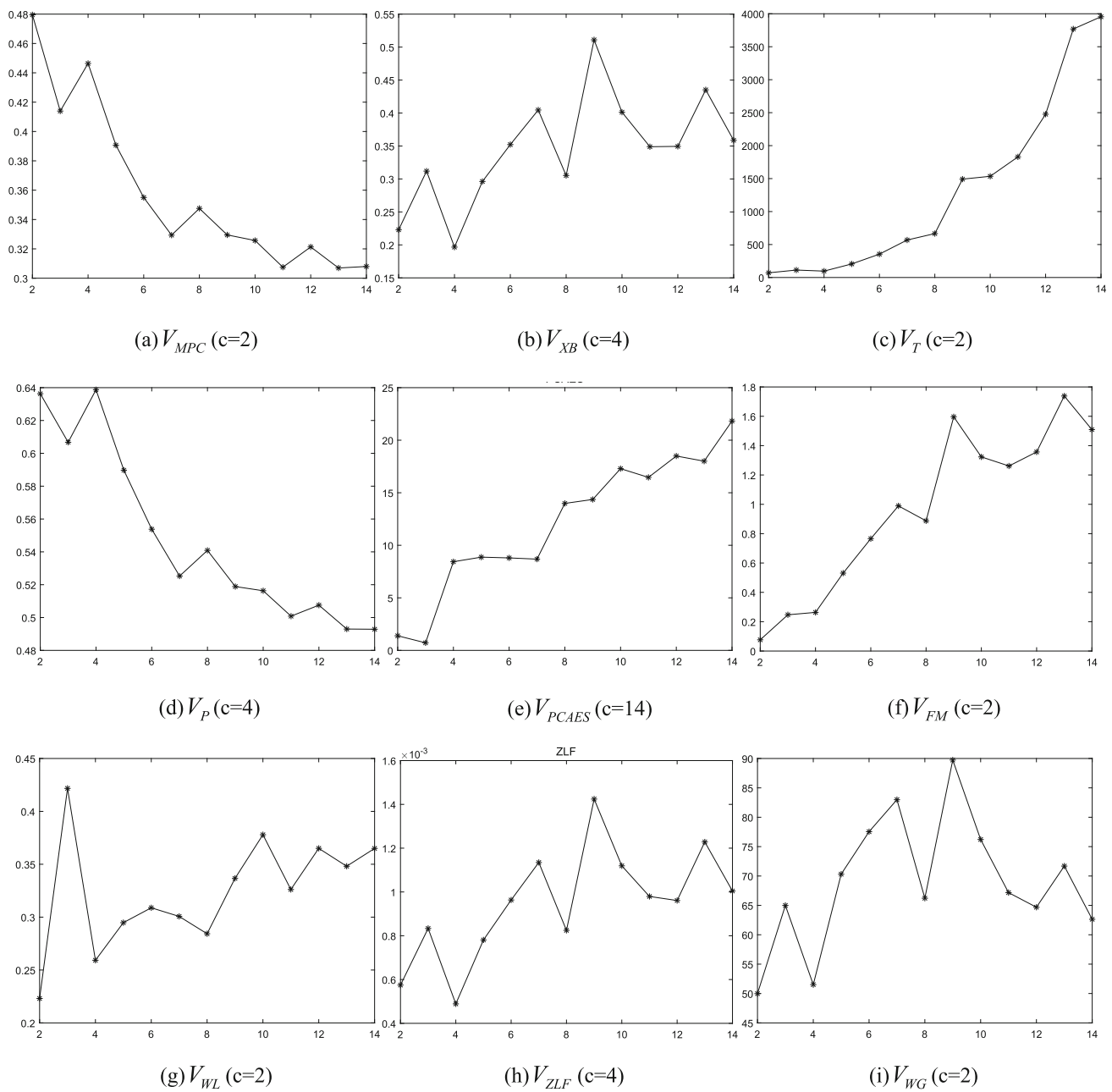
Fig. 13 Change trend of clustering validity functions under Haberman data set
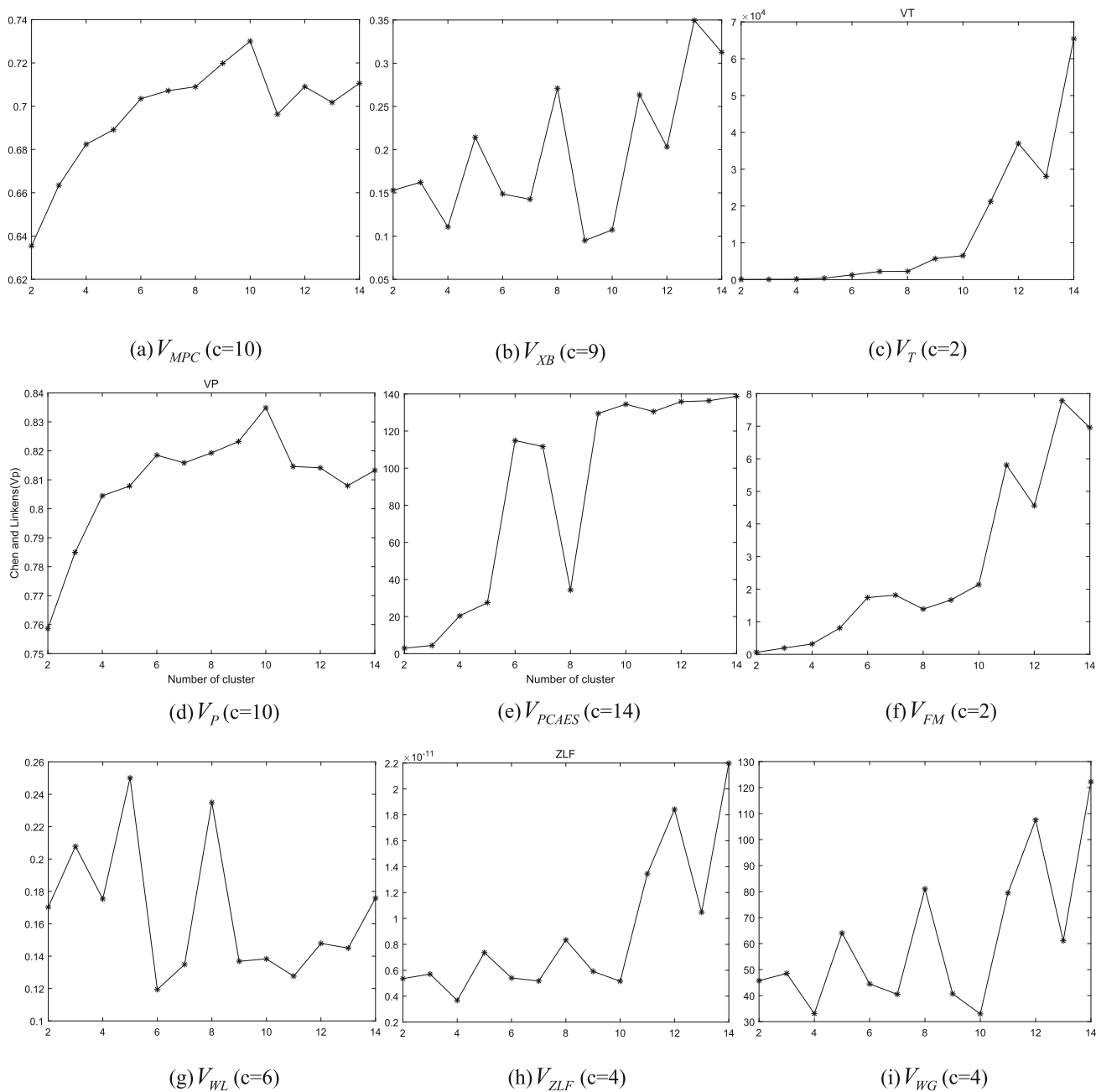
categories. In Fig. 13, V_{MPC}, V_T, V_{FM}, V_{WL} and V_{WG} can determine the best clustering result $c = 2$. Figure 14 shows that except for V_{PCAES}, other validity functions can classify correctly. Figure 15 shows that only V_{ZLF} and V_{WG} can obtain the best cluster number $c = 4$. Finally, in Fig. 16, still only V_{WG} can accurately classify 7 categories. Fig-

ures 15 and 16 are simulations for Hfcr and Segment data sets, and only V_{WG} can correctly classify them. This means that when the number of samples and complexity of the data set are relatively higher, the clustering effect of V_{WG} is significantly better than other validity functions. Finally, it can be seen from the experimental results in Fig. 17 that

(a) V_{MPC} (c=2)   (b) V_{XB} (c=2)   (c) V_T (c=2)

(d) V_P (c=2)   (e) V_{PCAES} (c=14)   (f) V_{FM} (c=2)

(g) V_{WL} (c=2)   (h) V_{ZLF} (c=2)   (i) V_{WG} (c=2)

Fig. 14 Change trend of clustering validity functions under HTUR2 data set

(a) V_{MPC} (c=10)

(b) V_{XB} (c=9)

(c) V_T (c=2)

(d) V_P (c=10)

(e) V_{PCAES} (c=14)

(f) V_{FM} (c=2)

(g) V_{WL} (c=6)

(h) V_{ZLF} (c=4)

(i) V_{WG} (c=4)

Fig. 15 Change trend of clustering validity functions under Hfcr data set

V_{XB}, V_T, V_{FM}, V_{WL}, V_{ZLF} and V_{WG} can be divided into two categories. From the experimental results in Fig. 10 to Fig. 17, it can be seen that only V_{WG} can find the best classification number of all UCI data sets, and when there are data sets with overlapping samples, noisy data, high dimensions and a large number of samples, the ideal number of clusters can still be found.
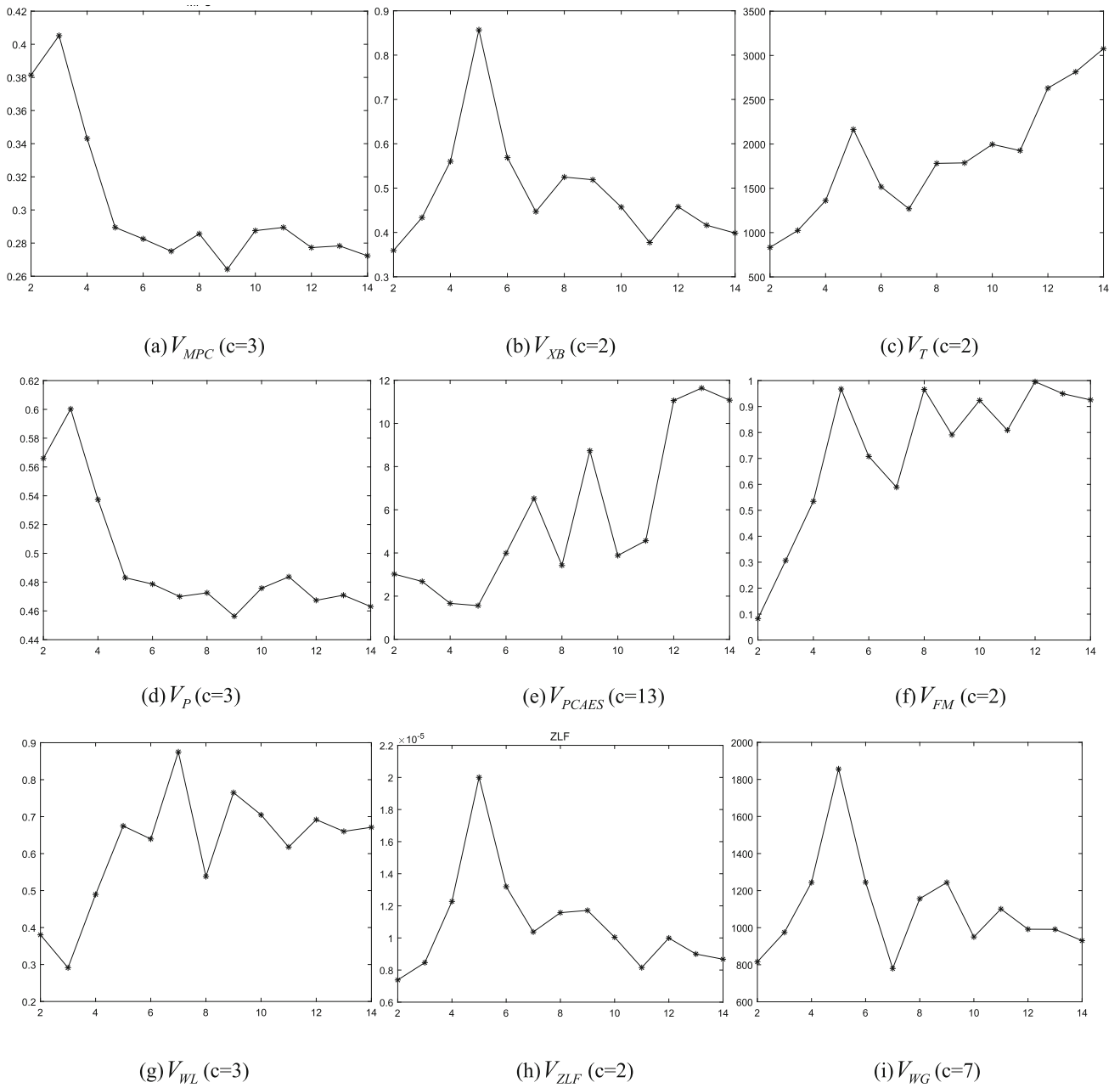
(a) V_{MPC} (c=3)

(b) V_{XB} (c=2)

(c) V_T (c=2)

(d) V_P (c=3)

(e) V_{PCAES} (c=13)

(f) V_{FM} (c=2)

(g) V_{WL} (c=3)

(h) V_{ZLF} (c=2)

(i) V_{WG} (c=7)

Fig. 16 Change trend of clustering validity functions under Segment data set

(a) V_{MPC} (c=10)

(b) V_{XB} (c=2)

(c) V_T (c=2)

(d) V_P (c=10)

(e) V_{PCAES} (c=3)

(f) V_{FM} (c=2)

(g) V_{WL} (c=2)

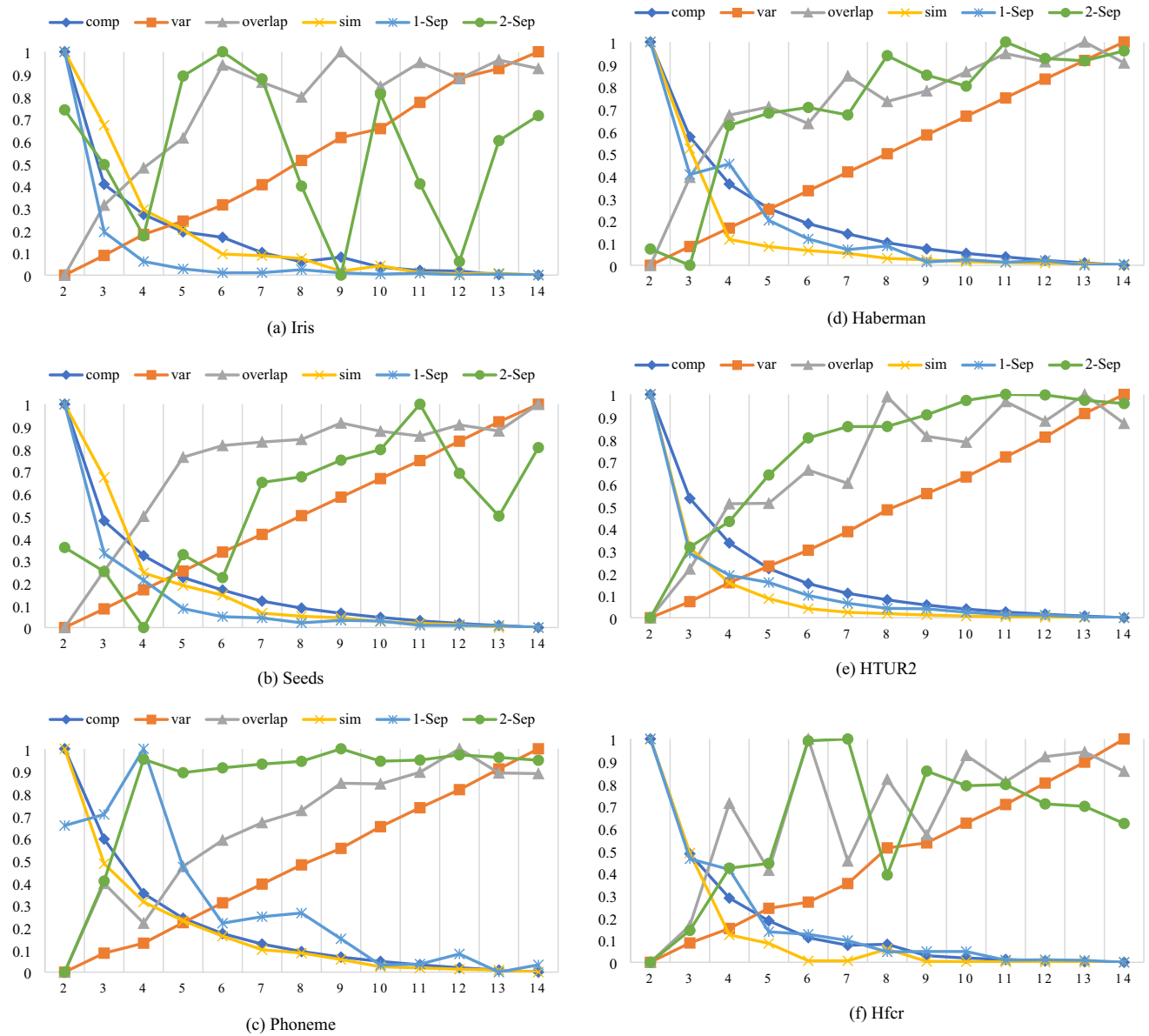(h) V_{ZLF} (c=2)

(i) V_{WG} (c=2)
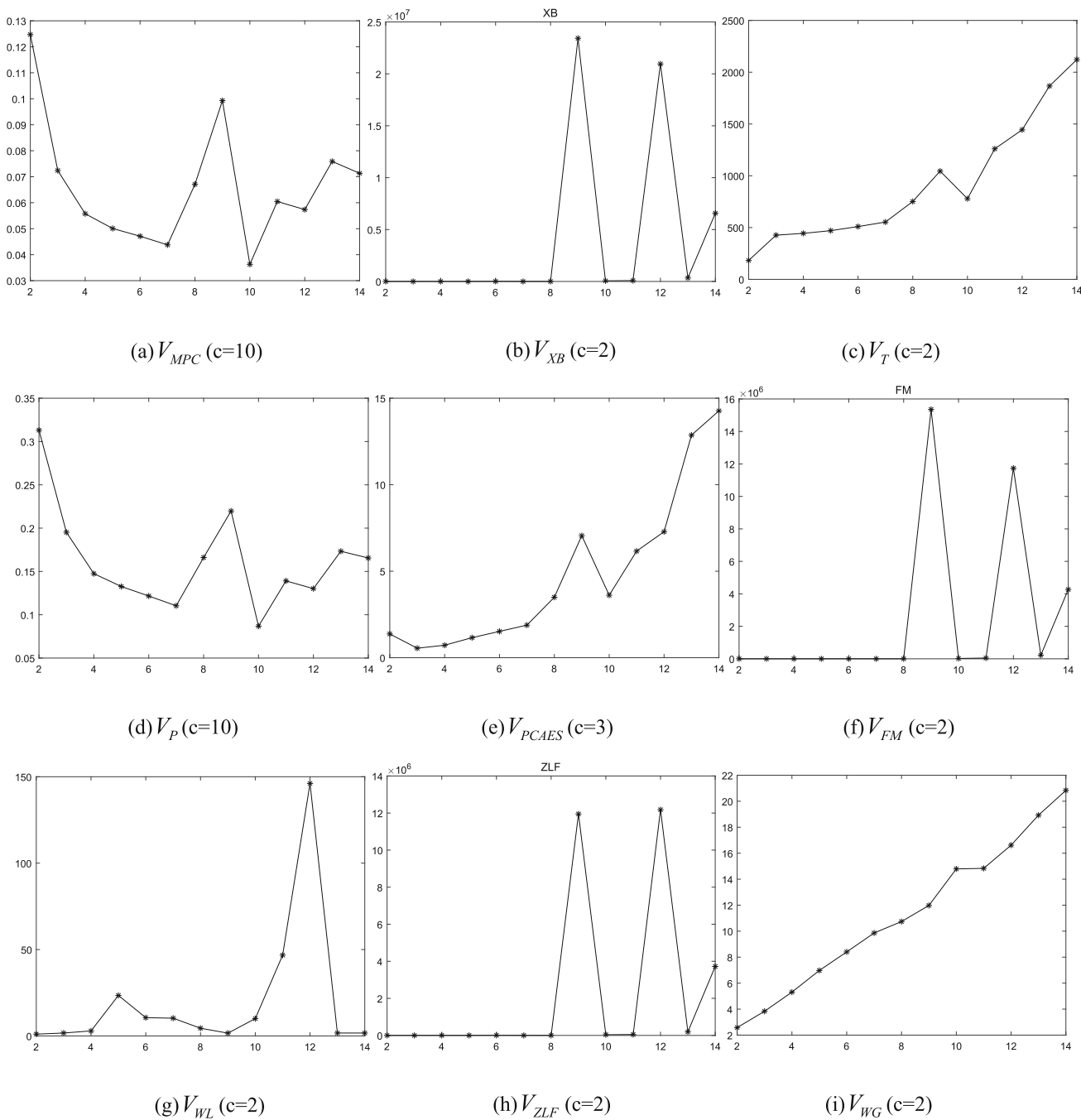
Fig. 17 Change trend of clustering validity functions under Heart data set

Fig. 18 Changing trend of each component in V_{WG} under UCI data sets

Similarly, in order to better observe the changing trend of each clustering performance evaluation component of V_{WG} when UCI data sets are adopted, six clustering performance evaluation components of V_{WG} are normalized and placed in the same coordinate system, which are shown in Fig. 18a-h. Then the function values of V_{MPC}, V_{XB}, V_T, V_P, V_{PCAES}, V_{FM}, V_{WL}, V_{ZLF} and V_{WG} on UCI data sets are put in the normalized coordinate system, which are shown in Fig. 19a–h. In this way, the clustering effect of V_{WG} and other clustering validity functions can be compared more intuitively. Finally, for different UCI data sets, the optimal
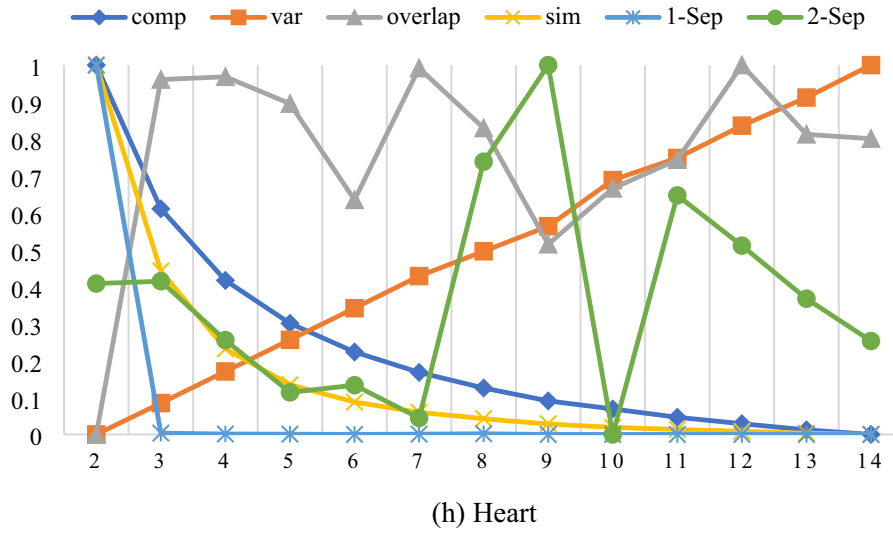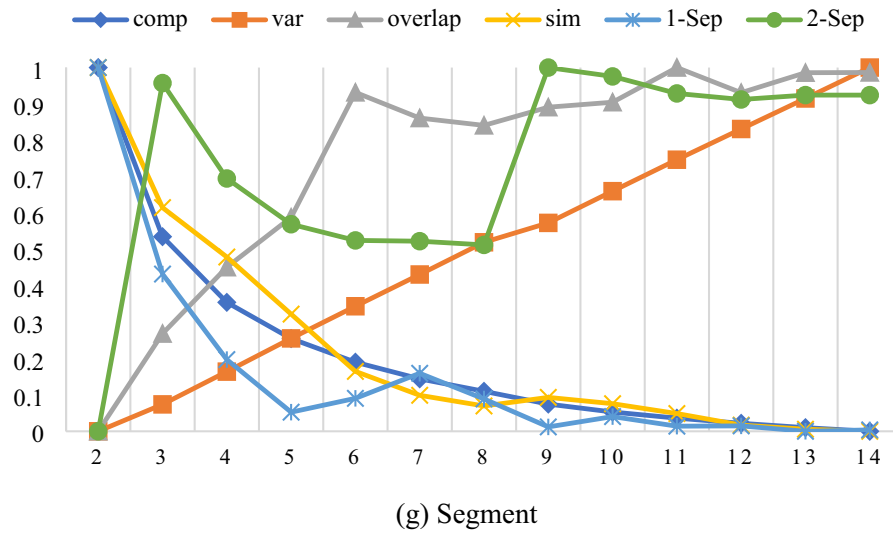
(g) Segment
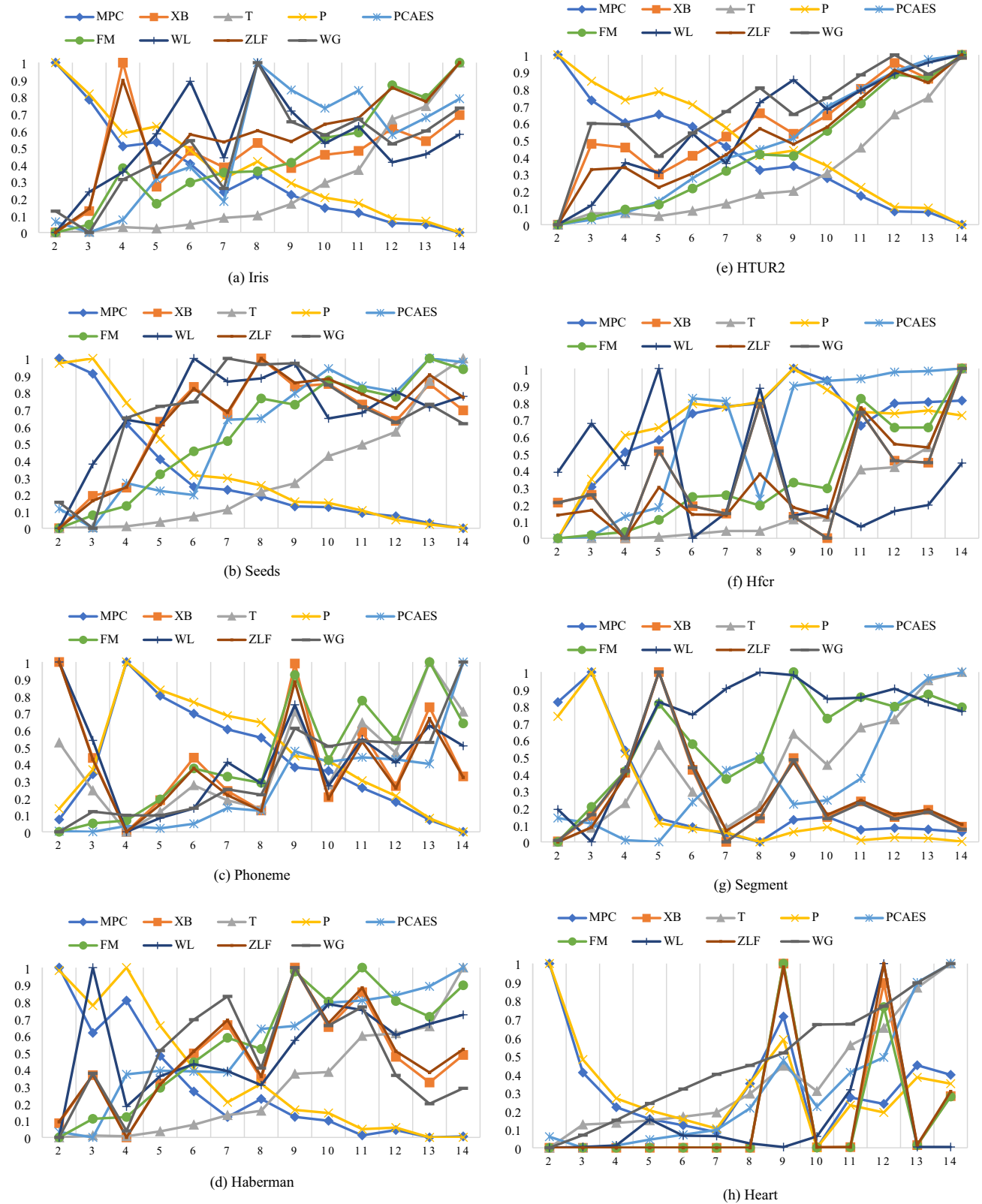

(h) Heart

Fig. 18 continued

Fig. 19 Changing trend of normalized clustering validity functions under UCI data sets

Table 4 Optimal number of clusters under different validity functions for UCI data sets

Data	Optimal c	V_{MPC}	V_{XB}	V_T	V_P	V_{PCAES}	V_{FM}	V_{WL}	V_{ZLF}	V_{WG}
Iris			2	2	2	14	2	2	2	**3**
	3	2								
Seeds			2	2	**3**	13	2	2	2	**3**
	3	2								
Phoneme			4	4	4	14	**2**	4	4	**2**
	2	4								
Haberman			4	**2**	4	14	**2**	**2**	4	**2**
	2	**2**								
HTUR2			**2**	**2**	**2**	14	**2**	**2**	**2**	**2**
	2	**2**								
Hfcr			9	2	10	14	2	6	**4**	**4**
	4	10								
Segment			2	2	3	13	2	3	2	**7**
	7	3								
Heart			**2**	**2**	10	3	**2**	**2**	**2**	**2**
	2	10								

The significance of bold indicate that clustering validity functions achieve the optimal cluster number

number of clusters of each cluster validity function are listed in Table 4.

5 Conclusions

In this paper, six evaluation components of clustering performance are defined based on compactness, variation, overlap, similarity and separation. After combining the six components, a new fuzzy clustering validity function V_{WG} is proposed, which is simulated on artificial data set and UCI data set. The experimental results show that V_{WG} can get accurate clustering number in Data_2_3(noise), Data_2_3(overlap), Data_2_5, Data_3_3, Data_3_6 and other manual data sets. V_{WG} for UCI data sets such as Iris, Seeds, Phoneme, Haberman, HTUR2, Hfcr, Segment and Heart can also be accurately divided. By comparison, it can be seen that the classification effect of V_{WG} is better than other traditional validity functions in data sets with noise data, overlapping data and high-dimensional data. Each clustering validity function is composed of various components, and different components play different roles in dividing a data set. Therefore, in the future work, we will analyze the rule of component synthesis of clustering validity function, and discuss the function and influence of different components through comparative experiments. At the same time, the idea of clustering validity function components can also be extended to the weighting and information integration among components to divide data sets.

Author Contributions GW participated in the data collection, analysis, algorithm simulation, and draft writing. J-SW participated in the concept, design, interpretation and commented on the manuscript. Hong-Yu Wang participated in the critical revision of this paper.

Declarations

References

1. Krista Rizman Žalik: Cluster validity index for estimation of fuzzy clusters of different sizes and densities. Pattern Recogn. **43**(10), 3374–3390 (2010)
2. Hartigan, J.A., Wong, M.A.: A K-Means Clustering Algorithm. J. R. Stat. Soc.: Ser. C: Appl. Stat. **28**(1), 100–108 (1979)
3. Lei, Y., Bezdek, J.C., Chan, J., Vinh, N.X., Romano, S., Bailey, J.: Extending information-theoretic validity indices for fuzzy clustering. IEEE Trans. Fuzzy Syst. **25**(4), 1013–1018 (2017)
4. Ruspini, E.H.: A new approach to clustering. Inf. Control **15**(1), 22–32 (1969)
5. Bezdek, J.C., Ehrlich, R., Full, W.: The fuzzy c-means clustering algorithm. Comput. Geosci. **10**(2–3), 191–203 (1984)
6. Fuzzy granular gravitational clustering algorithm for multivariate data: Mauricio A. Sanchez, Oscar Castillo, Juan R. Castro, Patricia Melin. Inf. Sci. **279**, 498–511 (2014)
7. Askari, S., Montazerin, N., Fazel Zarandi, M.H.: Generalized Possibilistic Fuzzy C-Means with novel cluster validity indices for clustering noisy data. Appl. Soft Comput. **53**, 262–283 (2017)
8. Rubio, E., Castillo, O., Valdez, F., Melin, P., Gonzalez, C.I., Martinez, G.: An extension of the fuzzy possibilistic clustering algorithm using Type-2 fuzzy logic techniques. Adv. Fuzzy Syst. **2017**, 23 (2017)
9. Farahani, F.V., Ahmadi, A., Zarandi, M.H.F.: Hybrid intelligent approach for diagnosis of the lung nodule from CT images using spatial kernelized fuzzy c-means and ensemble learning. Math. Comput. Simul. **149**, 48–68 (2018)
10. Yan, Bo., Na, Xu., Xu, L.P., Li, M.Q., Cheng, P.: An improved partitioning algorithm based on FCM algorithm for extended

target tracking in PHD filter. Digital Signal Processing **90**, 54–70 (2019)

11. Liang, H., Zou, J.: Rock image segmentation of improved semi-supervised SVM–FCM algorithm based on chaos. Circuits Syst Signal Process **39**, 571–585 (2020)

12. Bezdek, J.C., Moshtaghi, M., Runkler, T., Leckie, C.: The generalized c index for internal fuzzy cluster validity. IEEE Trans. Fuzzy Syst. **24**(6), 1500–1512 (2016)

13. Bezdek, J.C., Pal, N.R.: Some new indexes of cluster validity. IEEE Trans. Syst. Man Cybern. B Cybern. **28**(3), 301–315 (1998)

14. Simovici, D.A., Jaroszewicz, S.: An axiomatization of partition entropy. IEEE Trans. Inf. Theory **48**(7), 2138–2142 (2002)

15. Silva, L., Moura, R., Canuto, A.M.P., Santiago, R.H.N., Bedregal, B.: An Interval-based framework for fuzzy clustering applications. IEEE Trans. Fuzzy Syst. **23**(6), 2174–2187 (2015)

16. Fan, L., Xie, W.: Distance measure and induced fuzzy entropy. Fuzzy Sets Syst. **104**(2), 305–314 (1999)

17. Gaiyun, Gong, Xinbo, Gao (2004) Cluster validity function based on the partition fuzzy degree. Pattern Recognition and Artificial Intelligence, 412–416

18. Liu, Y., Zhang, X., Chen, J., Chao, H.: A Validity Index for Fuzzy Clustering Based on Bipartite Modularity. Journal of Electrical and Computer Engineering **2019**, 9 (2019)

19. J. Chen and D. Pi.(2013) A Cluster Validity Index for Fuzzy Clustering Based on Non-distance. International Conference on Computational and Information Sciences, 880–883

20. Joopudi, S., Rathi, S.S., Narasimhan, S., Rengaswamy, R.: A new cluster validity index for fuzzy clustering. IFAC Proceedings Volumes **46**(32), 325–330 (2013)

21. Zhang, D., Ji, M., Yang, J., Zhang, Y., Xie, F.: A novel cluster validity index for fuzzy clustering based on bipartite modularity. Fuzzy Sets Syst. **253**, 122–137 (2014)

22. XIE, Xuanli Lisa, BENI, Gerardo (1991) A validity measure for fuzzy clustering. IEEE Transactions on pattern analysis and machine intelligence, 841–847

23. Bensaid, A.M., et al.: Validity-guided (re)clustering with applications to image segmentation. IEEE Trans. Fuzzy Syst. **4**(2), 112–123 (1996)

24. KWON, Soon H.: Cluster validity index for fuzzy clustering. Electron. Lett. **34**, 2176–2177 (1998)

25. Kuo-Lung, Wu., Yang, M.-S.: A cluster validity index for fuzzy clustering. Pattern Recogn. Lett. **26**(9), 1275–1291 (2005)

26. Zhu, L.F., Wang, J.S., Wang, H.Y.: A novel clustering validity function of fcm clustering algorithm. IEEE Access **7**, 152289–152315 (2019)

27. Ouchicha, C., Ammor, O., Meknassi, M.: A new validity index in overlapping clusters for medical images. Control Comp, Sci. **54**, 238–248 (2020)

28. Liu, Y., Jiang, Y., Tao Hou, Fu.: A new robust fuzzy clustering validity index for imbalanced data sets. Inf. Sci. **547**, 579–591 (2021)

29. Wang, H.Y., Wang, J.S., Zhu, L.F.: A new validity function of FCM clustering algorithm based on the intra-class compactness and inter-class separation. Journal of Intelligent & Fuzzy Systems **40**(6), 12411–12432 (2021)

30. Wang, H.Y., Wang, J.S., Wang, G.: Combination Evaluation method of fuzzy C-mean clustering validity based on hybrid weighted strategy. IEEE Access **9**, 27239–27261 (2021)

31. Tasdemir, K., Merenyi, E.: A validity index for prototype-based clustering of data sets with complex cluster structures. IEEE Trans. Syst. Man Cybern. B Cybern. **41**(4), 1039–1053 (2011)

32. Yuangang Tang, Fuchun Sun and Zengqi Sun (2005) Improved validation index for fuzzy clustering. Proceedings of the 2005, American Control Conference, 1120–1125

33. Min-You Chen, D.A., Linkens,: Rule-base self-generation and simplification for data-driven fuzzy models. Fuzzy Sets Syst. **142**(2), 243–265 (2004)

34. Wu, C., Ouyang, C., Chen, L., Lu, L.: A new fuzzy clustering validity index with a median factor for centroid-based clustering. IEEE Trans. Fuzzy Syst. **23**(3), 701–718 (2004)

35. Meng, L., Chunchun, Hu.: Cluster validity index based on measure of fuzzy partition [J]. Comput. Eng. **33**(11), 15–17 (2007)

36. Pakhira, M.K., Bandyopadhyay, S., Maulik, U.: A study of some fuzzy cluster validity indices genetic clustering and application to pixel classification. Fuzzy Sets Syst. **155**(2), 191–214 (2005)

37. Zhang, Y., Wang, W., Zhang, X., Li, Yi.: A cluster validity index for fuzzy clustering. Inf. Sci. **178**(4), 1205–1218 (2008)

38. Rezaee, B.: A cluster validity index for fuzzy clustering. Fuzzy Sets Syst. **161**(23), 3014–3025 (2010)
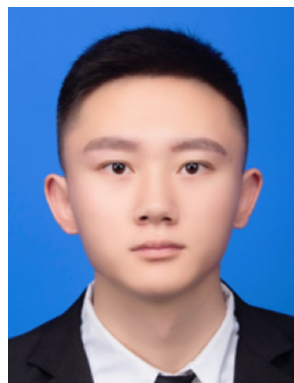
Guan Wang received the B. Eng. degree in Automation fromUniversity of Science and Technology Liaoning, China in 2020. He is currently pursuing the M. Eng. degree with the School of Control Science and Engineering, University of Science and Technology Liaoning, Anshan, China. His research interests include machine learning and intelligent optimization algorithm.

Jie-Sheng Wang received his B. Sc. And M. Sc. degrees in Control Science and Engineering from University of Science and Technology Liaoning, China in 1999 and 2002, respectively, and his Ph. D. degree in Control Science and Engineering from Dalian University of Technology, China in 2006. He is currently a professor and Doctor's Supervisor in School of Electronic and Information Engineering, University of Science and Technology Liaoning. His main research interest is modeling of complex industry process, intelligent control and Computer integrated manufacturing.

Hong-Yu Wang received his B. Eng. degree in Measuring Technology and Instrument from University of Science and Technology Liaoning, China in 2019. He is currently pursuing the M. Eng. degree with the School of Electronic and Information Engineering, University of Science and Technology Liaoning, Anshan, China. His main research interest is meta-heuristic algorithms and machine learning.