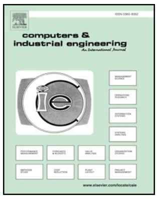
Computers & Industrial Engineering 127 (2019) 687–697

Contents lists available at ScienceDirect

Computers & Industrial Engineering

journal homepage: www.elsevier.com/locate/caie



A novel approach for fuzzy clustering based on neutrosophic association matrix

Hoang Viet Long[a,b], Mumtaz Ali[c], Le Hoang Son[d,*], Mohsin Khan[e], Doan Ngoc Tu[f]

[a] *Division of Computational Mathematics and Engineering, Institute for Computational Science, Ton Duc Thang University, Ho Chi Minh City, Vietnam*
[b] *Faculty of Mathematics and Statistics, Ton Duc Thang University, Ho Chi Minh City, Vietnam*
[c] *University of Southern Queensland, 4300 QLD, Australia*
[d] *VNU Information Technology Institute, Vietnam National University, Hanoi, Vietnam*
[e] *Abdul-Wali Khan University, Mardan 23200, Pakistan*
[f] *People's Police University of Technology and Logistics, Vietnam*

ARTICLE INFO

Keywords:
Fuzzy clustering
Neutrosophic set
Association matrix
Lambda-cutting matrix
Clustering quality

ABSTRACT

This paper proposes a fuzzy clustering algorithm through neutrosophic association matrix. In the first step, data are fuzzified into neutrosophic sets to create neutrosophic association matrix. By deriving a finite sequence of neutrosophic association matrices, the neutrosophic equivalence matrix is generated. Finally, the lambda-cutting is performed over the neutrosophic equivalence matrix to derive the final lambda-cutting matrix which is used to determine the clusters. Experimental results on several benchmark datasets using different clustering criteria show the advantage of the proposed clustering over the existing algorithms.

1. Introduction

In practice, data are often uncertain, inconsistency and uncompleted. To handle this problem, fuzzy set was proposed by Zadeh (1965) in which uncertainty is modeled as an elemental dependence of a set. Fuzzy sets have showed meaningful applications in many fields of study (Nguyen, Son, Ashour, & Dey, 2018; Ye & Du, 2017). One of the most essential utilization regarding the fuzzy set is the representation of information such as “non-membership” and “hesitancy”. For example, when diagnosing a patient, the doctor often concludes the patient's illness rate corresponds to the disease rather than indicating a complete or unspecified illness. There are several extensions of traditional fuzzy set have been proposed such as intuitionistic fuzzy sets (Atanassov, 1986) and neutrosophic fuzzy set (Smarandache, 1998). Neutrosophic set is the generalization of fuzzy set, intuitionistic fuzzy set and others. Neutrosophic set has been studied and applied in various fields such as the medical diagnosis (Mondaland and Pramanik, 2015), decision support systems (Pramanik and Chackrabarti, 2013), robots (Smarandache and Vladareanu, 2014), social and educational information analyzes, etc.

Clustering is an important concept along with fuzzy set theory. Several clustering algorithms based on fuzzy set have been proposed such as: Fuzzy C-Means (FCM) (Bezdek, Ehrlich, & Full, 1984), the methods proposed by Ye and Fu (2016), Ye and Fu (2016), Ye and Smarandache (2016), Ye and Zhang (2014), Ye (2014, 2016, 2017, 2018). Recently, neutrosophic association matrix usually is utilized as a tool in many fuzzy clustering algorithms. For the fuzzy clustering algorithm based on neutrosophic association matrix, the most important step is to evaluating the similarities in order to divide the elements into clusters. Ye and Smarandache (2016) proposed three types of measures including Jaccard, Dice and Cosine which then be used in multi-criteria decision making with simple neutrosophic dataset. In Ye (2014) and Ye and Zhang (2014); Ye continued to propose new neutrosophic fuzzy modification methods for decision-makers by combining above similar measures. On the other hand, Ma, Wang, Wang, and Wu (2015) investigate the similar measures of tangential function for medical applications. Other studies on neutrosophicfuzzy clustering algorithms can be found in Kuo, Potti, and Zulvia (2018), Wu, Wu, Zhou, Chen, and Guan (2017), Ye and Fu (2016), Ye and Zhang (2014), Ye (2016).

This article proposes a new fuzzy clustering using neutrosophic association matrix. The first step of the algorithm is to construct a neutrosophic association matrix from the data in the dataset. After that, a neutrosophic equivalent matrix is constructed from neutrosophic association matrix. Finally, the lambda-cutting matrix is built based on neutrosophic equivalent matrix by the lambda-cutting step. The result clusters are defined based on the lambda-cutting matrix.

Section 2 presents some background information and proposes a new neutrosophic clustering method though detailed analysis. Section 3

[*] Corresponding author at: 144 Xuan Thuy, Cau Giay, Hanoi, Vietnam.
E-mail addresses: hoangvietlong@tdtu.edu.vn (H.V. Long), Mumtaz.Ali@usq.edu.au (M. Ali), sonlh@vnu.edu.vn (L.H. Son), tudn@truongt36.edu.vn (D.N. Tu).

https://doi.org/10.1016/j.cie.2018.11.007
Received 8 May 2018; Received in revised form 17 September 2018; Accepted 5 November 2018
Available online 08 November 2018

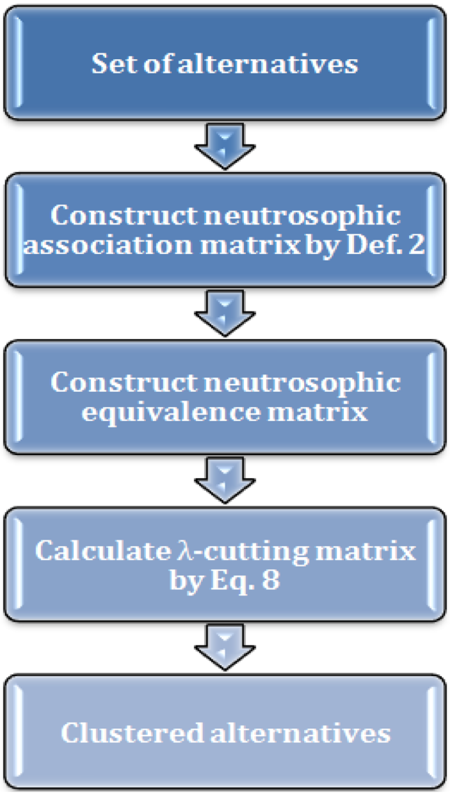

Set of alternatives
Construct neutrosophic association matrix by Def. 2
Construct neutrosophic equivalence matrix
Calculate λ-cutting matrix by Eq. 8
Clustered alternatives

Fig. 1. Flowchart of the proposed clustering algorithm.

shows the experimental result of proposed algorithm in comparison with other relevant methods on real data sets. Conclusions are in the Section 4.

2. The proposed clustering algorithm

2.1. Background of neutrosophic set

Let $\varepsilon \geq 0$ be a infinitesimal number (Smarandache, 1998), i.e., for all positive integers one has $\varepsilon < \frac{1}{n}$. Let $1^+ = 1 + \varepsilon$, where "1" and "ε" are its standard and non-standard parts respectively. Similarly, $(0^-) = 0 - \varepsilon$, and $]0^-, \ 1^+[$ is a non-standard unit interval.

A neutrosophic set Ain the universe $X$is characterized by a truth, indeterminacy, and falsehood membership functions $< T_A(x), I_A(x), F_A(x) >$ such that $T_A(x), I_A(x), F_A(x)\colon X \to]0^-, 1^+[$ and $0 \leq T_A(x) + I_A(x) + F_A(x) \leq 3$ (Smarandache, 1998).

Suppose that $A_1 = \{\langle x; T_1(x); I_1(x); F_1(x)\rangle | x \in X\}$ and $A_2 = \{\langle x; T_2(x); I_2(x); F_2(x)\rangle | x \in X\}$ be two neutrosophic sets. We recall some base relationship between neutrosophic sets (Smarandache, 1998):

$$A_1 \subseteq A_2 \text{iff } T_1(x) \leq T_2(x); I_1(x) \leq I_2(x); F_2(x) \leq F_1(x),$$
$$A_1^c = \{\langle x; F_1(x); I_1(x); T_1(x)\rangle | x \in X\},$$
$$A_1 \cap A_2 = \{\langle x; \min\{T_1(x); T_2(x)\}; \max\{I_1(x); I_2(x)\}; \max\{F_1(x); F_2(x)\}\rangle | x \in X\},$$
$$A_1 \cup A_2 = \{\langle x; \max\{T_1(x); T_2(x)\}; \min\{I_1(x); I_2(x)\}; \min\{F_1(x); F_2(x)\}\rangle | x \in X\}.$$

2.2. Construction of neutrosophic association matrices

Denote $N(X)$ by the set of all neutrosophic set.

Definition 1. Mapping $m\colon N(X) \times N(X) \to [0, 1]$ is defined a association coefficient function if it satisfies following properties for all $(A, B) \in N(X)$

(1) $0 \leq m(A, B) \leq 1$;

Table 1
The descriptions of experimental EPPO datasets.

Dataset	No. elements	No. attributes
eppo_standard_pp1	1452	289
eppo_standard_pm8	167	3
eppo_standard_pm4	555	35

Table 2
The descriptions of experimental UCI datasets.

Dataset	No. elements	No. attributes
Machine	209	10
Ecoli	336	9
Pima-indians-diabetes	768	9
Student	395	33
Transfusion	748	5
Voting-records	17	17
Climate model	540	22
Adult	806	14
Breast-cancer-wisconsin	699	11
Seed	210	8

(2) $m(A, B) = 1$ iff $A = B$;
(3) $m(A, B) = m(B, A) \ \forall \ m(B, A)$.

From this definition, we proposed the following notions and theorems which will be used in the main clustering algorithm later.

Definition 2. Let $B_j (j = 1, 2, \cdots, n)$ be neutrosophic sets. $M = (m_{ij})_{n \times n}$is called an **neutrosophic association matrix**, where $m_{ij} = m(B_i, B_j)$ is the association coefficients of B_i and B_j.

Definition 3. Let $M = (m_{ij})_{n \times n}$ be an association matrix. If $M^2 = M * M = (\bar{m}_{ij})_{n \times n}$, then M^2 is a **composition matrix** of M

$$\bar{m}_{ij} = \max_p\{\min\{m_{ip}, m_{pj}\}\}, \quad i, j = 1, 2, \cdots, n. \tag{1}$$

Theorem 1. *If $M = (m_{ij})_{n \times n}$ is an association matrix then M^2 is also an association matrix.*

Proof.

(a) For any $i, j = 1, 2, \cdots, n$, we have $0 \leq m_{ij} \leq 1$.

Thus,

$$0 \leq \bar{m}_{ij} = \max_p\{\min\{m_{ip}, m_{pj}\}\} \leq 1 \text{ for all } i, j = 1, 2, \cdots, n. \tag{2}$$

(b) Since $m_{ij} = 1$ *if and only if* $B_i = B_j$, $i, j = 1, 2, \cdots, n$, it yields

$$\bar{m}_{ij} = \max_p\{\min\{m_{ip}, m_{pj}\}\} = 1 \tag{3}$$
$$\text{if and only if } B_i = B_p = B_j \text{for some } p = 1, 2, \cdots, n.$$

(c) Since $m_{ij} = m_{ji}$ for all $i, j = 1, 2, \cdots, n$, we get

$$\bar{m}_{ij} = \max_p\{\min\{m_{ip}, m_{pj}\}\} = \max_p\{\min\{m_{pi}, m_{jp}\}\}$$
$$= \max_p\{\min\{m_{jp}, m_{pi}\}\} = \bar{m}_{ji}, \ i, j = 1, 2, \cdots, n \ \square \tag{4}$$

Theorem 2. *If $M = (m_{ij})_{n \times n}$is an association matrix then for a positive integer p,*

$$M^{2p+1} = M^{2p} * M^1 \tag{5}$$

is also an association matrix.

Proof. Straightforward. □

Definition 4. If $M^2 \subseteq M$ i.e.,

$$\max_p\{\min\{m_{ip}, m_{pj}\}\} \leq m_{ij} \tag{6}$$

for all $i, j = 1, 2, \cdots, n$, then an association matrix $M = (m_{ij})_{n \times n}$ is called

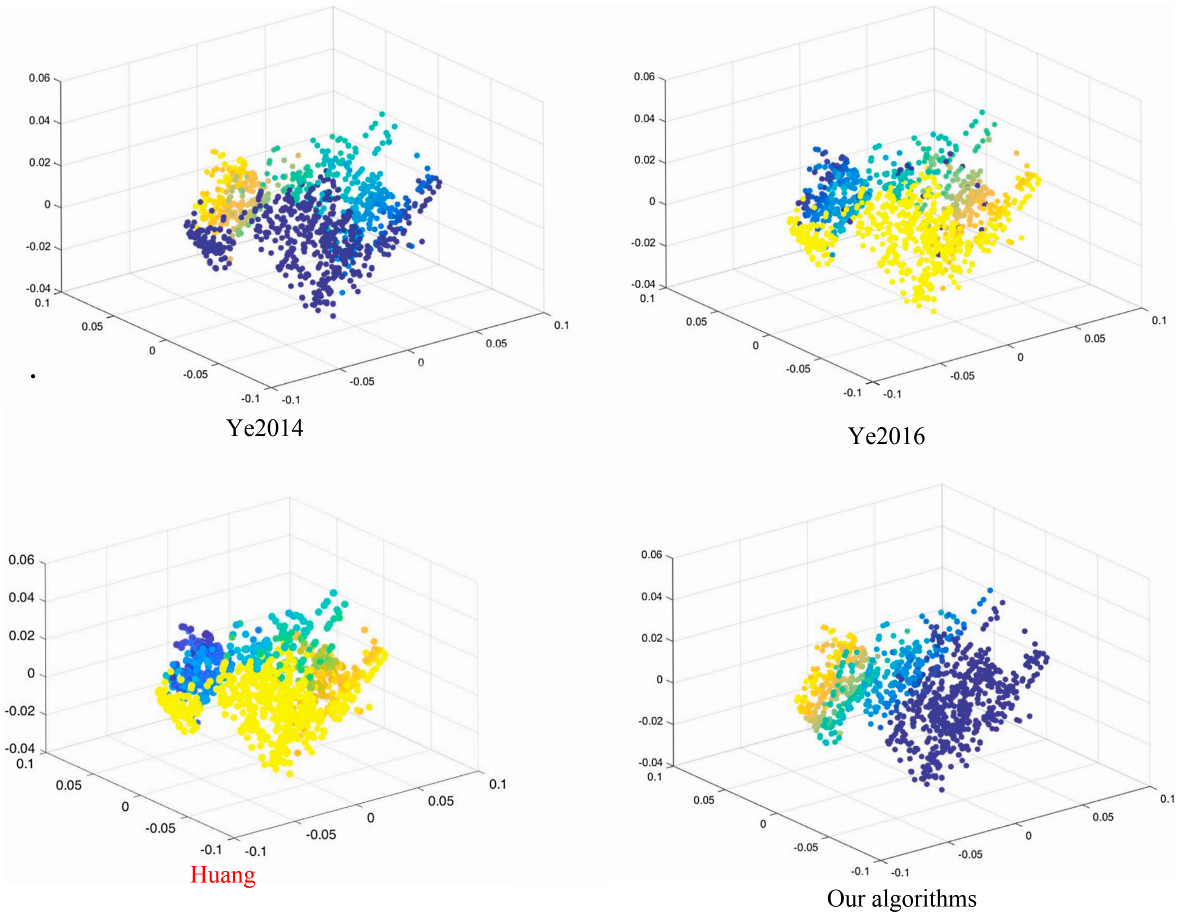


Fig. 2. Clustering result of 3 methods with eppo_standard_pp1 dataset.

an equivalent association matrix.

Theorem 3. *Let $M=(m_{ij})_{n\times n}$be an association matrix. After finite times of compositions:*

$$M\to M^2\to M^4\to \cdots\to M^{2p}\to \cdots \tag{7}$$

there exist p: $M^{2p}=M^{2(p+1)}$, and M^{2p} is an equivalent association matrix.

Definition 5. Let $M=(m_{ij})_{n\times n}$ be an equivalent association matrix. Then, $M_\lambda=(m_{ij}^{\lambda})_{n\times n}$ is called the **λ-cutting matrix** of M with $\lambda\in[0,1]$ being the confidence level.

$$m_{ij}^{\lambda}=\begin{cases}0 & if \quad m_{ij}<\lambda,\\ 1 & if \quad m_{ij}\geq\lambda,\end{cases} i,j=1,2,\cdots,n \tag{8}$$

2.3. Clustering algorithm based on association matrices of neutrosophic sets

Step 1: Let $U=\{u_1,u_2,\cdots,u_p\}$ be a universe of discourse, and $\alpha=(\alpha_1,\alpha_2,\cdots,\alpha_p)^T$ is the weight vector of $\alpha_l\,(l=1,2,\cdots,p)$, with $\alpha_l\in[0,1]$ for all $l=1,2,\cdots,p$, and $\sum_{l=1}^{p}\alpha_l=1$. Consider a collection of neutrosophic sets $B_j\,(j=1,2,\cdots,n)$, where

$$B_j=\{\langle y,T_{B_j}(y_l),I_{B_j}(y_l),F_{B_j}(y_l)\rangle|y_l\in U\},\ j=1,2,\cdots,n \tag{9}$$

$$\varphi_{B_j}(y_l)=3-T_{B_j}(y_l)-I_{B_j}(y_l)-F_{B_j}(y_l),$$

is the degree of uncertainty of y_l to B_j.

Step 2: Select a neutrosophic sets association measure, such as Eq. (10) below.

Let we noting that, by using well-known Cauchy-Schwarz inequalit

$$\sum_{i=1}^{p}a_ib_i\leq\sqrt{\left(\sum_{i=1}^{p}\alpha_i{a_i}^2\right)\left(\sum_{i=1}^{p}\alpha_i{b_i}^2\right)},$$

where $\sum_{i=1}^{p}\alpha_i=1$, we can show that $m(B_i,B_j)$ defined in eq. (10) satisfies Definition 1.

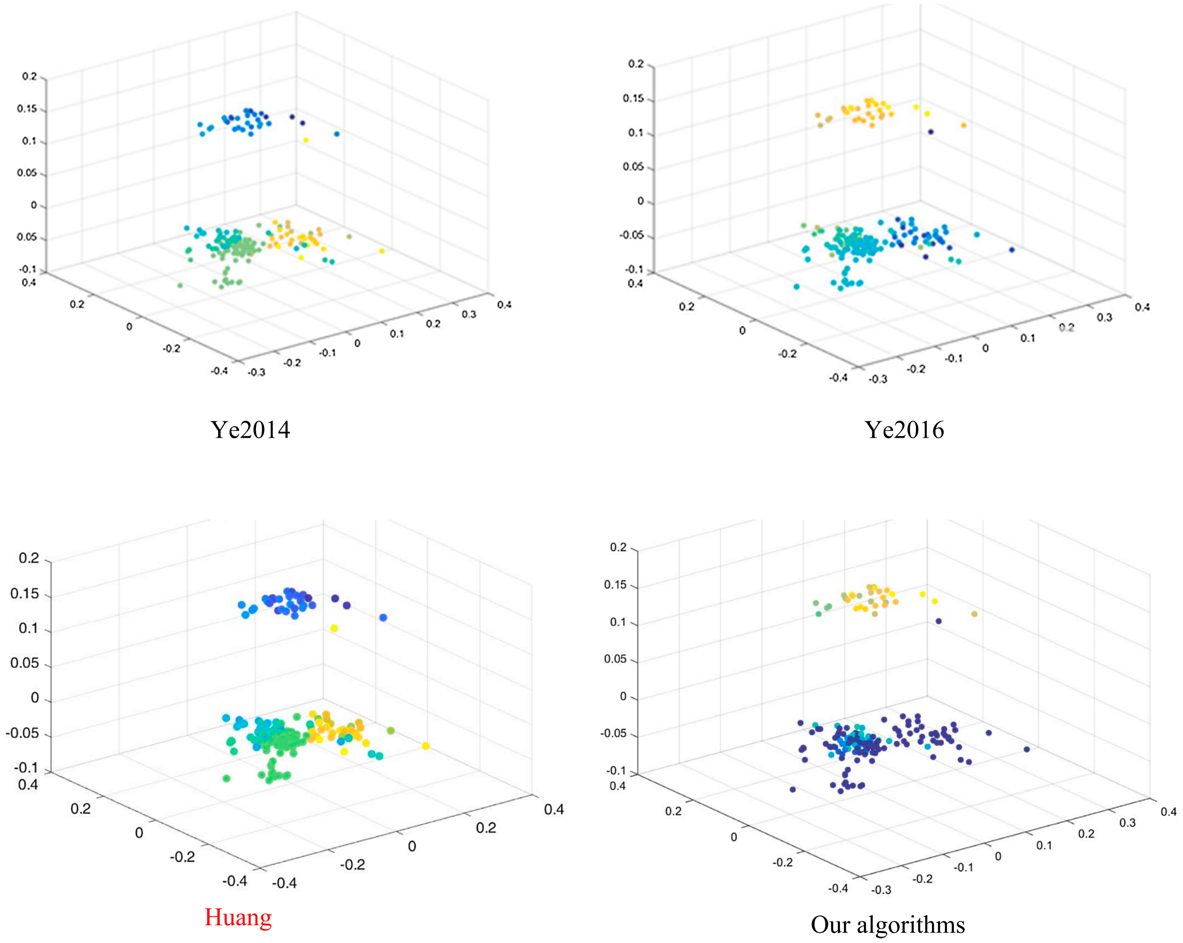
$$m(B_i,B_j)=\frac{\sum_{l=1}^{p}(T_{B_i}(y_l)^2T_{B_j}(y_l)^2+I_{B_i}(y_l)^2I_{B_j}(y_l)^2+F_{B_i}(y_l)^2F_{B_j}(y_l)^2)}{\max\Big(\sum_{l=1}^{p}\alpha_l({T^2}_{B_i}(y_l)+{I^2}_{B_i}(y_l)+{F^2}_{B_i}(y_l)+{\varphi^2}_{B_i}(y_l)),\sum_{l=1}^{p}\alpha_l({T^2}_{B_j}(y_l)+{I^2}_{B_j}(y_l)+{F^2}_{B_j}(y_l)+{\varphi^2}_{B_j}(y_l))\Big)} \tag{10}$$

Step 3: If $M=(m_{ij})_{p\times p}$ is an equivalent association matrix then build $M_\lambda=(m_{ij}^{\lambda})_{n\times n}$ using Eq. (8); otherwise derive an equivalent association matrix $\bar{M}$ by Eq. (7). Construct λ -cutting matrix $\bar{M}_\lambda=(\lambda\bar{m}_{ij})_{n\times n}$ of $\bar{M}$ by Eq. (8).

Step 4: If elements of the ith line in M_λ (or $\bar{M}_\lambda$) are the same as those of jth line then B_i and B_j are of the same type. By this principle, we can classify all these neutrosophic set $B_j\,(j=1,2,\cdots,n)$.

These steps of this clustering algorithm can be seen in the following Fig. 1.

By using the cutting matrix of the equivalent association matrix, the new algorithm classifies neutrosophic sets according to a given confidence level which is specified by elements of equivalent association matrices and actual situations.



Ye2014

Ye2016

Huang

Our algorithms

Fig. 3. Clustering result of 3 methods with eppo_standard_pm8 dataset.

3. Experiments

3.1. Experimental environments

The proposed algorithm has been implemented in addition to the methods of Ye (2014), Ye (2016) and Huang (2016) in Matlab 2015a programming language with a PC with CPU Intel(R) Core (TM) i5-2520 M@2.4 GHz, 4096 MB RAM, windows 7 Professional 64 bits.

In order to perform the evaluation, two kinds of datasets have been used. The first dataset is the set of EPPO standard dataset which is taken from EPPO Global Database. It provides a large dataset for variety types as agriculture, forestry and plan protection. Other 10 benchmark datasets (Machine, Ecoli, Pima-indians-diabetes, Student, Transfusion, Voting-records, Climate Model, Adult, Breast-cancer-wisconsin, Seed) have been taken from UCI dataset (UCI Machine Learning Datasets) (see Tables 1 and 2).

Experimental objectives: The quality of all clustering algorithms is evaluated by 3 indices namely DB, SSWC, IVF, VRC and BH.

(a) **Davies-Bouldin (DB)** (Davies and Bouldin, 1979):

Let x_i be an “n”-dimensional feature vector assigned to cluster C_i and $\bar{x}_i$ is the centroid of C_i. Denote $\bar{d}_l$, $\bar{d}_m$ by the average distances of clusters C_m and C_l,respectively and $d_{m,l}$ is the distance between them.

$$\bar{d}_l = \frac{1}{N_l}\sum_{x_i\in C_l} ||x_i - \bar{x}_l||;$$

$$d_{l,m} = ||\bar{x}_l - \bar{x}_m||.$$

If k is the number of clusters, then DB is called the Davies-Bouldin index with

$$DB = \frac{1}{k}\sum_{l=1}^{k} D_l \tag{11}$$

$$D_l = \max_{m\neq l}\{D_{l,m}\};$$

$$D_{l,m} = (\bar{d}_l + \bar{d}_m)/d_{m,l}$$

The lower value of DB criterion is better.

(b) **Simplified Silhouete Width Criterion (SSWC):**

Supposed that x_j is the point of cluster A and $a_{p,j}$ is the average distance of x_j to points in A, while $b_{p,j}$ is the minimum average distance from x_j to all other clusters.

Then the silhouette of x_j is defined by

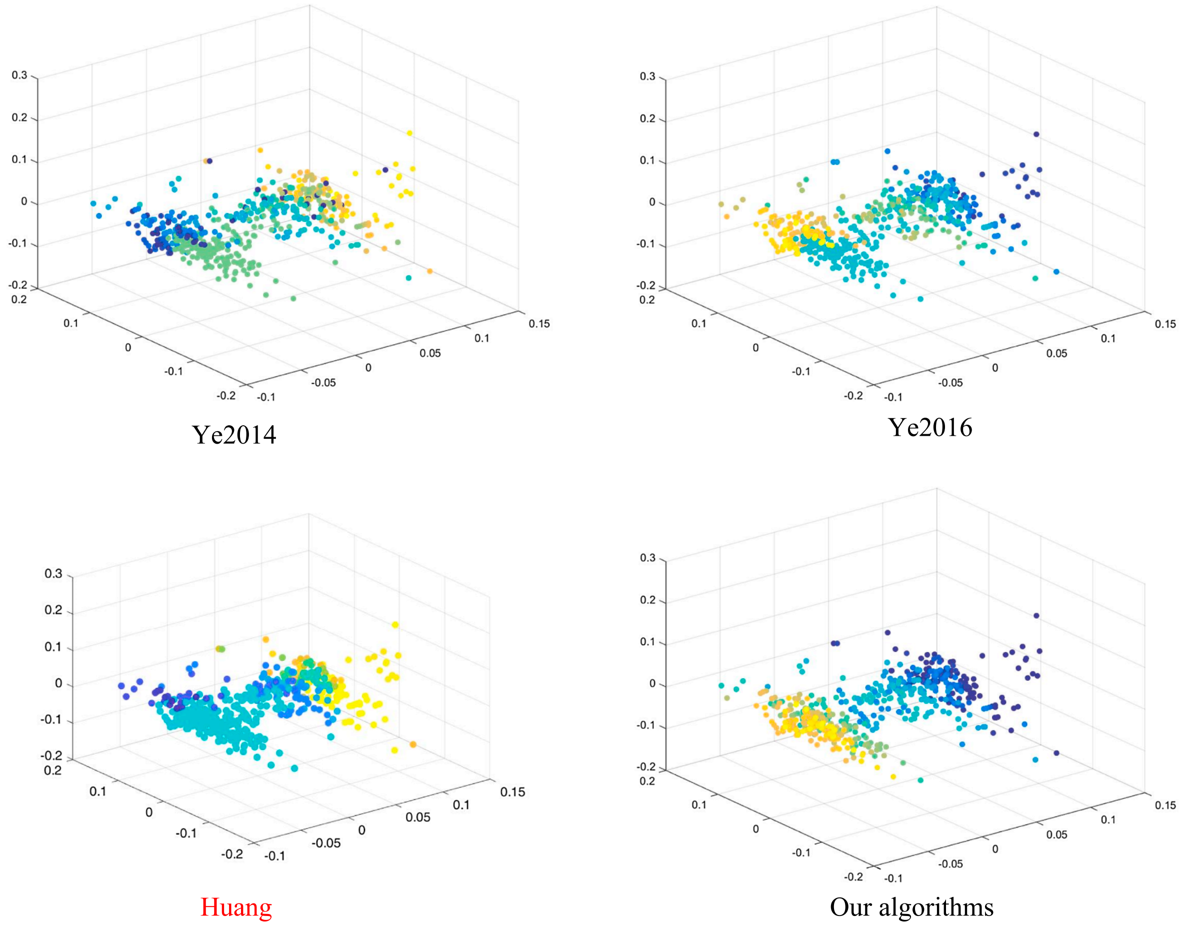
$$S_{xj} = \frac{b_{p,j} - a_{p,j}}{\max\{a_{p,j}, b_{p,j}\}}.$$

The Simplified Silhouete Width Criterion is

$$SSWC = \frac{1}{N}\sum_{j=1}^{N} S_{xj} \tag{12}$$

Using SSWC, the greater value shows more efficient algorithm.

(c) **IFV** (Atanassov, 1986):



Ye2014

Ye2016

Huang

Our algorithms

Fig. 4. Clustering result of 3 methods with eppo_standard_pm4 dataset.

$$IFV = \frac{1}{C}\sum_{j=1}^{C}\left\{\frac{1}{N}\sum_{k=1}^{N}u_{kj}^{2}\left[\log_2 C - \frac{1}{N}\sum_{k=1}^{N}\log_2 u_{kj}\right]^2\right\} \times \frac{SD_{max}}{\sigma_D} \tag{13}$$

where

$$SD_{max} = \max_{k\neq j}||V_k - V_j||^2,$$

$$\bar{\sigma}_D = \frac{1}{C}\sum_{j=1}^{C}\left(\frac{1}{N}\sum_{k=1}^{N}||X_k - V_j||^2\right)$$

Here, X_k is the element belonging to cluster k^{th} and V_k is the centroid of this cluster.

The maximal value of IFV indicates the better performance.

(d) **Calinski-Harabasz Criterion (VRC)** (Kaufman and Rouseeuw, 1991):

The Calinski-Harabasz criterion is called the variance ratio criterion (VRC). VRC is defined as

$$VRC_k = \frac{SS_B}{SS_W} \times \frac{(N-k)}{(k-1)} \tag{14}$$

where SS_B, SS_W are the overall between-cluster and within-cluster variance respectively, k and N are the number of clusters and observations. SS_B is defined as

$$SS_B = \sum_{i=1}^{k} n_i||m_i - m||^2$$

where k is the number of clusters, m_i is the centroid of cluster i, m is the overall mean of the data, and $||m_i - m||$ is the L^2 norm (Euclidean distance) between the two vectors. SS_W is defined as

$$SS_W = \sum_{i=1}^{k}\sum_{x\in c_i}||x - m_i||^2$$

where x is a data point, c_i is the ith cluster, m_i is the centroid of cluster i and $||x - m_i||$is Euclidean distance between the two vectors.

The maximal value of VRC show the better performance.

(e) **Ball-Hall criterion (BH)** (Atanassov, 1986):

The Ball-Hall criterion (BH) is the mean, through all the clusters, of their mean dispersion:

$$BH = \frac{1}{c}\sum_{i=1}^{c}\frac{1}{n_i}\sum_{j=1}^{n}u_{ij}\,||x_i - v_i||^2 \tag{15}$$

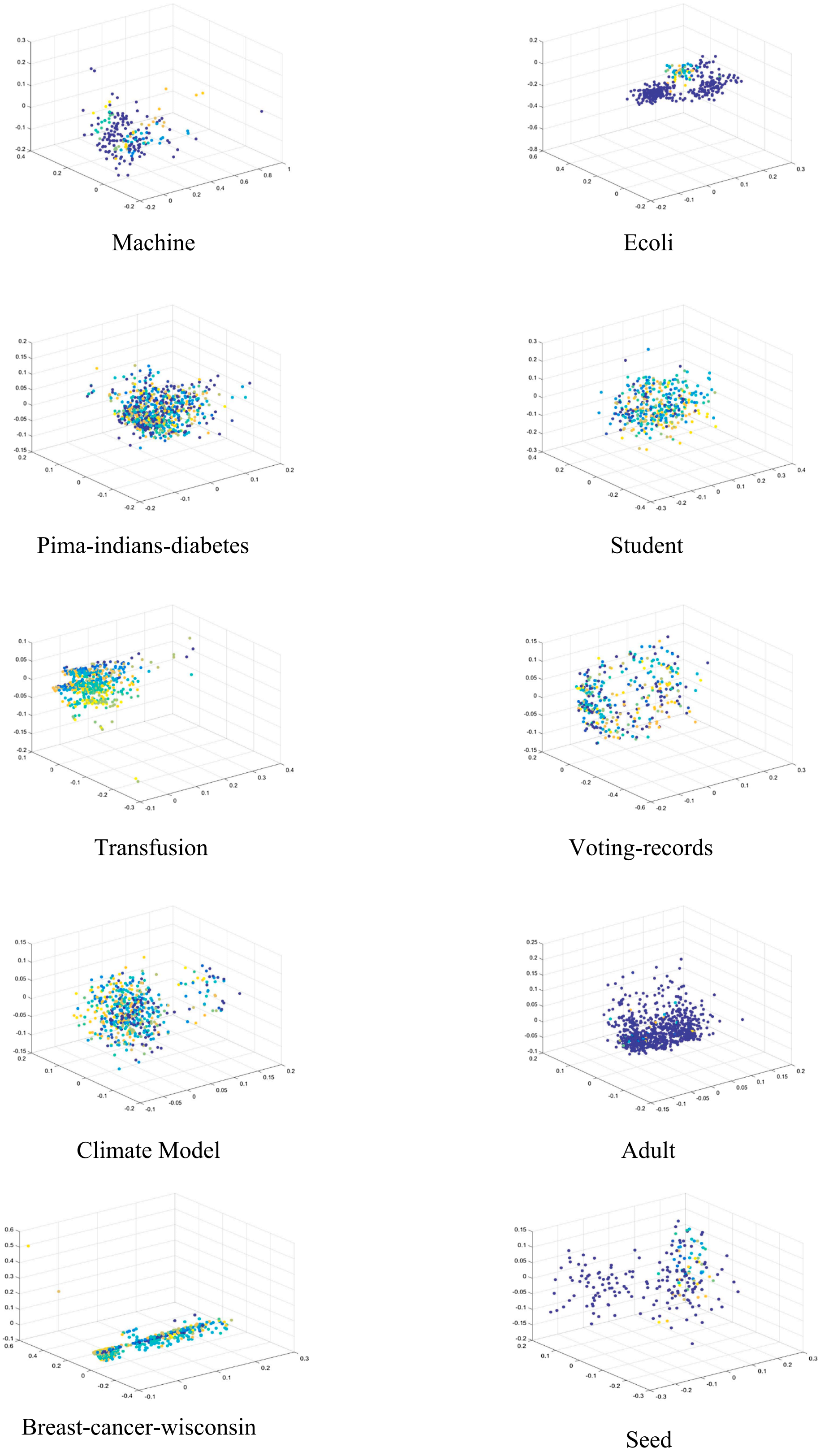
where n_i is the number of observations in the i-th cluster, and $u_{ij}$is the membership degree of $x_i$in the i-th cluster.

The maximal value of BH the better performance.

3.2. The comparison of performance

Figs. 2–4 show the result of clustering algorithms where a color represents a cluster. The number of clusters depends on each method and their configurations parameters. Here we choose the number of clusters in each algorithm by approximating each other. Through each figure and its sub-figure it is possible to see that the proposed algorithm expresses clusters more clearly than other algorithms.

Figs. 5–8 show the result of clustering algorithms for each UCI



Machine

Ecoli

Pima-indians-diabetes

Student

Transfusion

Voting-records

Climate Model

Adult

Breast-cancer-wisconsin

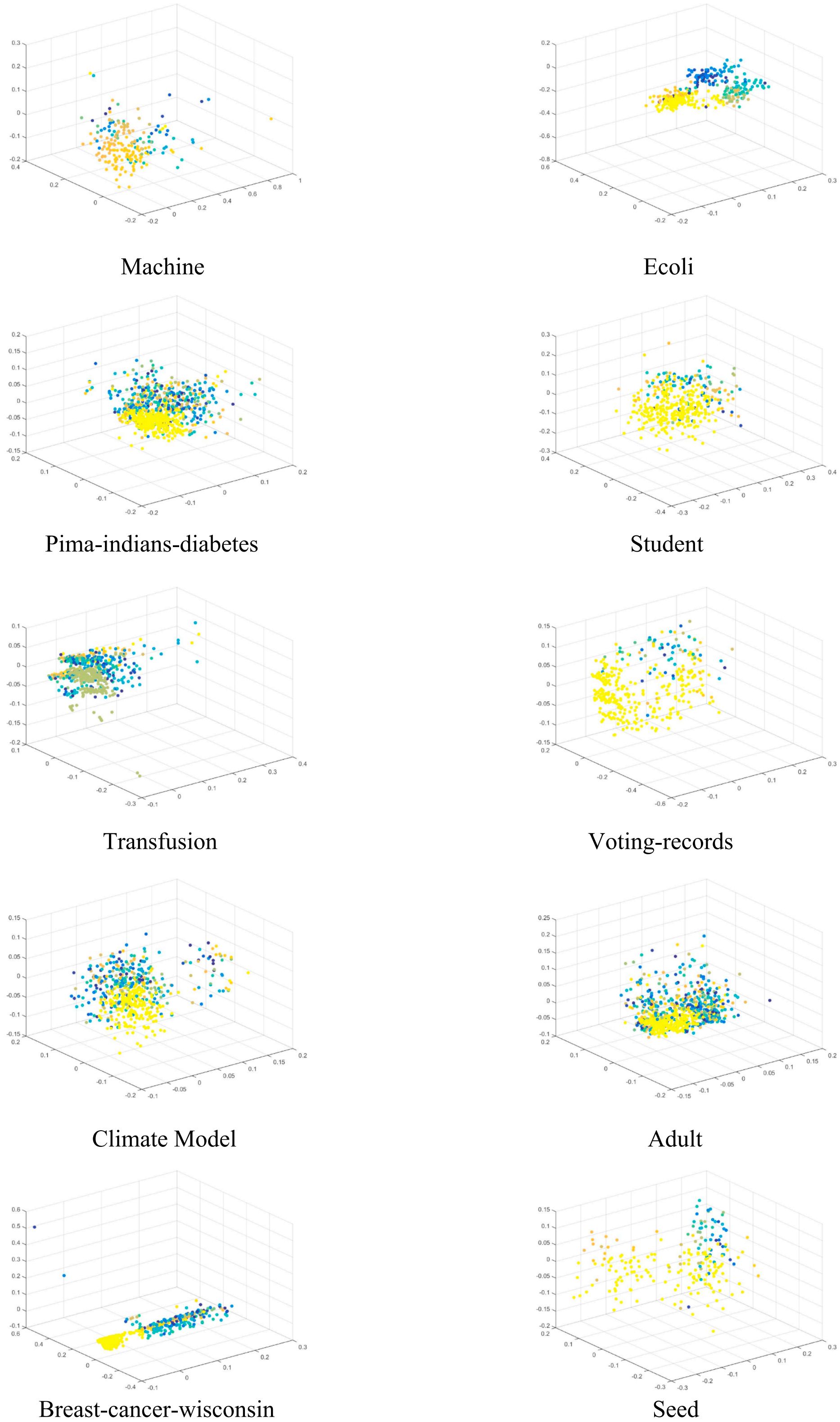
Seed

Fig. 5. Clustering result of our algorithm with 10 UCI dataset.

dataset. Tables 3 and 4 present some comparative results of proposed method with existing works on EPPO and UCI dataset.

Tables 5 and 6 present the clustering time of algorithms for 3 datasets of EPPO and UCI.

On the performance graph, the clustering results of the proposed algorithm are more obvious and less noise-intensive than those of the existing methods. The graphs show that the clustering results are nearest-neighbor groups will have the same color. It is clear that our algorithm genereates obvious clusters in the dataset compared to the other algorithms. Besides, it has less noise-intensive elements which are



Machine

Ecoli

Pima-indians-diabetes

Student

Transfusion

Voting-records

Climate Model

Adult

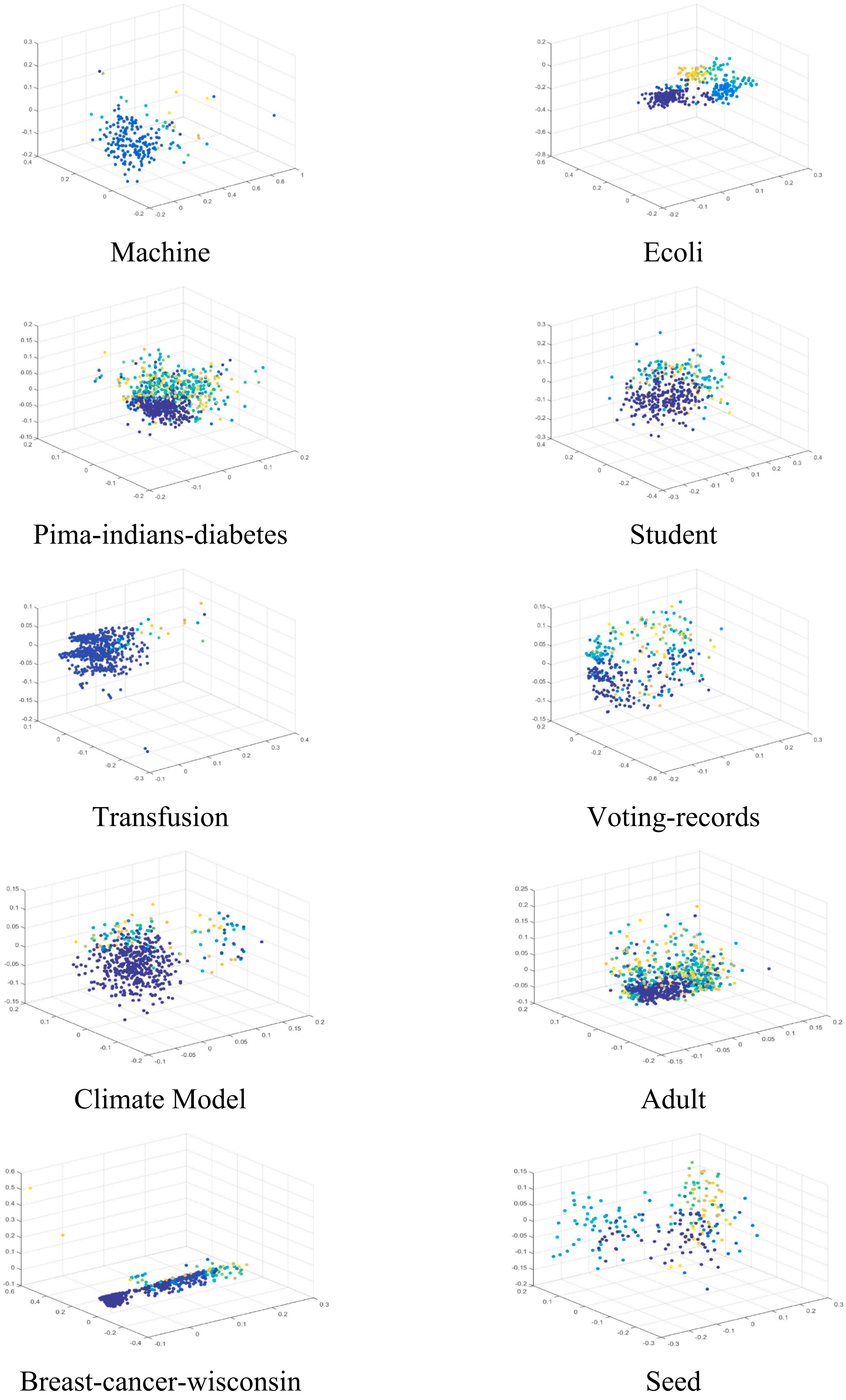
Breast-cancer-wisconsin

Seed

Fig. 6. Clustering result of Ye (2014) with 10 UCI dataset.

far from others.

For clustering quality, the proposed algorithm has higher SSWC, IFV, BH and VRC values than those of Ye (2014), Ye (2016) and Huang algorithms. The dataset which has the large number of elements such as: eppo_standard_pp1, Adult, Pima-indians-diabetes, proposed algorithm mostly show the better indices value compared to Ye (2014), Ye (2016) and Huang algorithms. About the running time of algorithms, most of the results show that the running time of the proposed algorithm is better than that of Huang algorithm and is longer than Ye (2016) and Ye (2014) algorithms. The evaluation indicators show that Ye (2016) algorithm has nearly similar indexes to Ye (2014) but has the better running time. Our algorithm has less running time compared to



Machine

Ecoli

Pima-indians-diabetes

Student

Transfusion

Voting-records

Climate Model

Adult

Breast-cancer-wisconsin

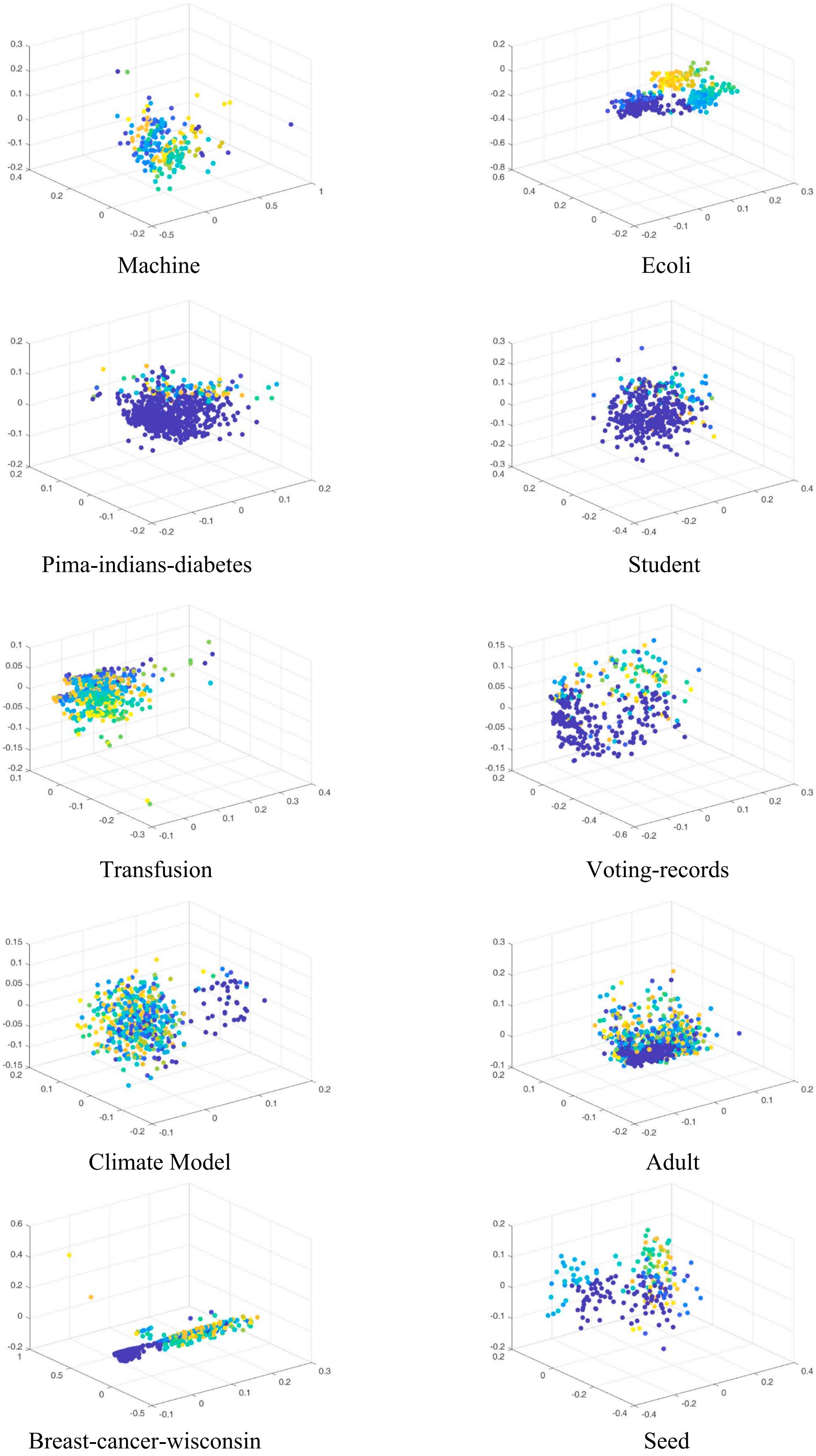
Seed

Fig. 7. Clustering result of Ye (2016) with 10 UCI dataset.

Ye (2014) and Huang algorithms with datasets such as eppo_standard_pm4, Ecoli, Transfusion, Adult, Breast-cancer-wisconsin.

4. Conclusions

This paper proposed a new fuzzy clustering algorithm based on association matrix using the neutrosophic set. After constructing a neutrosophic association matrix from the data, a neutrosophic equivalent matrix is designed based on association matrix. The next step is to construct the lambda-cutting matrix based on neutrosophic equivalent matrix by the lambda-cutting step. Finally, the clusters are defined on the basis of lambda-cutting matrix. To assess the quality of clusters,



Machine

Ecoli

Pima-indians-diabetes

Student

Transfusion

Voting-records

Climate Model

Adult

Breast-cancer-wisconsin

Seed

Fig. 8. Clustering result of Huang with 10 UCI dataset.

Table 3
Comparative result of proposed method with existing works on EPPO dataset (**Bold** shows the best results in a column).

Dataset	Algorithms	DB	SSWC	IFV	BH	VRC
eppo_standard_pp1	Ye (2014)	28.155232	0.557163	399.482864	168.9738	17.35721
	Ye (2016)	30.859587	0.712286	559.711	168.9738	5.687921
	Huang	**11.39557**	0.7122857	169.1306	502.946	17.39153
	Our algorithms	508.395223	**0.998623**	**589.791505**	**954.5629**	**549.1699**
eppo_standard_pm4	Ye (2014)	**5.720429**	0.660127	20036.600822	809.8115	2.439143
	Ye (2016)	**5.720429**	0.712286	559711.19511	809.8115	17.39153
	Huang	37.41549	0.5571625	120343.9	457.7613	14.17068
	Our algorithms	145.697383	**1**	**8769117.4027**	**2421.778**	**104.8944**
eppo_standard_pm8	Ye (2014)	38.40345	0.581924	13108.155304	276.2727	35.75515
	Ye (2016)	**3.323803**	0.671745	10186.476571	276.2727	51.26642
	Huang	9.254573	0.6717447	18058.79	**678.711**	29.77364
	Our algorithms	72.11886	**1**	**475193.42427**	**678.711**	**46.57749**

different clustering validaty indices are used.

The experimental results on the EPPO and UCI datasets show that the quality of the proposed algorithm is better than the comparative clustering algorithms. The clustering results are also well distributed and noises and exceptions. However, the runtime of our algorithm is usually longer than other algorithms. Therefore, in the future, we will study the improvement of the runtime of the fuzzy clustering algorithm on the neutrosophic fuzzy sets.

Table 4
Comparative result of proposed method with existing works on UCI dataset (**Bold** shows the best results in a column).

Dataset	Algorithms	DB	SSWC	IFV	BH	VRC
Machine	Ye (2014)	**5.915397**	0.713387	2202.371205	316.5159	9.838657
	Ye (2016)	**5.915397**	0.732526	11930.22742	316.5159	13.67363
	Huang	9.61929	0.7325257	38586.25	316.5159	11.85429
	Our algorithms	47.998991	**0.976077**	**982376.1554**	**350.2164**	**35.1964**
Ecoli	Ye (2014)	**3.565735**	0.627976	**5478.954648**	503.9417	59.81451
	Ye (2016)	**3.565735**	0.669643	1623.947206	503.9417	59.81451
	Huang	7.915799	0.6696428	86658.27	561.5002	24.56112
	Our algorithms	22.725662	**0.976190**	2044.066447	**629.8456**	**87.06884**
Pima-indians-diabetes	Ye (2014)	**13.353669**	0.582044	342.650457	987.0068	5.133967
	Ye (2016)	**13.353669**	0.604180	711.035495	987.0068	8.611765
	Huang	17.10831	0.6041797	105005.5	412.7064	8.93186
	Our algorithms	28.815353	**0.998698**	**11550516.8**	**1503.658**	**215.6549**
Student	Ye (2014)	10.481221	0.701304	2415.889826	321.3741	6.521691
	Ye (2016)	**9.481033**	0.739278	436.856690	321.3741	7.394256
	Huang	10.75008	0.7392785	32736.64	**559.4407**	5.65088
	Our algorithms	17.841839	**0.997468**	**1799002.62**	343.1385	**38.59601**
Transfusion	Ye (2014)	9.021250	0.449237	29.925621	**3890.629**	31.88744
	Ye (2016)	8.447838	0.473301	1122.64171	**3890.629**	164.6639
	Huang	**6.472221**	0.4733007	136553.8	3677.43	36.65331
	Our algorithms	235.937925	**0.998663**	**25249993.7**	2241.555	**200.3411**
Voting-records	Ye (2014)	**13.752153**	0.563254	1508.53905	276.1446	5.647962
	Ye (2016)	**13.752153**	0.593139	530.391568	276.1446	17.40125
	Huang	20.55124	0.5931388	43389.4	125.984	15.03395
	Our algorithms	28.490218	**0.997701**	**2491275.55**	**385.8623**	**362.5158**
Climate Model	Ye (2014)	18.176141	0.631502	110.930930	**765.7976**	**233.9937**
	Ye (2016)	15.753252	0.664836	478.678148	**765.7976**	**233.9937**
	Huang	**14.12625**	0.6648356	45979.65	332.3088	20.20007
	Our algorithms	41.314469	**0.979630**	**3640465.96**	703.2882	145.0514
Adult	Ye (2014)	**10.765577**	0.625310	561.240913	747.0846	9.08805
	Ye (2016)	**10.765577**	0.661290	842.216071	747.0846	9.548958
	Huang	11.57448	0.6612903	150282.1	1374.654	12.08014
	Our algorithms	22.284901	**0.998759**	**1075.57670**	**1753.58**	**174.5424**
Breast-cancer-wisconsin	Ye (2014)	**6.489717**	0.648971	235.262610	**5383.542**	48.67108
	Ye (2016)	**6.489717**	0.367668	1391.64800	**5383.542**	150.4453
	Huang	7.395256	0.3676681	91756.83	1993.873	124.6162
	Our algorithms	451.37170	**0.997139**	**29578245.5**	2421.802	**622.2076**
Seed	Ye (2014)	**12.447586**	0.762020	**2269.41915**	71.16163	5.195465
	Ye (2016)	12.703965	0.804877	231.222876	71.16163	7.294668
	Huang	15.25937	0.8048768	8260.108	79.94008	6.050869
	Our algorithms	19.157250	**0.995238**	258.476251	**118.9233**	**53.81705**

Table 5
Comparison of runtime (seconds) between 3 algorithms on EPPO dataset.

Dataset	Ye (2014)	Ye (2016)	Huang	Our algorithm
eppo_standard_pp1	1501.35048	283.43089	5129.41834	4222.5463
eppo_standard_pm8	41.350550	8.953963	37.795681	77.954504
eppo_standard_pm4	4535.71586	53.485130	2413.345142	1043.3757

Table 6
Comparision of runtime (seconds) between 3 algorithms on UCI dataset.

Dataset	Ye (2014)	Ye (2016)	Huang	Our algorithm
Machine	40.357924	42.789720	98,717846	43.409019
Ecoli	81.205389	30.233782	67.699230	40.295607
Pima-indians-diabetes	116.263368	59.601126	853.627924	293.19517
Student	110.543548	36.193778	144.462638	127.70387
Transfusion	513.618820	57.827674	373.033408	278.94937
Voting-records	37.587854	19.943098	124.481567	90.105930
Climate Model	49.108346	29.036644	315.563637	134.03757
Adult	460.867275	81.948437	1628.39179	443.47297
Breast-cancer-wisconsin	1320.93752	80.591459	239.630546	292.60151
Seed	9.700136	13.612500	23.851625	30.272458

Acknowledgement

The authors would like to thank Editor-in-chiefs, Associate editor and the anonymous referees for their helpful comments and valuable suggestions, which have greatly improved the paper. This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 102.05-2018.02.

References

EPPO Global Database, https://gd.eppo.int/.

UCI Machine Learning Datasets, Available: https://archive.ics.uci.edu/ml/datasets.html.

Atanassov, K. (1986). Intuitionistic fuzzy sets. *Fuzzy Sets and Systems, 20*, 87–96.

Bezdek, J. C., Ehrlich, R., & Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. *Computers & Geosciences, 10*, 191–203.

Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 1*, 224–227.

Huang, H. (2016). New distance measure of single-valued neutrosophic sets and its application. *International Journal of Intelligent Systems, 31*, 1021–1032.

Kaufman, L., & Rouseeuw, P. J. (1991). Finding groups in data: An introduction to cluster analysis. *Journal of the American Statistical Association, 86*, 830–832.

Kuo, J., Potti, Y., & Zulvia, F. (2018). Application of metaheuristic based Fuzzy K-modes algorithm to supplier clustering. *Computers & Industrial Engineering, 120*, 298–307.

Ma, Y., Wang, J., Wang, J., & Wu, X. (2015). An interval neutrosophic linguistic multicriteria group decision-making method and its application in selecting medical treatment options. *Neural Computing and Applications, 28*, 2745–2765.

Mondaland, K., & Pramanik, S. (2015). Weighted fuzzy similarity measure based on tangent function and its application to medical diagnosis. *International Journal of Innovative Research in Science, Engineering and Technology, 4*, 158–164.

Nguyen, G. N., Son, L. H., Ashour, A. S., & Dey, N. (2018). A survey of the state-of-the-arts on neutrosophic sets in biomedical diagnoses. *International Journal of Machine Learning and Cybernetics, 8*, 1–13.

Pramanik, S., & Chackrabarti, S. N. (2013). A study on problems of construction workers in West Bengal based on neutrosophic cognitive maps. *International Journal of Innovative Research in Science, Engineering and Technology, 2*(11), 6387–6394.

Smarandache, F. (1998). *Neutrosophy: Neutrosophic probability, set, and logic.* American Research Press105.

Smarandache, F., & Vladareanu, L. (2014). Applications of Neutrosophic logic to robotics. *NeutrosophicTheoryand its Applications, 1*, 61–66.

Wu, Q., Wu, P., Zhou, L., Chen, H., & Guan, X. (2017). Some new Hamacher aggregation operators under single-valued neutrosophic 2-tuple linguistic environment and their applications to multi-attribute group decision making. *Computers & Industrial Engineering, 116*, 144–162.

Ye, J. (2014). Clustering methods using distance-based similarity measures of single-valued Neutrosophic sets. *Journal of Intelligent Systems, 23*, 379–389.

Ye, J. (2016). A netting method for clustering-simplfiedneutrosophic information. *Soft Computing, 21*, 7571–7577.

Ye, J. (2017). Single-valued neutrosophic similarity measures based on cotangent function and their application in the fault diagnosis of steam turbine. *Soft Computing, 21*, 817–825.

Ye, J. (2018). Operations and aggregation method of neutrosophic cubic numbers for multiple attribute decision-making. *Soft Computing, 22*, 1–10.

Ye, J., & Du, S. (2017). Some distances, similarity and entropy measures for interval-valued neutrosophic sets and their relationship. *International Journal of Machine Learning and Cybernetics, 8*, 1–9.

Ye, J., & Fu, J. (2016). Multi-period medical diagnosis method using a single valued neutrosophic similarity measure based on tangent function. *Computer Methods and Programs in Biomedicine, 123*, 142–149.

Ye, J., & Smarandache, F. (2016). Similarity measure of refined single-valued neutrosophic sets and its multicriteria decision making method. *Neutrosophic Sets and Systems, 12*, 41–44.

Ye, J., & Zhang, Q. S. (2014). Single valued neutrosophic similarity measures for multiple attribute decision making. *Neutrosophic Sets and Systems, 2*, 48–54.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control, 8*, 338–353.