DE GRUYTER DOI 10.1515/jisys-2013-0091 — Journal of Intelligent Systems 2014; 23(4): 379–389

Jun Ye*

Clustering Methods Using Distance-Based Similarity Measures of Single-Valued Neutrosophic Sets

Abstract: Clustering plays an important role in data mining, pattern recognition, and machine learning. Single-valued neutrosophic sets (SVNSs) are useful means to describe and handle indeterminate and inconsistent information that fuzzy sets and intuitionistic fuzzy sets cannot describe and deal with. To cluster the data represented by single-valued neutrosophic information, this article proposes single-valued neutrosophic clustering methods based on similarity measures between SVNSs. First, we define a generalized distance measure between SVNSs and propose two distance-based similarity measures of SVNSs. Then, we present a clustering algorithm based on the similarity measures of SVNSs to cluster single-valued neutrosophic data. Finally, an illustrative example is given to demonstrate the application and effectiveness of the developed clustering methods.

Keywords: Neutrosophic set, single-valued neutrosophic set, clustering algorithm, distance measure, similarity measure.

***Corresponding author: Jun Ye,** Department of Electrical and Information Engineering, Shaoxing University, 508 Huancheng West Road, Shaoxing, Zhejiang Province 312000, P.R. China, Tel.: +86-575-88327323, e-mail: yehjun@aliyun.com

1 Introduction

Clustering plays an important role in data mining, pattern recognition, information retrieval, microbiology analysis, and machine learning. Clustering data sets into disjoint groups is a problem arising in many domains. Generally, the goal of clustering is to find groups that are both homogeneous and well separated; that is, entities within the same group should be similar and entities in different groups dissimilar. However, because of the fuzziness and uncertainty of many practical problems in the real world, Zadeh [12] first proposed the theory of fuzzy sets, which has achieved great success in various fields. Fuzzy clustering analysis is a fundamental but important tool in fuzzy data analysis. Thus, Ruspini [3] first presented the concept of fuzzy division and a fuzzy clustering approach. Later, the intuitionistic fuzzy set (IFS) introduced by Atanassov [1] has been found to be highly useful in dealing with vagueness. The concept of IFSs is a generalization of that of fuzzy sets. The IFSs consider three aspects of information: membership, non-membership, and hesitancy. Therefore, it is much more flexible and practical than traditional fuzzy sets in dealing with vagueness and uncertainty problems. Hence, Zhang et al. [13] and Xu et al. [7] proposed clustering algorithms for IFSs based on association coefficients and similarity measures of IFSs, and then extended the algorithms to cluster interval-valued IFSs (IVIFSs) proposed by Atanassov and Gargov [2]. However, in the above clustering technique, fuzzy sets, IFSs, and IVIFSs cannot describe and deal with indeterminate information and inconsistent information that exist in the real world. To represent uncertain, imprecise, incomplete, and inconsistent information, Smarandache [4] gave the concept of a neutrosophic set from a philosophical point of view. The neutrosophic set is a powerful general formal framework that generalizes the concept of the classic set, fuzzy set, interval-valued fuzzy set, IFS, IVIFS, paraconsistent set, dialetheist set, paradoxist set, and tautological set [4]. In the neutrosophic set, truth-membership, indeterminacy-membership, and falsity-membership are represented independently. However, the neutrosophic set generalizes the above-mentioned sets from a philosophical point of view and its functions $T_A(x)$, $I_A(x)$, and $F_A(x)$ are real standard or non-standard subsets of $]^-0, 1^+[$, i.e., $T_A(x): X \to\ ]^-0, 1^+[$, $I_A(x): X \to\ ]^-0, 1^+[$, and $F_A(x): X \to\ ]^-0, 1^+[$, and there is no restriction on the sum of $T_A(x)$, $I_A(x)$, and $F_A(x)$, i.e., $^-0 \le \sup T_A(x) + \sup I_A(x) + \sup F_A(x) \le 3^+$. Thus, it will be difficult to apply in real scientific and engineering areas [6]. Thus, Wang et al. [6] introduced a single-valued neutrosophic set (SVNS), which is an instance of a neutrosophic set. It can describe and handle indeterminate information and inconsistent information.

The IFS contains both the truth-membership $t_A(x)$ and the falsity-membership $f_A(x)$ with $t_A(x), f_A(x) \in [0, 1]$, and $0 \leq t_A(x) + f_A(x) \leq 1$, and can only handle incomplete information (set incompletely known) but cannot handle the indeterminate information that is the zone of ignorance of a proposition's value between truth and falsehood (inconsistent information). The indeterminacy in an IFS is $1 - t_A(x) - f_A(x)$ (i.e., hesitancy or unknown degree) by default, while the indeterminacy in a neutrosophic set is quantified explicitly, and then the component of the indeterminacy $I(x)$ can be split into more subcomponents in order to better catch the vague information in the real world [4]. However, the truth-membership, the indeterminacy-membership, and the falsity-membership are independently represented in the neutrosophic set. Its components, $T(x)$, $I(x)$, $F(x)$, are non-standard subsets included in the unitary non-standard interval$]0^-, 1^+[$ or standard subsets included in the unitary standard interval [0, 1] as in the IFS. Furthermore, the connectors in the IFS are only defined by $T(x)$ and $F(x)$ (i.e., truth-membership and falsity-membership); hence, the indeterminacy $I(x)$ is what is left from 1, while in the neutrosophic set, they can be defined by any of them (no restriction) [4]. For example, when we ask the opinion of an expert about a certain statement, he/she may say that the possibility in which the statement is true is 0.6 and the statement is false is 0.5 and the degree in which he/she is not sure is 0.2. For a neutrosophic notation, it can be expressed as $x(0.6, 0.2, 0.5)$. For another example, suppose there are 10 voters during a voting process. Five vote "aye," two vote "blackball," and three are undecided. For neutrosophic notation, it can be expressed as $x(0.5, 0.3, 0.2)$. However, these expressions are beyond the scope of the IFS. Therefore, the notion of a neutrosophic set is more general and overcomes the aforementioned issues.

Recently, Ye [8, 9] presented the correlation coefficient of SVNSs and the cross-entropy measure of SVNSs and applied them to single-valued neutrosophic decision-making problems. Wang et al. [5] proposed the theory and application of interval neutrosophic sets. Then, Ye [11] proposed similarity measures between interval neutrosophic sets and their applications in multicriteria decision making. Furthermore, Ye [10] introduced the concept of simplified neutrosophic sets and simplified neutrosophic weighted aggregation operators, and then applied them to multicriteria decision-making problems under a simplified neutrosophic environment.

Yet, until now, there have been no studies on clustering of data represented by single-valued neutrosophic information. However, the existing clustering algorithms cannot cluster single-valued neutrosophic data. Motivated by intuitionistic fuzzy clustering algorithms [7, 13], this article proposes a single-valued neutrosophic clustering algorithm to deal with data represented by SVNSs. To do so, the rest of the article is organized as follows. Section 2 introduces some basic concepts of SVNSs. Section 3 defines a generalized distance measure between SVNSs and proposes two distance-based similarity measures. In Section 4, single-valued neutrosophic clustering methods are proposed based on the similarity measures of SVNSs as an extension of intuitionistic fuzzy clustering algorithms. Section 5 gives an illustrative example and a discussion of the clustering analyses. Conclusions and further research are contained in Section 6.

2 Basic Concepts of SVNSs

The neutrosophic set is a part of neutrosophy and generalizes fuzzy sets, interval-valued fuzzy set, IFS, and IVIFS from a philosophical point of view [4]. Smarandache [4] originally gave the definition of a neutrosophic set.

Definition 1 ([4]). Let X be a space of points (objects), with a generic element in X denoted by x. A neutrosophic set A in X is characterized by a truth-membership function $T_A(x)$, an indeterminacy-membership function $I_A(x)$, and a falsity-membership function $F_A(x)$. The functions $T_A(x)$, $I_A(x)$, and $F_A(x)$ are real standard or non-standard subsets of $]^-0, 1^+[$. That is, $T_A(x): X \to]^-0, 1^+[$, $I_A(x): X \to]^-0, 1^+[$, and $F_A(x): X \to]^-0, 1^+[$. Thus, there is no restriction on the sum of $T_A(x)$, $I_A(x)$, and $F_A(x)$, so $^-0 \leq \sup T_A(x) + \sup I_A(x) + \sup F_A(x) \leq 3^+$.

Obviously, it is difficult to apply in real scientific and engineering applications [6]. Hence, Wang et al. [6] proposed an SVNS as a subclass of a neutrosophic set and introduced the definition of an SVNS.

Definition 2 ([6]). Let X be a space of points (objects) with generic elements in X denoted by x. An SVNS A in X is characterized by truth-membership function $T_A(x)$, indeterminacy-membership function $I_A(x)$, and falsity-membership function $F_A(x)$. Then, an SVNS A can be denoted by

$$A=\{\langle x, T_A(x), I_A(x), F_A(x)\rangle | x\in X\},$$

where $T_A(x), I_A(x), F_A(x) \in [0, 1]$ for each point x in X. Therefore, the sum of $T_A(x)$, $I_A(x)$, and $F_A(x)$ satisfies the condition $0 \le T_A(x) + I_A(x) + F_A(x) \le 3$.

Definition 3 ([6]). The complement of an SVNS A is denoted by A^c and is defined as $T_A{}^c(x) = F_A(x)$, $I_A{}^c(x) = 1 - I_A(x)$, $F_A{}^c(x) = T_A(x)$ for any x in X. Then

$$A^c=\{\langle x, F_A(x), 1-I_A(x), T_A(x)\rangle | x\in X\}.$$

Definition 4 ([6]). An SVNS A is contained in the other SVNS, B; $A \subseteq B$, if and only if $T_A(x) \le T_B(x)$, $I_A(x) \ge I_B(x)$, $F_A(x) \ge F_B(x)$ for any x in X.

Definition 5 ([6]). Two SVNSs A and B are equal, written as $A = B$, if and only if $A \subseteq B$ and $B \subseteq A$.

3 Distance-Based Similarity Measures between SVNSs

For two SVNSs A and B in a universe of discourse $X = \{x_1, x_2,\dots,x_n\}$, which are denoted by $A = \{\langle x_i, T_A(x_i), I_A(x_i), F_A(x_i)\rangle | x_i \in X\}$ and $B = \{\langle x_i, T_B(x_i), I_B(x_i), F_B(x_i)\rangle | x_i \in X\}$, where $T_A(x_i), I_A(x_i), F_A(x_i), T_B(x_i), I_B(x_i), F_B(x_i) \in [0, 1]$ for every $x_i \in X$. Let us consider the weight w_i $(i = 1, 2,\dots,n)$ of an element x_i $(i = 1, 2,\dots,n)$, with $w_i \ge 0$ $(i = 1, 2,\dots,n)$ and $\sum_{i=1}^{n} w_i = 1$. Then, we define the generalized single-valued neutrosophic weighted distance measure:

$$d_p(A, B)=\left\{\frac{1}{3}\sum_{i=1}^{n} w_i[|T_A(x_i)-T_B(x_i)|^p+|I_A(x_i)-I_B(x_i)|^p+|F_A(x_i)-F_B(x_i)|^p]\right\}^{1/p}, \tag{1}$$

where $p > 0$.

As the Hamming distance and Euclidean distance, which are two typical distance measures, are usually used in practical applications [11], when $p = 1, 2$, we can obtain the single-valued neutrosophic weighted Hamming distance and the single-valued neutrosophic weighted Euclidean distance, respectively, as follows:

$$d_1(A, B)=\frac{1}{3}\sum_{i=1}^{n} w_i[|T_A(x_i)-T_B(x_i)|+|I_A(x_i)-I_B(x_i)|+|F_A(x_i)-F_B(x_i)|], \tag{2}$$

$$d_2(A, B)=\left\{\frac{1}{3}\sum_{i=1}^{n} w_i[|T_A(x_i)-T_B(x_i)|^2+|I_A(x_i)-I_B(x_i)|^2+|F_A(x_i)-F_B(x_i)|^2]\right\}^{1/2}. \tag{3}$$

Therefore, Eqs. (2) and (3) are the special cases of Eq. (1).

Then, for the distance measure, we have the following proposition.

Proposition 1. *The above-defined distance $d_p(A, B)$ for $p > 0$ satisfies the following properties:*

(*DP*1) $0 \le d_p(A, B) \le 1$;

(*DP*2) $d_p(A, B) = 0$ *if and only if* $A = B$;

(*DP*3) $d_p(A, B) = d_p(B, A)$;

(*DP*4) *If* $A \subseteq B \subseteq C$, *C is an SVNS in X, then* $d_p(A, C) \ge d_p(A, B)$ *and* $d_p(A, C) \ge d_p(B, C)$.

Proof. It is easy to see that $d_p(A, B)$ satisfies the properties (DP1)–(DP3). Therefore, we only prove (DP4). Let $A \subseteq B \subseteq C$, then, $T_A(x_i) \le T_B(x_i) \le T_C(x_i)$, $I_A(x_i) \ge I_B(x_i) \ge I_C(x_i)$, and $F_A(x_i) \ge F_B(x_i) \ge F_C(x_i)$ for every $x_i \in X$. Then, we obtain the following relations:

$$|T_A(x_i)-T_B(x_i)|^p \le |T_A(x_i)-T_C(x_i)|^p,\ |T_B(x_i)-T_C(x_i)|^p \le |T_A(x_i)-T_C(x_i)|^p,$$
$$|I_A(x_i)-I_B(x_i)|^p \le |I_A(x_i)-I_C(x_i)|^p,\ |I_B(x_i)-I_C(x_i)|^p \le |I_A(x_i)-I_C(x_i)|^p,$$
$$|F_A(x_i)-F_B(x_i)|^p \le |F_A(x_i)-F_C(x_i)|^p,\ |F_B(x_i)-F_C(x_i)|^p \le |F_A(x_i)-F_C(x_i)|^p.$$

Hence,

$$\begin{aligned}&|T_A(x_i)-T_B(x_i)|^p+|I_A(x_i)-I_B(x_i)|^p+|F_A(x_i)-F_B(x_i)|^p\\ &\le |T_A(x_i)-T_C(x_i)|^p+|I_A(x_i)-I_C(x_i)|^p+|F_A(x_i)-F_C(x_i)|^p,\end{aligned}$$

$$\begin{aligned}&|T_B(x_i)-T_C(x_i)|^p+|I_B(x_i)-I_C(x_i)|^p+|F_B(x_i)-F_C(x_i)|^p\\ &\le |T_A(x_i)-T_C(x_i)|^p+|I_A(x_i)-I_C(x_i)|^p+|F_A(x_i)-F_C(x_i)|^p.\end{aligned}$$

Combining the above inequalities with the above-defined distance formula (1), we can obtain $d_p(A, B) \le d_p(A, C)$ and $d_p(B, C) \le d_p(A, C)$ for $p > 0$.

Thus, the property (DP4) is satisfied.

This completes the proof. □

Note that similarity and distance (dissimilarity) measures are complementary: when the first increases, the second decreases. Normalized distance measure and similarity measure are dual concepts. Thus, $S(A, B) = 1 - d(A, B)$ and vice versa. The properties of distance measures below are complementary to those of similarity measures.

Proposition 2. *Let A and B be two SVNSs in a universe of discourse* $X = \{x_1, x_2, \ldots, x_n\}$; $S(A, B)$ *is called a single-valued neutrosophic similarity measure, which should satisfy the following properties:*

(*SP*1) $0 \le S(A, B) \le 1$;

(*SP*2) $S(A, B) = 1$ *if and only if* $A = B$;

(*SP*3) $S(A, B) = S(B, A)$;

(*SP*4) $S(A, C) \le S(A, B)$ *and* $S(A, C) \le S(B, C)$ *if* $A \subseteq B \subseteq C$ *for an SVNS C.*

Assume that there are two SVNSs $A = \{\langle x_i, T_A(x_i), I_A(x_i), F_A(x_i)\rangle | x_i \in X\}$ and $B = \{\langle x_i, T_B(x_i), I_B(x_i), F_B(x_i)\rangle | x_i \in X\}$ in a universe of discourse $X = \{x_1, x_2, \ldots, x_n\}$. Thus, according to the relationship between the distance and the similarity measure, we can obtain the following single-valued neutrosophic similarity measure:

$$\begin{aligned}S_1(A, B) &= 1 - d_p(A, B)\\ &= 1 - \left\{\frac{1}{3}\sum_{i=1}^{n} w_i\left[|T_A(x_i)-T_B(x_i)|^p+|I_A(x_i)-I_B(x_i)|^p+|F_A(x_i)-F_B(x_i)|^p\right]\right\}^{1/p}.\end{aligned} \tag{4}$$

Obviously, we can easily prove that $S_1(A, B)$ satisfied the properties (SP1)–(SP4) in Proposition 2 by the relationship between the distance and the similarity measure and the proof of Proposition 1, which is omitted here.

Furthermore, we can also propose another single-valued neutrosophic similarity measure:

$$S_2(A, B)=\frac{1-d_p(A, B)}{1+d_p(A, B)}$$
$$=\frac{1-\left\{\frac{1}{3}\sum_{i=1}^{n} w_i[|T_A(x_i)-T_B(x_i)|^p+|I_A(x_i)-I_B(x_i)|^p+|F_A(x_i)-F_B(x_i)|^p]\right\}^{1/p}}{1+\left\{\frac{1}{3}\sum_{i=1}^{n} w_i[|T_A(x_i)-T_B(x_i)|^p+|I_A(x_i)-I_B(x_i)|^p+|F_A(x_i)-F_B(x_i)|^p]\right\}^{1/p}}. \tag{5}$$

Then, the similarity measure $S_2(A, B)$ also satisfied the properties (SP1)–(SP4) in Proposition 2.

Proof. It is easy to see that $S_2(A, B)$ satisfies the properties (SP1)–(SP3). Therefore, we only prove the property (SP4).

As we obtain $d_p(A, B) \le d_p(A, C)$ and $d_p(B, C) \le d_p(A, C)$ for $p > 0$ from the property (DP4) in Proposition 1, there are $1 - d_p(A, B) \ge 1 - d_p(A, C)$, $1 - d_p(B, C) \ge 1 - d_p(A, C)$, $1 + d_p(A, B) \le 1 + d_p(A, C)$, and $1 + d_p(B, C) \le 1 + d_p(A, C)$. Then, there are the following inequalities:

$$\frac{1-d_p(A, B)}{1+d_p(A, B)}\ge\frac{1-d_p(A, C)}{1+d_p(A, C)} \text{ and } \frac{1-d_p(B, C)}{1+d_p(B, C)}\ge\frac{1-d_p(A, C)}{1+d_p(A, C)}.$$

Then, there are $S(A, C) \le S(A, B)$ and $S(A, C) \le S(B, C)$. Hence, the property (SP4) is satisfied.
This completes the proof. □

Example 1. Assume that we have the following three SVNSs in a universe of discourse $X = \{x_1, x_2\}$:
$A = \{<x_1, 0.1, 0.5, 0.6>, <x_2, 0.2, 0.5, 0.7>\}$,
$B = \{<x_1, 0.3, 0.4, 0.5>, <x_2, 0.5, 0.3, 0.4>\}$,
$C = \{<x_1, 0.6, 0.1, 0.2>, <x_2, 0.8, 0.1, 0.3>\}$.

Then, there are $A \subseteq B \subseteq C$, with $T_A(x_i) \le T_B(x_i) \le T_C(x_i)$, $I_A(x_i) \ge I_B(x_i) \ge I_C(x_i)$, and $F_A(x_i) \ge F_B(x_i) \ge F_C(x_i)$ for each x_i in $X = \{x_1, x_2\}$, and the weight vector $w = (0.5, 0.5)^T$.

By applying Eq. (4) (take $p = 1$), the similarity measures between the SVNSs are as follows:
$S_1(A, B) = 0.8$, $S_1(B, C) = 0.75$, and $S_1(A, C) = 0.55$.
Thus, $S_1(A, C) \le S_1(A, B)$ and $S_1(A, C) \le S_1(B, C)$.
When $p = 2$, the similarity measures between the SVNSs are as follows:
$S_1(A, B) = 0.784$, $S_1(B, C) = 0.7386$, and $S_1(A, C) = 0.5436$.
Hence, $S_1(A, C) \le S_1(A, B)$ and $S_1(A, C) \le S_1(B, C)$.

By applying Eq. (5) for $p = 1$, the similarity measures between the SVNSs are as follows:
$S_2(A, B) = 0.6667$, $S_2(B, C) = 0.6$, and $S_2(A, C) = 0.3793$.
Thus, $S_2(A, C) \le S(A, B)$ and $S_2(A, C) \le S_2(B, C)$.
When $p = 2$, the similarity measures between the SVNSs are as follows:
$S_2(A, B) = 0.6447$, $S_2(B, C) = 0.5855$, and $S_2(A, C) = 0.3732$.
Hence, $S_2(A, C) \le S_2(A, B)$ and $S_2(A, C) \le S_2(B, C)$.

4 Clustering Algorithm Based on the Similarity Measures of SVNSs

In this section, we can apply the proposed similarity measures of SVNSs to clustering analysis under a single-valued neutrosophic environment.

On the basis of the intuitionistic fuzzy clustering algorithm proposed by Zhang et al. [13] and Xu et al. [7], we first introduce the following definitions.

Definition 6. Assume that $A = (A_1, A_2,\ldots,A_m)$ is a set of SVNSs and $C = (s_{ij})_{m\times m}$ is a similarity matrix, where $s_{ij} = S_k(A_i, A_j)$ $(k = 1, 2)$ and $s_{ij} \in [0, 1]$ for $i, j = 1, 2,\ldots,m$, with $s_{ii} = 1$ for $i = 1, 2,\ldots,m$, and $s_{ij} = s_{ji}$ for $i, j = 1, 2,\ldots,m$.

Definition 7 ([7, 13]). Let $C = (s_{ij})_{m\times m}$ be a similarity matrix, if $C^2 = C \circ C = (\bar{s}_{ij})_{m\times m}$, then C^2 is called a composition matrix of C, where $\bar{s}_{ij} = \max_k\{\min(s_{ik}, s_{kj})\}$ for $i, j = 1, 2,\ldots,m$.

Definition 8 ([7, 13]). Let $C = (s_{ij})_{m\times m}$ be a similarity matrix, if $C^2 \subseteq C$, i.e., $\bar{s}_{ij} \le s_{ij}$ for $i, j = 1, 2,\ldots,m$, then C is called an equivalent similarity matrix.

Definition 9 ([7, 13]). Let $C = (s_{ij})_{m\times m}$ be a similarity matrix. Then, after finite time compositions of C:

$$C \to C^2 \to C^4 \to \ldots \to C^{2^k} \to \ldots, \tag{6}$$

there must exist a positive integer k such that $C^{2^k} = C^{2^{(k+1)}}$, then C^{2^k} is also an equivalent similarity matrix.

Definition 10 ([7, 13]). Let $C = (s_{ij})_{m\times m}$ be an equivalent similarity matrix. Then, $C_\lambda = (s_{ij}^\lambda)_{m\times m}$ is called the λ-cutting matrix of C, where

$$s_{ij}^\lambda = \begin{cases} 0, & s_{ij} < \lambda; \\ 1, & s_{ij} \ge \lambda \end{cases} \quad \text{for } i,\ j = 1, 2,\ldots,m, \tag{7}$$

and λ is the confidence level with $\lambda \in [0, 1]$.

Assume that $A = (A_1, A_2,\ldots,A_m)$ is a set of SVNSs, where $A_j = \{\langle x_i, T_{A_j}(x_i), I_{A_j}(x_i), F_{A_j}(x_i)\rangle \mid x_i \in X\}$ $(j = 1, 2,\ldots,m)$ in a universe of discourse $X = \{x_1, x_2,\ldots,x_n\}$ is an SVNS. Let w_i be the weight for each element x_i $(i = 1, 2,\ldots,n)$, with $w_i \in [0, 1]$, and $\sum_{i=1}^n w_i = 1$. Then, we can give the algorithm of clustering SVNSs as follows:

Step 1. By use of Eqs. (4) or (5), one can calculate the similarity measure degrees of SVNSs, and then construct a similarity matrix $C = (s_{ij})_{m\times m}$, where $s_{ij} = S_k(A_i, A_j)$ $(k = 1, 2)$ for $i, j = 1, 2,\ldots,m$.

Step 2. The process of building the composition matrices is repeated until it holds that

$$C \to C^2 \to C^4 \to \ldots \to C^{2^k} = C^{2^{(k+1)}},$$

which implies that C^{2^k} is an equivalent similarity matrix, which is denoted by $\bar{C} = (\bar{s}_{ij})_{m\times m}$.

Step 3. For the equivalent similarity matrix $\bar{C} = (\bar{s}_{ij})_{m\times m}$, we can construct a λ-cutting matrix $\bar{C}_\lambda = (\bar{s}_{ij}^\lambda)_{m\times m}$ of $\bar{C}$ by Eq. (7); if all the elements of the ith row or column in $\bar{C}_\lambda$ are the same as the corresponding elements of the jth row or column, we conceive object sets A_i and A_j are the same class.

5 Illustrative Example and Discussion

In this section, a real example adapted from Zhang et al. [13] is employed to demonstrate the application and effectiveness of the proposed clustering methods under a single-valued neutrosophic data environment.

A car market is going to classify five different cars of A_j $(j = 1, 2,\ldots,5)$. Every car has six evaluation factors (attributes): (i) x_1, fuel consumption; (ii) x_2, coefficient of friction; (iii) x_3, price; (iv) x_4, comfortable degree; (v) x_5, design; (vi) x_6, security coefficient. The characteristics of each car under the six attributes are represented by the form of SVNSs, and then the single-valued neutrosophic data are as follows:

$A_1 = \{<x_1, 0.3, 0.2, 0.5>, <x_2, 0.6, 0.3, 0.1>, <x_3, 0.4, 0.3, 0.3>, <x_4, 0.8, 0.1, 0.1>, <x_5, 0.1, 0.3, 0.6>, <x_6, 0.5, 0.2, 0.4>\}$,

$A_2 = \{\langle x_1, 0.6, 0.3, 0.3\rangle, \langle x_2, 0.5, 0.4, 0.2\rangle, \langle x_3, 0.6, 0.2, 0.1\rangle, \langle x_4, 0.7, 0.2, 0.1\rangle, \langle x_5, 0.3, 0.1, 0.6\rangle, \langle x_6, 0.4, 0.3, 0.3\rangle\}$,

$A_3 = \{\langle x_1, 0.4, 0.2, 0.4\rangle, \langle x_2, 0.8, 0.2, 0.1\rangle, \langle x_3, 0.5, 0.3, 0.1\rangle, \langle x_4, 0.6, 0.1, 0.2\rangle, \langle x_5, 0.4, 0.1, 0.5\rangle, \langle x_6, 0.3, 0.2, 0.2\rangle\}$,

$A_4 = \{\langle x_1, 0.2, 0.4, 0.4\rangle, \langle x_2, 0.4, 0.5, 0.1\rangle, \langle x_3, 0.9, 0.2, 0.0\rangle, \langle x_4, 0.8, 0.2, 0.1\rangle, \langle x_5, 0.2, 0.3, 0.5\rangle, \langle x_6, 0.7, 0.3, 0.1\rangle\}$,

$A_5 = \{\langle x_1, 0.5, 0.3, 0.2\rangle, \langle x_2, 0.3, 0.2, 0.6\rangle, \langle x_3, 0.6, 0.1, 0.3\rangle, \langle x_4, 0.7, 0.1, 0.1\rangle, \langle x_5, 0.6, 0.2, 0.2\rangle, \langle x_6, 0.5, 0.2, 0.3\rangle\}$.

If the weight vector of the attribute x_i ($i = 1, 2,\ldots,6$) is $w = (1/6, 1/6, 1/6, 1/6, 1/6, 1/6)^T$, then we utilize the two single-valued neutrosophic similarity measures to classify the five different cars of A_j ($j = 1, 2,\ldots,5$) by the single-valued neutrosophic clustering algorithms.

5.1 Clustering Analysis Using Eq. (4)

Step 1. Utilize the similarity measure formula (4) (take $p = 2$) to calculate the similarity measures between each pair of SVNSs A_i and A_j ($i, j = 1, 2, 3, 4, 5$) and construct the following similarity matrix:

$$C=\begin{bmatrix} 1 & 0.8528 & 0.8528 & 0.8085 & 0.7631 \\ 0.8528 & 1 & 0.8709 & 0.8317 & 0.8174 \\ 0.8528 & 0.8709 & 1 & 0.7853 & 0.7814 \\ 0.8085 & 0.8317 & 0.7853 & 1 & 0.7585 \\ 0.7631 & 0.8174 & 0.7814 & 0.7585 & 1 \end{bmatrix}.$$

Step 2. Obtain equivalent similarity matrices by limited time compositions of C:

$$C^2=\begin{bmatrix} 1 & 0.8528 & 0.8528 & 0.8317 & 0.8174 \\ 0.8528 & 1 & 0.8709 & 0.8317 & 0.8174 \\ 0.8528 & 0.8709 & 1 & 0.8317 & 0.8174 \\ 0.8317 & 0.8317 & 0.8317 & 1 & 0.8174 \\ 0.8174 & 0.8174 & 0.8174 & 0.8174 & 1 \end{bmatrix},$$

$$C^4=\begin{bmatrix} 1 & 0.8528 & 0.8528 & 0.8317 & 0.8174 \\ 0.8528 & 1 & 0.8709 & 0.8317 & 0.8174 \\ 0.8528 & 0.8709 & 1 & 0.8317 & 0.8174 \\ 0.8317 & 0.8317 & 0.8317 & 1 & 0.8174 \\ 0.8174 & 0.8174 & 0.8174 & 0.8174 & 1 \end{bmatrix}.$$

Obviously, $C^4 = C^2$ implies that C^2 is an equivalent similarity matrix, denoted by $\bar{C}$.

Step 3. When λ has different values, we can construct a λ-cutting matrix $\bar{C}_\lambda = (\bar{s}_{ij}^{\lambda})_{m\times m}$ of $\bar{C}$ by Eq. (7) and obtain different categories, which give the following discussion:

(i) If $0 \le \lambda \le 0.8174$, $\bar{C}_\lambda = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$, then the cars are the same category: $\{A_1, A_2, A_3, A_4, A_5\}$.

(ii) If $0.8174 < \lambda \le 0.8317$, $\bar{C}_\lambda = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$,

then the cars can be divided into two categories: $\{A_1, A_2, A_3, A_4\}$, $\{A_5\}$.

(iii) If $0.8317 < \lambda \le 0.8528$, $\bar{C}_\lambda = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$,

then the cars can be divided into three categories: $\{A_1, A_2, A_3\}$, $\{A_4\}$, $\{A_5\}$.

(iv) If $0.8528 < \lambda \le 0.8709$, $\bar{C}_\lambda = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$,

then the cars can be divided into four categories: $\{A_1\}$, $\{A_2, A_3\}$, $\{A_4\}$, $\{A_5\}$.

(v) If $0.8709 < \lambda \le 1$, $\bar{C}_\lambda = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$,

then the cars can be divided into five categories: $\{A_1\}$, $\{A_2\}$, $\{A_3\}$, $\{A_4\}$, $\{A_5\}$.

5.2 Clustering Analysis Using Eq. (5)

Step 1. Utilize the similarity measure formula (5) (take $p = 2$) to calculate the similarity measures between each pair of SVNSs A_i and A_j $(i, j = 1, 2, 3, 4, 5)$ and construct the following similarity matrix:

$$C = \begin{bmatrix} 1 & 0.7434 & 0.7434 & 0.6786 & 0.6170 \\ 0.7434 & 1 & 0.7713 & 0.7119 & 0.6912 \\ 0.7434 & 0.7713 & 1 & 0.6464 & 0.6413 \\ 0.6786 & 0.7119 & 0.6464 & 1 & 0.6109 \\ 0.6170 & 0.6912 & 0.6413 & 0.6109 & 1 \end{bmatrix}.$$

Step 2. Obtain equivalent similarity matrices by limited time compositions of C:

$$C^2 = \begin{bmatrix} 1 & 0.7434 & 0.7434 & 0.7119 & 0.6912 \\ 0.7434 & 1 & 0.7713 & 0.7119 & 0.6912 \\ 0.7434 & 0.7713 & 1 & 0.7119 & 0.6912 \\ 0.7119 & 0.7119 & 0.7119 & 1 & 0.6912 \\ 0.6912 & 0.6912 & 0.6912 & 0.6912 & 1 \end{bmatrix},$$

$$C^4 = \begin{bmatrix} 1 & 0.7434 & 0.7434 & 0.7119 & 0.6912 \\ 0.7434 & 1 & 0.7713 & 0.7119 & 0.6912 \\ 0.7434 & 0.7713 & 1 & 0.7119 & 0.6912 \\ 0.7119 & 0.7119 & 0.7119 & 1 & 0.6912 \\ 0.6912 & 0.6912 & 0.6912 & 0.6912 & 1 \end{bmatrix}.$$

Obviously, $C^4 = C^2$ implies that C^2 is an equivalent similarity matrix, denoted by $\overline{C}$.

Step 3. When λ has different values, we can construct a λ-cutting matrix $\overline{C}_\lambda = (\overline{s}_{ij}^{\lambda})_{m\times m}$ of $\overline{C}$ by Eq. (7) and can obtain different categories, which make the following discussion:

(i) If $0 \le \lambda \le 0.6912$, $\overline{C}_\lambda = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$,

then the cars are the same category: $\{A_1, A_2, A_3, A_4, A_5\}$.

(ii) If $0.6912 < \lambda \le 0.7119$, $\overline{C}_\lambda = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$,

then the cars can be divided into two categories: $\{A_1, A_2, A_3, A_4\}$, $\{A_5\}$.

(iii) If $0.7119 < \lambda \le 0.7434$, $\overline{C}_\lambda = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$,

then the cars can be divided into three categories: $\{A_1, A_2, A_3\}$, $\{A_4\}$, $\{A_5\}$.

(iv) If $0.7434 < \lambda \le 0.7713$, $\overline{C}_\lambda = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$, then the cars can be divided into four categories: $\{A_1\}$, $\{A_2, A_3\}$, $\{A_4\}$, $\{A_5\}$.

(v) If $0.7713 < \lambda \le 1$, $C_\lambda = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$,

then the cars can be divided into five categories: $\{A_1\}$, $\{A_2\}$, $\{A_3\}$, $\{A_4\}$, $\{A_5\}$.

5.3 Discussion

From the above clustering results, we know the two similarity measures can be applied to clustering SVNSs and these clustering results using the two similarity measures are the same. However, the literature [13] obtained three situations by the clustering algorithm based on the similar measure of IFSs; however, we can obtain five situations by the clustering algorithm based on the proposed similarity measures of SVNSs. Hence, we can see that the clustering algorithms based on the two similarity measures of SVNSs have better accuracy in clustering problems.

As mentioned above, the single-valued neutrosophic information is a generalization of intuitionistic fuzzy information, and intuitionistic fuzzy information is a further generalization of fuzzy information. On the one hand, an SVNS is an instance of a neutrosophic set, which gives us an additional possibility to represent uncertain, imprecise, incomplete, and inconsistent information that exist in the real world. It can describe and handle indeterminate information and inconsistent information. However, the connector in the fuzzy set is defined with respect to T, i.e., membership only; hence, the information of indeterminacy and non-membership is lost. The connectors in the IFS are defined with respect to T and F, i.e., membership and non-membership only; hence, the indeterminacy is what is left from 1, and then the IFS can only handle incomplete information but not the indeterminate information and inconsistent information. While in the SVNSs, its truth-membership, indeterminacy-membership, and falsity-membership are represented independently, and then they can be defined with respect to any of them (no restriction). Thus, the notion of SVNSs is more general. On the other hand, the clustering analysis under a single-valued neutrosophic environment is suitable for capturing imprecise, uncertain, and inconsistent information in clustering the data. Thus, the clustering algorithm based on the similarity measures of SVNSs not only can cluster the single-valued neutrosophic information but also can cluster the intuitionistic fuzzy information and the fuzzy information. Obviously, the proposed single-valued neutrosophic clustering algorithm is the extension of both fuzzy clustering algorithm and intuitionistic fuzzy clustering algorithm. Therefore, compared with the intuitionistic fuzzy clustering algorithm and the fuzzy clustering algorithm, the single-valued neutrosophic clustering algorithm is more general. Furthermore, when we encounter some situations that are represented by indeterminate information and inconsistent information, the single-valued neutrosophic clustering algorithm can demonstrate its great superiority in clustering those single-valued neutrosophic data.

6 Conclusion

This article introduced a generalized single-valued neutrosophic weighted distance measure and presented two distance-based similarity measures in a single-valued neutrosophic setting. Then, a single-valued neutrosophic clustering algorithm was established on the basis of the two similarity measures. Finally, an illustrative example was given to demonstrate the application and effectiveness of the single-valued neutrosophic clustering methods. The clustering results have shown that the single-valued neutrosophic clustering algorithm is more general than the intuitionistic fuzzy clustering algorithm and the fuzzy clustering algorithm. Furthermore, in the situations that are represented by indeterminate information and inconsistent information, the single-valued neutrosophic clustering algorithm can demonstrate its great superiority in clustering those single-valued neutrosophic data, as the SVNSs are a powerful tool to deal with uncertain, imprecise, incomplete, and inconsistent information. In the future, the developed clustering algorithm will be extended to clustering problems of interval-valued neutrosophic sets and further applied to many areas such as information retrieval, investment decision making, and data mining.

Received November 15, 2013; previously published online March 7, 2014.

Bibliography

[1] K. Atanassov, Intuitionistic fuzzy sets, *Fuzzy Sets Syst.* **20** (1986), 87–96.
[2] K. Atanassov and G. Gargov, Interval valued intuitionistic fuzzy sets, *Fuzzy Sets Syst.* **31** (1989), 343–349.
[3] E. H. Ruspini, A new approach to clustering, *Inform. Control* **15** (1969), 22–32.
[4] F. Smarandache, *A unifying field in logics. Neutrosophy: neutrosophic probability, set and logic*, American Research Press, Rehoboth, 1999.
[5] H. Wang, F. Smarandache, Y. Q. Zhang and R. Sunderraman, *Interval neutrosophic sets and logic: theory and applications in computing*, Hexis, Phoenix, AZ, 2005.
[6] H. Wang, F. Smarandache, Y. Q. Zhang and R. Sunderraman, Single valued neutrosophic sets, *Multispace Multistruc.* **4** (2010), 410–413.
[7] Z. S. Xu, J. Chen and J. J. Wu, Clustering algorithm for intuitionistic fuzzy sets, *Inform. Sci.* **19** (2008), 3775–3790.
[8] J. Ye, Multicriteria decision-making method using the correlation coefficient under single-valued neutrosophic environment, *Int. J. General Syst.* **42** (2013), 386–394.
[9] J. Ye, Single valued neutrosophic cross-entropy for multicriteria decision making problems, *Appl. Math. Model.* **38** (2014), 1170–1175.
[10] J. Ye, A multicriteria decision-making method using aggregation operators for simplified neutrosophic sets, *J. Intell. Fuzzy Syst.* (2013), doi: 10.3233/IFS-130916.
[11] J. Ye, Similarity measures between interval neutrosophic sets and their applications in multicriteria decision-making, *J. Intell. Fuzzy Syst.* **26** (2014), 165–172.
[12] L. A. Zadeh, Fuzzy sets, *Inform. Control* **8** (1965), 338–353.
[13] H. M. Zhang, Z. S. Xu and Q. Chen, Clustering method of intuitionistic fuzzy sets, *Control Decision* **22** (2007), 882–888.