NSS

Neutrosophic Hierarchical Clustering Algoritms

Rıdvan Şahin

Department of Mathematics, Faculty of Science, Ataturk University, Erzurum, 25240, Turkey. mat.ridone@gmail.com

Abstract. Interval neutrosophic set (INS) is a generalization of interval valued intuitionistic fuzzy set (IVIFS), whose the membership and non-membership values of elements consist of fuzzy range, while single valued neutrosophic set (SVNS) is regarded as extension of intuitionistic fuzzy set (IFS). In this paper, we extend the hierarchical clustering techniques proposed for IFSs and IVIFSs to SVNSs and INSs respectively. Based on the traditional hierarchical clustering procedure, the single valued neutrosophic aggregation operator, and the basic distance measures between SVNSs, we define a single valued neutrosophic hierarchical clustering algorithm for clustering SVNSs. Then we extend the algorithm to classify an interval neutrosophic data. Finally, we present some numerical examples in order to show the effectiveness and availability of the developed clustering algorithms.

Keywords: Neutrosophic set, interval neutrosophic set, single valued neutrosophic set, hierarchical clustering, neutrosophic aggregation operator, distance measure.

1 Introduction

Clustering is an important process in data mining, pattern recognition, machine learning and microbiology analysis [1, 2, 14, 15, 21]. Therefore, there are various types of techniques for clustering data information such as numerical information, interval-valued information, linguistic information, and so on. Several of them are clustering algorithms such as partitional, hierarchical, density-based, graph-based, model-based. To handle uncertainty, imprecise, incomplete, and inconsistent information which exist in real world, Smarandache [3, 4] proposed the concept of neutrosophic set (NS) from philosophical point of view. A neutrosophic set [3] is a generalization of the classic set, fuzzy set [13], intuitionistic fuzzy set [11] and interval valued intuitionistic fuzzy set [12]. It has three basic components independently of one another, which are truht-membership, indeterminacy-membership, and falsity-membership. However, the neutrosophic sets is be diffucult to use in real scientiffic or engineering applications. So Wang et al. [5, 6] defined the concepts of single valued neutrosophic set (SVNS) and interval neutrosophic set (INS) which is an instance of a neutrosophic set. At present, studies on the SVNSs and INSs is progressing rapidly in many different aspects [7, 8, 9, 10, 16, 18]. Yet, until now there has been little study on clustering the data represented by neutrosophic information [9]. Therefore, the existing clustering algorithms cannot cluster the neutrosophic data, so we need to develop some new techniques for clustering SVNSs and INSs.

2 Preliminaries

In this section we recall some definitions, operations and properties regarding NSs, SVNSs and INSs, which will be used in the rest of the paper.

2.1 Neutrosophic sets

Definition 1. [3] Let X be a space of points (objects) and $x \in X$. A neutrosophic set N in X is characterized by a truth-membership function T_N, an indeterminacy-membership function I_N and a falsity-membership function F_N, where $T_N(x)$, $I_N(x)$ and $F_N(x)$ are real standard or non-standard subsets of $]0^-, 1^+[$. That is, $T_N : U \to]0^-, 1^+[$, $I_N : U \to]0^-, 1^+[$ and $F_N : U \to]0^-, 1^+[$.

There is no restriction on the sum of $T_N(x)$, $I_N(x)$ and $F_N(x)$, so

$$0^- \leq \sup T_N(x) + \sup I_N(x) + \sup F_N(x) \leq 3^+.$$

Neutrosophic sets is difficult to apply in real scientific and engineering applications [5]. So Wang et al. [5] proposed the concept of SVNS, which is an instance of neutrosophic set.

2.2 Single valued neutrosophic sets

A single valued neutrosophic set has been defined in [5] as follows:

Definition 2. Let X be a universe of discourse. A single valued neutrosophic set A over X is an object having the form:

$$A = \{\langle x, u_A(x), w_A(x), v_A(x) \rangle : x \in X\},$$

where $u_A: X \to [0,1]$, $w_A: X \to [0,1]$ and $v_A: X \to [0,1]$ with the condition

$$0 \leq u_A(x) + w_A(x) + v_A(x) \leq 3, \quad \forall x \in X.$$

The numbers $u_A(x)$, $w_A(x)$ and $v_A(x)$ denote the degree of truth-membership, indeterminacy membership and falsity-membership of x to X, respectively.

Definition 3. Let A and B be two single valued neutrosophic sets,

$$A = \{\langle x, u_A(x), w_A(x), v_A(x)\rangle : x \in X\}$$
$$B = \{\langle x, u_B(x), w_B(x), v_B(x)\rangle : x \in X\}$$

Then we can give two basic operations of A and B as follows:

1. $A + B = \{< x, u_A(x) + u_B(x) - u_A(x).u_B(x), w_A(x).w_B(x), v_A(x).v_B(x) >: x \in X\};$
2. $\lambda A = \{< x, 1-(1-u_A(x))^{\lambda}, (w_A(x))^{\lambda}, (v_A(x))^{\lambda} >: x \in X \text{ and } \lambda > 0\}$

Definition 4. Let $X = \{x_1, x_2, \ldots, x_n\}$ be a universe of discourse. Consider that the elements x_i (i = 1,2,...,n) in the universe X may have different importance, let $\omega = (\omega_1, \omega_2, \ldots, \omega_n)^T$ be the weight vector of x_i (i = 1,2,...,n), with $\omega_i \geq 0$, i = 1,2,...,n, $\sum_{i=1}^{n} \omega i = 1$. Assume that

$$A = \{\langle x, u_A(x), w_A(x), v_A(x)\rangle : x \in X\} \text{ and}$$
$$B = \{\langle x, u_B(x), w_B(x), v_B(x)\rangle : x \in X\}$$

be two SVNSs. Then we give the following distance measures:
The weighted Hamming distance and normalized Hamming distance [9]

$$e_1^{\omega}(A,B) = (\tfrac{1}{3}\textstyle\sum_{i=1}^{n} \omega_i(|u_A(x_i) - u_B(x_i)| + |w_A(x_i) - w_B(x_i)| + |v_A(x_i) - v_B(x_i)|)). \quad (1)$$

Assume that $\omega = (1/n, 1/n, \ldots, 1/n)^T$, then Eq. (1) is reduced to the normalized Hamming distance

$$e_2^{n}(A,B) = (\tfrac{1}{3n}\textstyle\sum_{i=1}^{n}(|u_A(x_i) - u_B(x_i)| + |w_A(x_i) - w_B(x_i)| + |v_A(x_i) - v_B(x_i)|)) \quad (2)$$

The weighted Euclidean distance and normalized Euclidean distance [7]

$$e_3^{\omega}(A,B) = \left(\tfrac{1}{3}\textstyle\sum_{i=1}^{n} \omega_i(|u_A(x_i) - u_B(x_i)|)^2 + (|w_A(x_i) - w_B(x_i)|)^2 + (|v_A(x_i) - v_B(x_i)|)^2\right)^{\frac{1}{2}} \quad (3)$$

Assume that $\omega = (1/n, 1/n, \ldots, 1/n)^T$, then Eq. (3) is reduced to the normalized Euclidean distance

$$e_4^{n}(A,B) = \left(\tfrac{1}{3n}\textstyle\sum_{i=1}^{n}(|u_A(x_i) - u_B(x_i)|)^2 + (|w_A(x_i) - w_B(x_i)|)^2 + (|v_A(x_i) - v_B(x_i)|)^2\right)^{\frac{1}{2}} \quad (4)$$

2.3 Interval neutrosophic sets

Definition 5. [3] Let X be a set and Int[0,1] be the set of all closed subsets of [0,1]. An INS $\tilde{A}$ in X is defined with the form

$$\tilde{A} = \{\langle x, u_{\tilde{A}}(x), w_{\tilde{A}}(x), v_{\tilde{A}}(x)\rangle : x \in X\}$$

where $u_{\tilde{A}}: X \to \text{Int}[0,1]$, $w_{\tilde{A}}: X \to \text{Int}[0,1]$ and $v_{\tilde{A}}: X \to \text{Int}[0,1]$ with the condition

$$0 \leq \sup u_{\tilde{A}}(x) + \sup w_{\tilde{A}}(x) + \sup v_{\tilde{A}}(x) \leq 3,$$

for all $x \in X$.

The intervals $u_{\tilde{A}}(x)$, $w_{\tilde{A}}(x)$ and $v_{\tilde{A}}(x)$ denote the truth-membership degree, the indeterminacy membership degree and the falsity-membership degree of x to $\tilde{A}$, respectively.

For convenience, if let

$$u_{\tilde{A}}(x) = [u_{\tilde{A}}^{+}(x), u_{\tilde{A}}^{-}(x)]$$
$$w_{\tilde{A}}(x) = [w_{\tilde{A}}^{+}(x), w_{\tilde{A}}^{-}(x)]$$
$$v_{\tilde{A}}(x) = [v_{\tilde{A}}^{+}(x), v_{\tilde{A}}^{-}(x)]$$

then

$$\tilde{A} = \{\langle x, [u_{\tilde{A}}^{-}(x), u_{\tilde{A}}^{+}(x)], [w_{\tilde{A}}^{-}(x), w_{\tilde{A}}^{+}(x)], [v_{\tilde{A}}^{-}(x), v_{\tilde{A}}^{+}(x)]\rangle\} : x \in X\}$$

with the condition

$$0 \leq \sup u_{\tilde{A}}^{+}(x) + \sup w_{\tilde{A}}^{+}(x) + \sup v_{\tilde{A}}^{+}(x) \leq 3,$$

for all $x \in X$. If $w_{\tilde{A}}(x) = [0,0]$ and $\sup u_{\tilde{A}}^{+}(x) + \sup v_{\tilde{A}}^{+} \leq 1$ then $\tilde{A}$ reduces to an interval valued intuitionistic fuzzy set.

Definition 6. [20] Let $\tilde{A}$ and $\tilde{B}$ be two interval neutrosophic sets,

$$\tilde{A} = \{\langle x, [u_{\tilde{A}}^{-}(x), u_{\tilde{A}}^{+}(x)], [w_{\tilde{A}}^{-}(x), w_{\tilde{A}}^{+}(x)], [v_{\tilde{A}}^{-}(x), v_{\tilde{A}}^{+}(x)]\rangle : x \in X\},$$
$$\tilde{B} = \{\langle x, [u_{\tilde{B}}^{-}(x), u_{\tilde{B}}^{+}(x)], [w_{\tilde{B}}^{-}(x), w_{\tilde{B}}^{+}(x)], [v_{\tilde{B}}^{-}(x), v_{\tilde{B}}^{+}(x)]\rangle : x \in X\}.$$

Then two basic operations of $\tilde{A}$ and $\tilde{B}$ are given as follows:

1. $\tilde{A} + \tilde{B} = \{< x, [u_{\tilde{A}}^{-}(x) + u_{\tilde{B}}^{-}(x) -, u_{\tilde{A}}^{+}(x) \cdot u_{\tilde{B}}^{-}(x), u_{\tilde{A}}^{+}(x) + u_{\tilde{B}}^{+}(x) - u_{\tilde{A}}^{+}(x) \cdot u_{\tilde{B}}^{+}(x)], [u_{\tilde{A}}^{-}(x) \cdot w_{\tilde{B}}^{-}(x), w_{\tilde{A}}^{+}(x) \cdot w_{\tilde{B}}^{+}(x)], [v_{\tilde{A}}^{-}(x) \cdot v_{\tilde{B}}^{-}(x), v_{\tilde{A}}^{+}(x) \cdot v_{\tilde{B}}^{+}(x)] : x \in X\}$
2. $\lambda\tilde{A} = \{< x, \left[1-\left(1-u_{\tilde{A}}^{-}(x)\right)^{\lambda}, 1-\left(1-u_{\tilde{A}}^{+}(x)\right)^{\lambda}\right], \left[\left(w_{\tilde{A}}^{-}(x)\right)^{\lambda}, \left(w_{\tilde{A}}^{+}(x)\right)^{\lambda}\right], \left[\left(v_{\tilde{A}}^{-}(x)\right)^{\lambda}, \left(v_{\tilde{A}}^{+}(x)\right)^{\lambda}\right] >: x \in X \text{ and } \lambda > 0\}.$

Definition 7. Let $X = \{x_1, x_2, \ldots, x_n\}$ be a universe of discourse. Consider that the elements x_i (i = 1,2,...,n) in the universe X may have different importance, let $\omega = (\omega_1, \omega_2, \ldots, \omega_n)^T$ be the weight vector of x_i (i = 1,2,...,n), with $\omega_i \geq 0$, i = 1,2,...,n, $\sum_{i=1}^{n} \omega_i = 1$. Suppose that $\tilde{A}$ and $\tilde{B}$ are two interval neutrosophic sets. Ye [6] has defined the distance measures for INSs as follows:
The weighted Hamming distance and normalized Hamming distance:

$$d_1^{\omega}(\tilde{A},\tilde{B}) = (\tfrac{1}{6}\textstyle\sum_{i=1}^{n} \omega_i(|u_{\tilde{A}}^{-}(x) - u_{\tilde{B}}^{-}(x)| + |u_{\tilde{A}}^{+}(x) - u_{\tilde{B}}^{+}(x)| + |w_{\tilde{A}}^{+}(x) - w_{\tilde{B}}^{+}(x)| + |v_{\tilde{A}}^{-}(x) - u_{\tilde{B}}^{-}(x)| + |v_{\tilde{A}}^{+}(x) - u_{\tilde{B}}^{+}(x)|) \quad (5)$$

Assume that $\omega = (1/n, 1/n, \ldots, 1/n)^T$, then Eq. (5) is reduced to the normalized Hamming distance

$$d_2^{\omega}(\tilde{A},\tilde{B}) = (\tfrac{1}{6n}\textstyle\sum_{i=1}^{n}(|u_{\tilde{A}}^{-}(x) - u_{\tilde{B}}^{-}(x)| + |u_{\tilde{A}}^{+}(x) - u_{\tilde{B}}^{+}(x)| + |w_{\tilde{A}}^{+}(x) - w_{\tilde{B}}^{+}(x)| + |v_{\tilde{A}}^{-}(x) - u_{\tilde{B}}^{-}(x)| +$$

$|v_{\tilde{A}}^{+}(x)-u_{\tilde{B}}^{+}(x)|)$ (6)

The weighted Euclidean distance and normalized Hamming distance

$$d_3^{\omega}(\tilde{A},\tilde{B})=\Big(\tfrac{1}{6}\textstyle\sum_{i=1}^{n}\omega_i(\left|u_{\tilde{A}}^{-}(x)-u_{\tilde{B}}^{-}(x)\right|^2+\left|u_{\tilde{A}}^{+}(x)-u_{\tilde{B}}^{+}(x)\right|^2+\left|w_{\tilde{A}}^{-}(x)-w_{\tilde{B}}^{-}(x)\right|^2+\left|w_{\tilde{A}}^{+}(x)-w_{\tilde{B}}^{+}(x)\right|^2+\left|v_{\tilde{A}}^{-}(x)-u_{\tilde{B}}^{-}(x)\right|^2+|v_{\tilde{A}}^{+}(x)-u_{\tilde{B}}^{+}(x)|^2)\Big)^{\frac{1}{2}} \quad (7)$$

Assume that $\omega=(1/n,1/n,\ldots,1/n)^T$, then Eq. (7) is reduced to the normalized Hamming distance

$$d_4^{n}(\tilde{A},\tilde{B})=\Big(\tfrac{1}{6n}\textstyle\sum_{i=1}^{n}(\left|u_{\tilde{A}}^{-}(x)-u_{\tilde{B}}^{-}(x)\right|^2+\left|u_{\tilde{A}}^{+}(x)-u_{\tilde{B}}^{+}(x)\right|^2+\left|w_{\tilde{A}}^{-}(x)-w_{\tilde{B}}^{-}(x)\right|^2+\left|w_{\tilde{A}}^{+}(x)-w_{\tilde{B}}^{+}(x)\right|^2+\left|v_{\tilde{A}}^{-}(x)-u_{\tilde{B}}^{-}(x)\right|^2+\left|v_{\tilde{A}}^{+}(x)-u_{\tilde{B}}^{+}(x)\right|^2)\Big)^{\frac{1}{2}} \quad (8)$$

Definition 8. [20] Let

$$\tilde{A}_k=\langle[u_{\tilde{A}}^{-}(x),u_{\tilde{A}}^{+}(x)],[w_{\tilde{A}}^{-}(x),w_{\tilde{A}}^{+}(x)],[v_{\tilde{A}}^{-}(x),v_{\tilde{A}}^{+}(x)]\rangle$$

$k=1,2,\ldots,.n)$ be a collection of interval neutrosophic sets. A mapping $\tilde{F}\omega: INS^n\to INS$ is called an interval neutrosophic weighted averaging operator of dimension n if it is satisfies

$$\tilde{F}\omega(\tilde{A}_1,\tilde{A}_2,\ldots,\tilde{A}_n)=\textstyle\sum_{k=1}^{n}\omega_k\tilde{A}_k$$

where $\omega=(\omega_1,\omega_2,\ldots,\omega_n)^T$ is the weight vector of $\tilde{A}_k$ $(k=1,2,\ldots,n)$, $\omega_k\in[0,1]$ and $\sum_{k=1}^{n}\omega_k=1$.

Theorem 1. [20] Suppose that

$$\tilde{A}_k=\langle[u_{\tilde{A}}^{-}(x),u_{\tilde{A}}^{+}(x)],[w_{\tilde{A}}^{-}(x),w_{\tilde{A}}^{+}(x)],[v_{\tilde{A}}^{-}(x),v_{\tilde{A}}^{+}(x)]\rangle$$

$k=1,2,\ldots,.n)$ are interval neutrosophic sets. Then the aggregation result through using the interval neutrosophic weighted averaging operator F_ω is an interval neutrosophic set and

$$\begin{aligned}\tilde{F}\omega(\tilde{A}_1,\tilde{A}_2,\ldots,\tilde{A}_n)&=\tilde{A}_k\\ =<&\Big[1-\textstyle\prod_{k=1}^{n}\big(1-u_{\tilde{A}_k}^{-}(x)\big)^{\omega_k},1-\prod_{k=1}^{n}\big(1-u_{\tilde{A}_k}^{+}(x)\big)^{\omega_k}\Big],\\ &\Big[\textstyle\prod_{k=1}^{n}\big(w_{\tilde{A}_k}^{-}(x)\big)^{\omega_k},\prod_{k=1}^{n}\big(w_{\tilde{A}_k}^{+}(x)\big)^{\omega_k}\Big],\\ &\Big[\textstyle\prod_{k=1}^{n}\big(v_{\tilde{A}_k}^{-}(x)\big)^{\omega_k},\prod_{k=1}^{n}\big(v_{\tilde{A}_k}^{+}(x)\big)^{\omega_k}\Big]>\end{aligned} \quad (9)$$

where $\omega=(\omega_1,\omega_2,\ldots,\omega_n)^T$ is the weight vector of $\tilde{A}_k$ $(k=1,2,\ldots,n)$, $\omega_k\in[0,1]$ and $\sum_{k=1}^{n}\omega_k=1$.

Suppose that $\omega=(1/n,1/n,\ldots,1/n)^T$ then the $\tilde{F}\omega$ is called an arithmetic average operator for INSs.

Since INS is a generalization of SVNS, according to Definition 8 and Theorem 1, the single valued neutrosophic weighted averaging operator can be easily obtained as follows.

Definition 9. *Let*

$$A_k=\langle u_{A_k},w_{A_k},v_{A_k}\rangle$$

$(k=1,2,\ldots,n)$ *be* a collection single valued neutrosophic sets. A mapping $F\omega: SVNS^n\to SVNS$ *is* called a single valued neutrosophic weighted averaging operator of dimension n if it is satisfies

$$F\omega(A_1,A_2,\ldots,A_n)=\textstyle\sum_{k=1}^{n}\omega_kA_k$$

where $\omega=(\omega_1,\omega_2,\ldots,\omega_n)^T$ is the weight vector of $A_k(k=1,2,\ldots,n)$, $\omega_k\in[0,1]$ and $\sum_{k=1}^{n}\omega_k=1$.

Theorem 2. Suppose that

$$A_k=\langle u_{A_k},w_{A_k},v_{A_k}\rangle$$

$(k=1,2,\ldots,n)$ are single valued neutrosophic sets. Then the aggregation result through using the single valued neutrosophic weighted averaging operator $F\omega$ is single neutrosophic set and

$$\begin{aligned}F\omega(A_1,A_2,\ldots,A_n)&=A_k\\ &=<1-\textstyle\prod_{k=1}^{n}\big(1-u_{A_k}(x)\big)^{\omega_k},\\ &\textstyle\prod_{k=1}^{n}\big(w_{A_k}(x)\big)^{\omega_k},\prod_{k=1}^{n}\big(v_{A_k}(x)\big)^{\omega_k}>\end{aligned} \quad (10)$$

where $\omega=(\omega_1,\omega_2,\ldots,\omega_n)^T$ is the weight vector of $A_k(k=1,2,\ldots,n)$, $\omega_k\in[0,1]$ and $\sum_{k=1}^{n}\omega_k=1$.

Suppose that $\omega=(1/n,1/n,\ldots,1/n)^T$, then the $F\omega$ is called an arithmetic average operator for SVNSs.

3 Neutrosophic hierarchical algorithms

The traditional hierarchical clustering algorithm [17, 19] is generally used for clustering numerical information. By extending the traditional hierarchical clustering algorithm, Xu [22] introduced an intuitionistic fuzzy hierarchical clustering algorithm for clustering IFSs and extended it to IVIFSs. However, they fail to deal with the data information expressed in neutrosophic environment. Based on extending the intuitionistic fuzzy hierarchical clustering algorithm and its extended form, we propose the neutrosophic hierarchical algorithms which are called the single valued neutrosophic hierarchical clustering algorithm and interval neutrosophic hierarchical clustering algorithm.

Algorithm 1. Let us consider a collection of n SVNSs $A_k(k=1,2,\ldots,n)$. In the first stage, the algorithm starts by assigning each of the n SVNSs to a single cluster. Based on the weighted Hamming distance (1) or the weighted Euclidean distance (3), the SVNSs $A_k(k=1,2,\ldots,n)$ are then compared among themselves and are merged them into a single cluster according to the closest (with smaller distance) pair of clusters. The process are continued until all the SVNSs A_k are merged into one cluster i.e., clustered into a single cluster of size *n*. In each stage, only two clusters can be merged and they cannot be separated after they are merged, and the center of each cluster is recalculated by using the arithmetic average (from Eq. (10)) of the SVNSs proposed to the cluster. The distance between the centers of each cluster is considered as the distance between two clusters.

However, the clustering algorithm given above cannot cluster the interval neutrosophic data. Therefore, we need another clustering algorithm to deal with the data represented by INSs.

Algorithm 2. Let us consider a collection of n INSs $\tilde{A}_k(k = 1,2,\dots,n)$. In the first stage, the algorithm starts by assigning each of the n INSs to a single cluster. Based on the weighted Hamming distance (5) or the weighted Euclidean distance (7), the INSs $\tilde{A}_k(k = 1,2,\dots,n)$ are then compared among themselves and are merged them into a single cluster according to the closest (with smaller distance) pair of clusters. The process are continued until all the INSs $\tilde{A}_k$ are merged into one cluster i.e., clustered into a single cluster of size n. In each stage, only two clusters can be merged and they cannot be separated after they are merged, and the center of each cluster is recalculated by using the arithmetic average (from Eq. (9)) of the INSs proposed to the cluster. The distance between the centers of each cluster is considered as the distance between two clusters.

3.1 Numerical examples.

Let us consider the clustering problem adapted from [21].

Example 1. Assume that five building materials: sealant, floor varnish, wall paint, carpet, and polyvinyl chloride flooring, which are represented by the SVNSs $A_k(k = 1,2,\dots,5)$ in the feature space $X = \{x_1, x_2, \dots, x_8\}$. $\omega = (0.15,0.10,0.12,0.15,0.10,0.13,0.14,0.11)$ is the weight vector of $x_i(i = 1,2,\dots,8)$, and the given data are listed as follows:

$$A_1 = \{(x_1, 0.20,0.05,0.50), (x_2, 0.10,0.15,0.80), (x_3, 0.50,0.05,0.30), (x_4, 0.90,0.55,0.00), (x_5, 0.40,0.40,0.35), (x_6, 0.10,0.40,0.90), (x_7, 0.30,0.15,0.50), (x_8, 1.00,0.60,0.00),\}$$

$$A_2 = \{(x_1, 0.50,0.60,0.40), (x_2, 0.60,0.30,0.15)\}, (x_3, 1.00,0.60,0.00), (x_4, 0.15,0.05,0.65), (x_5, 0.00,0.25,0.80), (x_6, 0.70,0.65,0.15), (x_7, 0.50,0.50,0.30), (x_8, 0.65,0.05,0.20)\}$$

$$A_3 = \{(x_1, 0.45,0.05,0.35), (x_2, 0.60,0.50,0.30)\}, (x_3, 0.90,0.05,0.00), (x_4, 0.10,0.60,0.80), (x_5, 0.20,0.35,0.70), (x_6, 0.60,0.40,0.20), (x_7, 0.15,0.05,0.80), (x_8, 0.20,0.60,0.65)\}$$

$$A_4 = \{(x_1, 1.00,0.65,0.00), (x_2, 1.00,0.25,0.00)\}, (x_3, 0.85,0.65,0.10), (x_4, 0.20,0.05,0.80), (x_5, 0.15,0.30,0.85), (x_6, 0.10,0.60,0.70), (x_7, 0.30,0.60,0.70), (x_8, 0.50,0.35,0.70)\}$$

$$A_5 = \{(x_1, 0.90,0.20,0.00), (x_2, 0.90,0.40,0.10), (x_3, 0.80,0.05,0.10), (x_4, 0.70,0.45,0.20), (x_5, 0.50,0.25,0.15), (x_6, 0.30,0.30,0.65), (x_7, 0.15,0.10,0.75), (x_8, 0.65,0.50,0.80)\}$$

Now we utilize Algorithm 1 to classify the building materials $A_k(k = 1,2,\dots,5)$:

Step1 In the first stage, each of the SVNSs $A_k(k = 1,2,\dots,5)$ is considered as a unique cluster $\{A_1\}, \{A_2\}, \{A_3\}, \{A_4\}, \{A_5\}$.

Step2 Compare each SVNS A_k with all the other four SVNSs by using Eq. (1):

$$e_1^\omega(A_1, A_2) = d_1(A_2, A_1) = 0.6403$$
$$e_1^\omega(A_1, A_3) = d_1(A_3, A_1) = 0.5191$$
$$e_1^\omega(A_1, A_4) = d_1(A_4, A_1) = 0.7120$$
$$e_1^\omega(A_1, A_5) = d_1(A_5, A_1) = 0.5435$$
$$e_1^\omega(A_2, A_3) = d_1(A_3, A_2) = 0.5488$$
$$e_1^\omega(A_2, A_4) = d_1(A_4, A_2) = 0.4546$$
$$e_1^\omega(A_2, A_5) = d_1(A_5, A_2) = 0.6775$$
$$e_1^\omega(A_3, A_4) = d_1(A_4, A_3) = 0.3558$$
$$e_1^\omega(A_3, A_5) = d_1(A_5, A_3) = 0.2830$$
$$e_1^\omega(A_4, A_5) = d_1(A_5, A_4) = 0.3117$$

and hence

$$e_1^\omega(A_1, A_3) = \min\{e_1^\omega(A_1, A_2), e_1^\omega(A_1, A_3), e_1^\omega(A_1, A_4), e_1^\omega(A_1, A_5)\} = 0.5191,$$

$$e_1^\omega(A_2, A_4) = \min\{e_1^\omega(A_2, A_1), e_1^\omega(A_2, A_3), e_1^\omega(A_2, A_4), e_1^\omega(A_2, A_5)\} = 0.4546,$$

$$e_1^\omega(A_3, A_5) = \min\{e_1^\omega(A_3, A_1), e_1^\omega(A_3, A_2), e_1^\omega(A_3, A_4), e_1^\omega(A_3, A_5)\} = 0.2830.$$

Then since only two clusters can be merged in each stage, the SVNSs $A_k(k = 1,2,\dots,5)$ can be clustered into the following three clusters at the second stage $\{A_1\}, \{A_2, A_4\}, \{A_3, A_5\}$.

Step3 Calculate the center of each cluster by using Eq. (10)

$$c\{A_1\} = A_1$$

$$c\{A_2, A_4\} = F_\omega(A_2, A_4) = \{(x_1, 1.00,0.62,0.00), (x_2, 1.00,0.27,0.00), (x_3, 1.00,0.62,0.00), (x_4, 0.17,0.05,0.72), (x_5, 0.07,0.27,0.82), (x_6, 0.48,0.62,0.32), (x_7, 0.40,0.54,0.45), (x_8, 0.58,0.13,0.37)\}$$

$$c\{A_3, A_5\} = F_\omega(A_3, A_5) = \{(x_1, 0.76,0.10,0.00), (x_2, 0.80,0.44,0.17), (x_3, 0.85,0.05,0.00), (x_4, 0.48,0.51,0.40), (x_5, 0.36,0.29,0.32), (x_6, 0.47,0.34,0.36), (x_7, 0.15,0.07,0.77), (x_8, 0.47,0.54,0.72)\}.$$

and then compare each cluster with the other two clusters by using Eq. (1):

$$e_1^{\omega}(c\{A_1\}, c\{A_2, A_4\}) = e_1^{\omega}(c\{A_2, A_4\}, c\{A_1\}) = 0.7101,$$
$$e_1^{\omega}(c\{A_1\}, c\{A_3, A_5\}) = e_1^{\omega}(c\{A_3, A_5\}, c\{A_1\}) = 5266,$$
$$e_1^{\omega}(c\{A_2, A_4\}, c\{A_3, A_5\}) = e_1^{\omega}(c\{A_3, A_5\}, c\{A_2, A_4\}) = 0.4879.$$

Subsequently, the SVNSs $A_k(k = 1,2,\ldots,5)$ can be clustered into the following two clusters at the third stage $\{A_1\}, \{A_2, A_3, A_4, A_5\}$.

Finally, the above two clusters can be further clustered into a unique cluster $\{A_1, A_2, A_3, A_4, A_5\}$.

All the above processes can be presented as in Fig. 1.

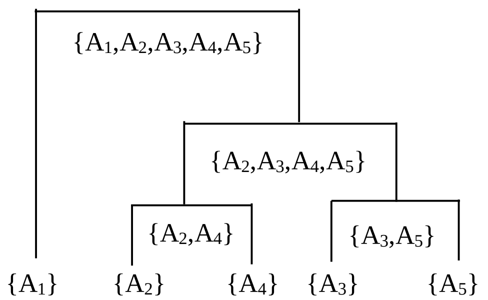

FIGURE 1: Classification of the building materials $A_k(k = 1,2,\ldots,5)$

Example 2. Consider four enterprises, represented by the INSs $\tilde{A}_k(k = 1,2,3,4)$ in the attribute set $X = \{x_1, x_2, \ldots, x_6\}$, where (1) x_1–the ability of sale; (2) x_2–the ability of management; (3) x_3–the ability of production; (4) x_4–the ability of technology; (5) x_5–the ability of financing; (6) x_6–the ability of risk bearing (the weight vector of $x_i(i = 1,2,\ldots,6)$ *is* $\omega = (0.25,0.20,15,0.10,0.15,0.15)$. The given data are listed as follows.

$$\tilde{A}_1 = \{(x_1, [0.70,0.75], [0.25,0.45], [0.10,0.15]), (x_2, [0.00,0.10], [0.15,0.15], [0.80,0.90]), (x_3, [0.15,0.20], [0.05,0.35], [0.60,0.65]), (x_4, [0.50,0.55], [0.45,0.55], [0.30,0.35]), (x_5, [0.10,0.15], [0.40,0.60], [0.50,0.60]), (x_6, [0.70,0.75], [0.20,0.25], [0.10,0.15])\}$$

$$\tilde{A}_2 = (x_1, [0.40,0.45], [0.00,0.15], [0.30,0.35]), (x_2, [0.60,0.65], [0.10,0.25], [0.20,0.30]), (x_3, [0.80,1.00], [0.05,0.75], [0.00,0.00]), (x_4, [0.70,0.90], [0.35,0.65], [0.00,1.00]), (x_5, [0.70,0.75], [0.15,0.55], [0.10,0.20]), (x_6, [0.90,1.00], [0.30,0.35], [0.00,0.00])\}.$$

$$\tilde{A}_3 = (x_1, [0.20,0.30], [0.85,0.60], [0.40,0.45), (x_2, [0.80,0.90], [0.10,0.25], [0.00,0.10]), (x_3, [0.10,0.20], [0.00,0.05], [0.70,0.80]), (x_4, [0.15,0.20], [0.25,0.45], [0.70,0.75]), (x_5, [0.00,0.10], [0.25,0.35], [0.80,0.90]), (x_6, [0.60,0.70], [0.15,0.25], [0.20,0.30])\}.$$

$$\tilde{A}_4 = (x_1, [0.60,0.65], [0.05,0.10], [0.30,0.35]), (x_2, [0.45,0.50], [0.45,0.55], [0.30,0.40]), (x_3, [0.20,0.25], [0.05,0.25], [0.65,0.70]), (x_4, [0.20,0.30], [0.35,0.45], [0.50,0.60]), (x_5, [0.00,0.10], [0.35,0.75], [0.75,0.80]), (x_6, [0.50,0.60], [0.00,0.05], [0.20,0.25])\}.$$

Here Algorithm 2 can be used to classify the enterprises$\tilde{A}_k(k = 1,2,3,4)$:

Step 1 In the first stage, each of the INSs $\tilde{A}_k(k = 1,2,3,4)$ is considered as a unique cluster $\{\tilde{A}_1\}, \{\tilde{A}_2\}, \{\tilde{A}_3\}, \{\tilde{A}_4\}$

Step 2 Compare each INS $\tilde{A}_k$ with all the other three INSs by using Eq. (5)

$$d_1^{\omega}(\tilde{A}_1, \tilde{A}_2) = d_1^{\omega}(\tilde{A}_2, \tilde{A}_1) = 0.3337,$$
$$d_1^{\omega}(\tilde{A}_1, \tilde{A}_3) = d_1^{\omega}(\tilde{A}_3, \tilde{A}_1) = 0.2937,$$
$$d_1^{\omega}(\tilde{A}_1, \tilde{A}_4) = d_1^{\omega}(\tilde{A}_4, \tilde{A}_1) = 0.2041,$$
$$d_1^{\omega}(\tilde{A}_2, \tilde{A}_3) = d_1^{\omega}(\tilde{A}_3, \tilde{A}_2) = 0.3508,$$
$$d_1^{\omega}(\tilde{A}_2, \tilde{A}_4) = d_1^{\omega}(\tilde{A}_4, \tilde{A}_2) = 0.2970,$$
$$d_1^{\omega}(\tilde{A}_3, \tilde{A}_4) = d_1^{\omega}(\tilde{A}_4, \tilde{A}_3) = 0.2487,$$

then the INSs $\tilde{A}_k(k = 1,2,3,4)$ can be clustered into the following three clusters at the second stage $\{\tilde{A}_1, \tilde{A}_4\}, \{\tilde{A}_2\}, \{\tilde{A}_3\}$.

Step 3 Calculate the center of each cluster by using Eq. (9)

$$c\{\tilde{A}_2\} = \{\tilde{A}_2\}, c\{\tilde{A}_3\} = \{\tilde{A}_3\},$$
$$c\{\tilde{A}_1, \tilde{A}_4\} = F_{\omega}(\tilde{A}_1, \tilde{A}_4) = (x_1, [0.60,0.70], [0.11,0.21], [0.17,0.22]), (x_2, [0.25,0.32], [0.25,0.28], [0.48,0.60]), (x_3, [0.17,0.22], [0.05,0.29], [0.62,0.67]), (x_4, [0.36,0.43], [0.39,0.49], [0.38,0.45]), (x_5, [0.05,0.12], [0.37,0.67], [0.61,0.69]), (x_6, [0.61,0.68], [0.00,0.011], [0.14,0.19])\}.$$

and then compare each cluster with the other two clusters by using Eq. (5)

$d_1^{\omega}(c\{\widetilde{A}_2\}, c\{\widetilde{A}_3\}) = d_1^{\omega}(c\{\widetilde{A}_3\}, c\{\widetilde{A}_2\}) = 0.3508$
$d_1^{\omega}(c\{\widetilde{A}_2\}, c\{\widetilde{A}_1, \widetilde{A}_4\}) = d_1^{\omega}(c\{\widetilde{A}_4, \widetilde{A}_1\}, c\{\widetilde{A}_2\}) = 0.3003$
$d_1^{\omega}(c\{\widetilde{A}_3\}, c\{\widetilde{A}_1, \widetilde{A}_4\}) = d_1^{\omega}(c\{\widetilde{A}_4, \widetilde{A}_1\}, c\{\widetilde{A}_3\}) = 0.2487.$

then the INSs $\tilde{A}_k (k = 1,2,3,4)$ can be clustered into the following two clusters in the third stage $\{\tilde{A}_2\}, \{\tilde{A}_1, \tilde{A}_3, \tilde{A}_4\}$.

In the final stage, the above two clusters can be further clustered into a unique cluster $\{\tilde{A}_1, \tilde{A}_2, \tilde{A}_3, \tilde{A}_4\}$.

Note that the clustering results obtained in Example 1 and 2 are different from ones in [21].

All the above processes can be presented as in Fig. 2.

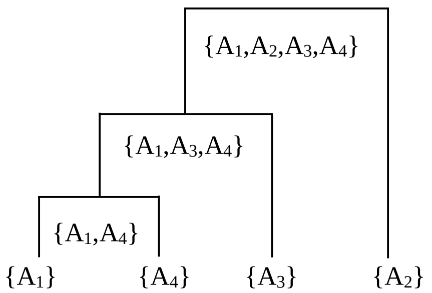

FIGURE 2: Classification of the enterprises $\tilde{A}_k (k = 1,2,3,4)$

Interval neutrosophic information is a generalization of interval valued intuitionistic fuzzy information while the single valued neutrosophic information extends the intuitionistic fuzzy information. In other words, The components of IFS and IVIFS are defined with respect to T and F, i.e., membership and nonmembership only, so they can only handle incomplete information but not the indetermine information. Hence INS and SVNS, whose components are the truth membership, indeterminacy-membership and falsity membership functions, are more general than others that do not include the indeterminacy-membership. Therefore, it is a natural outcome that the neutrosophic hierarchical clustering algorithms developed here is the extension of both the intuitionistic hierarchical clustering algorithm and its extend form. The above expression clearly indicates that clustering analysis under neutrossophic environment is more general and more practical than existing hierarchical clustering algorithms.

4 Conclusion

To cluster the data represented by neutrosophic information, we have discussed on the clustering problems of SVNSs and INSs. Firstly, we have proposed a single valued neutrosophic hierarchical algorithm for clustering SVNSs, which is based on the traditional hierarchical clustering procedure, the single valued neutrosophic aggregation operator, and the basic distance measures between SVNSs. Then, we have extented the algorithm to INSs for clustering interval neutrosophic data. Finally, an illustrative example is presented to demonstrate the application and effectiveness of the developed clustering algorithms. Since the NSs are a more general platform to deal with uncertainties, the proposed neutrosophic hierarchical algorithms are more priority than the other ones. In the future we will focus our attention on the another clustering methods of neutrosophic information.

References

[1] A.K. Jain, M.N. Murty and P.J. Flynn, Data clustering: A view. ACM Computing Surveys, 31 (1999) 264–323.
[2] B. S. Everitt, S. Landau, M. Leese. Cluster analysis. New York: Oxford University Press, (2001).
[3] F. Smarandache, A unifying field in logic neutrosophy: Neutrosophic probability, set and logic, American Research Pres, Rehoboth, 1999.
[4] F. Smarandache, A Unifying Field in Logics Neutrosophic Logic. Neutrosophy, Neutrosophic Set, Neutrosophic Probability, American Research Press, 2003.
[5] H. Wang, F. Smarandache, Y. Q. Zhang and R. Sunderraman, Single valued neutrosophic sets, Multispace and Multistructure 4 (2010) 410-413.
[6] H. Wang, F. Smarandache, Y.Q. Zhang, and R. Sunderraman, Interval neutrosophic sets and logic: Theory and applications in computing, Hexis, Phoenix, AZ, (2005).
[7] H. D. Cheng and Y. Guo, A new neutrosophic approach to image thresholding. New Math. Nat. Comput. 4(3) (2008) 291-308.
[8] J. Ye, Similarity measures between interval neutrosophic sets and their applications in multicriteria decisionmaking, Journal of Intelligent and Fuzzy Systems (2013) doi: 10.3233/IFS-120724.
[9] J. Ye, Single-Valued Neutrosophic Minimum Spanning Tree and Its Clustering Method, Journal of Intelligent Systems, (2014) DOI: 10.1515/jisys-2013-0075.
[10] J. Ye, Single valued neutrosophic cross-entropy for multicriteria decision making problems, Applied Mathematical Modelling (2013), DOI:10.1016/j.apm.2013.07.020.
[11] K. Atanassov, Intuitionistic fuzzy sets, Fuzzy Sets and Systems 20 (1986), 87-96.
[12] K. Atanassov and G. Gargov, Interval valued intuitionistic fuzzy sets, Fuzzy Sets and Systems 31 (1989), 343-349.
[13] L.A. Zadeh, Fuzzy Sets, Information and Control 8 (1965) 338-353.
[14] M. Eissen, P. Spellman, P. Brown and D. Botstein, Cluster analysis and display of genome- wide expression patterns. Proceeding of National Academy of Sciences of USA, (1998) 95: 14863–14868.
[15] M. R. Anderberg, Cluster analysis for applications. New York: Academic Press, (1972).
[16] P. Majumdar, S. K. Samant, On similarity and entropy of neutrosophic sets, Journal of Intelligent and fuzzy Systems, (2013) Doi:10.3233/IFS-130810.
[17] S.A. Mingoti, J.O. Lima, Comparing SOM neural network

with fuzzy c-means k-means and traditional hierarchical clustering algorithms. European Journal of Operational Research, (2006) 174: 1742–1759.

[18] Y. Guo and H. D. Cheng, New neutrosophic approach to image segmentation. Pattern Recognition, 42 (2009) 587-595.

[19] Y. H. Dong, Y. T. Zhuang, K. Chen and X. Y. Tai (2006). A hierarchical clustering algorithm based on fuzzy graph connectedness. Fuzzy Sets and Systems, 157, 1760–1774.

[20] Z. Hongyu, J. Qiang Wang, and X. Chen, Interval Neutrosophic Sets and its Application in Multi-criteria Decision Making Problems, The Scientific World Journal, to appear (2013).

[21] Z. M. Wang, Yeng Chai Soh, Qing Song, Kang Sim, Adaptive spatial information-theoretic clustering for image segmentation, Pattern Recognition, 42 (2009) 2029-2044.

[22] Z. S. Xu, Intuitionistic fuzzy hierarchical clustering algorithms, Journal of Systems Engineering and Electronics 20 (2009) 90–97.

Received: January 13th, 2014. Accepted: February 3th, 2014