Advances in Fuzzy Systems

Fuzzy Methods for Data Analysis

Guest Editors: Ferdinando Di Martino, Irina Perfilieva, and Salvatore Sessa

Fuzzy Methods for Data Analysis

Advances in Fuzzy Systems

Fuzzy Methods for Data Analysis

Guest Editors: Ferdinando Di Martino, Irina Perfilieva, and Salvatore Sessa

Editorial Board

Contents

Hindawi Publishing Corporation
Advances in Fuzzy Systems
Volume 2015, Article ID 957856, 1 page
http://dx.doi.org/10.1155/2015/957856

Editorial

Fuzzy Methods for Data Analysis

Ferdinando Di Martino,[1] Irina Perfilieva,[2] and Salvatore Sessa[1,3]

[1] *Dipartimento di Costruzioni e Metodi Matematici in Architettura, Università degli Studi di Napoli Federico II, Via Monteoliveto 3, 80134 Napoli, Italy*
[2] *Centre of Excellence IT4Innovations, Division of the University of Ostrava, Institute for Research and Applications of Fuzzy Modeling, 30. Dubna 22, 701 00 Ostrava 1, Czech Republic*
[3] *Centro Interdipartimentale di Ricerca per l'Analisi e la Progettazione Urbana Luigi Pisciotti, Via Toledo 402, 80134 Napoli, Italy*

Correspondence should be addressed to Ferdinando Di Martino; fdimarti@unina.it

Received 9 December 2014; Accepted 9 December 2014

This special issue contains various papers where the topic of data analysis is studied with various fuzzy tools. The following is a short summary of what has been considered.

The paper of R. Li and G. Li deals with 20 samples and 44 samples of terracotta warriors and horses of Qin Shi Huang's Mausoleum 20 samples of clay near Qin's Mausoleum, and 2 samples of Yaozhou porcelain bodies, and in each of them the contents of 32 elements are determined by using the Neutron Activation Analysis whose data are analyzed with a fuzzy cluster algorithm: the conclusion is that the terracotta of the above artefacts originates from different near sites.

The paper of J. Davis et al. deals with fuzzy probability of screening a survey data across relevant criteria for selecting suppliers based on fuzzy expected values that are influenced by a delivery performance.

The paper of A. Chaudhuri deals with an intuitionistic fuzzy possibilistic *C*-means algorithm which determines membership values of objects to each cluster as intervals instead of single numerical values, so classifying datasets with labelled patterns.

The paper of M. Burda provides a correct definition of lift, leverage, and conviction measures for fuzzy association rules, studying the related properties.

The paper of P. Shanmugasundaram et al. deals with a revised intuitionistic fuzzy max-min average composition method in order to select well-trained students on the basis of their skills by the recruiters by means of intuitionistic fuzzy soft matrices.

We hope that this issue evokes deep interest among fuzzy authors and not alone.

Ferdinando Di Martino
Irina G. Perfilieva
Salvatore Sessa

Hindawi Publishing Corporation
Advances in Fuzzy Systems
Volume 2015, Article ID 238237, 17 pages
http://dx.doi.org/10.1155/2015/238237

Research Article

Intuitionistic Fuzzy Possibilistic C Means Clustering Algorithms

Arindam Chaudhuri

Samsung Research & Development Institute, Noida 201304, India

Correspondence should be addressed to Arindam Chaudhuri; arindamphdthesis@gmail.com

Received 24 August 2014; Accepted 4 October 2014

Academic Editor: Ferdinando Di Martino

Intuitionistic fuzzy sets (IFSs) provide mathematical framework based on fuzzy sets to describe vagueness in data. It finds interesting and promising applications in different domains. Here, we develop an intuitionistic fuzzy possibilistic C means (IFPCM) algorithm to cluster IFSs by hybridizing concepts of FPCM, IFSs, and distance measures. IFPCM resolves inherent problems encountered with information regarding membership values of objects to each cluster by generalizing membership and nonmembership with hesitancy degree. The algorithm is extended for clustering interval valued intuitionistic fuzzy sets (IVIFSs) leading to interval valued intuitionistic fuzzy possibilistic C means (IVIFPCM). The clustering algorithm has membership and nonmembership degrees as intervals. Information regarding membership and typicality degrees of samples to all clusters is given by algorithm. The experiments are performed on both real and simulated datasets. It generates valuable information and produces overlapped clusters with different membership degrees. It takes into account inherent uncertainty in information captured by IFSs. Some advantages of algorithms are simplicity, flexibility, and low computational complexity. The algorithm is evaluated through cluster validity measures. The clustering accuracy of algorithm is investigated by classification datasets with labeled patterns. The algorithm maintains appreciable performance compared to other methods in terms of pureness ratio.

1. Introduction

Clustering algorithms [1, 2] form an integral part of computational intelligence and pattern recognition research. Clustering analysis is commonly used as an important tool to classify collection of objects into homogeneous groups, such that objects within a given group are similar to each other whereas objects within different groups are dissimilar to each other. The concept is based on the notion of similarity, which is a basic component of intelligence and ubiquitous to scientific endeavor. Clustering finds numerous applications [3] across a variety of disciplines such as taxonomy, image processing, information retrieval, data mining, pattern recognition, microbiology, archaeology and geographical analysis, and so forth. It is an exploratory tool for deducing the nature of data by providing labels to individual objects that describe how the data separate into groups. It has improved the performance of other systems by separating the problem domain into manageable subgroups [4]. Often researchers are confronted with the challenging datasets that are large and unlabeled. There are many methods available in exploratory data analysis [5, 6] by which researchers can elucidate these data.

Clustering an unlabeled dataset $X = \{x_1, \ldots, x_n\} \subset \mathfrak{R}^p$ is partitioning of X into $1 < c < n$ subgroups such that each subgroup represents natural substructure in X. This is done by assigning labels to vectors in X and hence to objects generating X. A c partition of is a set of cn values that can be conveniently represented as $c \times n$ matrix $= [u_{ik}]$. There are generally three sets of partition matrices [7, 8]:

$$M_{pcn} = \left\{U \in \mathfrak{R}^{cn} : 0 \le u_{ik} \le 1 \ \forall i, k;\ \forall k\ \exists i \ni u_{ik} > 0\right\} \quad (1)$$

$$M_{fcn} = \left\{U \in M_{pcn} : \sum_{i=1}^{c} u_{ik} = 1\ \forall k;\ \sum_{i=1}^{c} u_{ik} > 0\ \forall i\right\} \quad (2)$$

$$M_{hcn} = \left\{U \in M_{fcn} : u_{ik} = 0\ \vee 1\ \forall i\ \wedge k\right\}. \quad (3)$$

The matrix M_{pcn} in (1) has the property that for any k there exists at least an index i such that u_{ik} is greater than 0. The matrix M_{fcn} in (2) states that if $\sum_{i=1}^{c} u_{ik}$ is equal to 1 for any k, it is obvious that $\sum_{i=1}^{c} u_{ik}$ is greater than 0.

The matrix M_{hcn} in (3) is formed by boolean matrices as a subset of matrix M_{fcn} in (2). The equations (1), (2), and (3) thus define the sets of possibilistic, fuzzy, or probabilistic and crisp c partitions of X, respectively. Hence, there are thus four kinds of label vectors, but fuzzy and probabilistic label vectors are mathematically identical having entries between 0 and 1 that sum to 1 over each column. The reason these matrices are called partitions follows from the interpretation of their entries. If U is crisp or fuzzy, u_{ik} is taken as a membership of $\mathbf{x}_k$ in ith partitioning fuzzy subset or cluster of X. If U in M_{fcn} is probabilistic, u_{ik} is the posterior probability $\text{Prob}(i/\mathbf{x}_k)$. If U in M_{pcn} is possibilistic, it has entries between 0 and 1 that do not necessarily sum to 1 over any column. In this case, u_{ik} is taken as the possibility that $\mathbf{x}_k$ belongs to class i. An alternate interpretation of possibility u_{ik} is that it measures the typicality of $\mathbf{x}_k$ to cluster i. It is observed that $M_{hcn} \subset M_{fcn} \subset M_{pcn}$. A clustering algorithm $\mathbf{C}$ finds $U \in M_{hcn}(M_{fcn}, M_{pcn})$ which best explains and represents an unknown structure in X with respect to the model that defines $\mathbf{C}$. For U in M_{fcn}, $c = 1$ is represented uniquely by the hard 1 partition, $\mathbf{1}_n = \underbrace{[1 \cdots 1]}_{n \text{ times}}$ which unequivocally assigns all n objects to a single cluster, and $c = n$ is represented uniquely by $U = I_n$, the $n \times n$ identity matrix up to a permutation of columns. In this case, each object is in its own singleton cluster. Choosing $c = 1$ or $c = n$ rejects the hypothesis that X contains clusters.

In the last few decades, variety of clustering techniques [3, 5, 6, 9, 10] has been developed to classify data. Clustering techniques are broadly divided into hierarchical and partition methods. Hierarchical clustering [5] generates hierarchical tree of clusters called dendrogram which can be either divisive or agglomerative [3]. The former is a top-down splitting technique which starts with all objects in one cluster and forms hierarchy by dividing objects into smaller clusters in an iterative procedure until the desired number of clusters is achieved or considered objects which is constituted as unique cluster. The latter starts by considering each object as cluster, followed by comparing them amongst themselves using distance measure. The clusters with smaller distance are considered as constituting unique group and then merged. The merging procedure is repeated until the desirable number of clusters is achieved or only one cluster is left with all considered objects. Partition clustering method gives single c partition of objects, with c being the predefined number of clusters [11]. One of the most widely used partition clustering algorithms is fuzzy C means (FCM). FCM is a combination of k means clustering algorithm and fuzzy logic [1, 7]. It works iteratively in which the desired number of clusters c and initial seeds are predefined. FCM algorithm assigns memberships to $\mathbf{x}_k$ which are inversely proportional to relative distance of $\mathbf{x}_k$ to c point prototypes $\{\mathbf{v}_i\}$ that are cluster centers. For $c = 2$, if $\mathbf{x}_k$ is equidistant from two prototypes, the membership of $\mathbf{x}_k$ in each cluster will be the same regardless of absolute value of the distance of $\mathbf{x}_k$ from two centroids as well as from other points. The problem this creates is noise points which are far but equidistant from central structure of two clusters that can never be given equal membership, when it seems far more natural that such points are given very low or no membership in either cluster. This problem was overcome by Krishnapuram and Keller [8], who proposed possibilistic C means (PCM) which relaxes column sum constraint in (2) so that sum of each column satisfies the constraint $0 < \sum_{i=1}^{c} u_{ik} \leq c$. In other words, each element of kth column can be between 0 and 1, as long as at least one of them is positive. They suggested that value should be interpreted as typicality of x_k relative to cluster i. They interpreted each row of U as possibility distribution over X. The objective function of PCM algorithm sometimes helps to identify outliers, that is, noise points. However, Barni et al. [12] pointed that PCM pays price for its freedom to ignore noise points such that it is very sensitive to initializations and sometimes generates coincident clusters. Moreover, typicality can be very sensitive to the choice of additional parameters needed by PCM algorithm. The coincident cluster problem of PCM algorithm was avoided by two possibilistic fuzzy clustering algorithms proposed by Timm et al. [13–15]. They modified PCM objective function by adding an inverse function of distances between the cluster centers. This term acts in repulsive nature and avoids coincident clusters. In [13, 14], Timm et al. used the same concept to modify objective function as used by Gustafson and Kessel [16] clustering algorithm. These algorithms exploit benefits of both fuzzy and possibilistic clustering. Pal et al. [17] justified the need for both possibility, that is, typicality and membership values, and proposed a model and corresponding algorithm called fuzzy possibilistic C means (FPCM). This algorithm normalizes possibility values, so that the sum of possibilities of all data points in a cluster is 1. Although FPCM is much less prone to errors encountered by FCM and PCM, possibility values are very small when size of dataset increases.

The notion of intuitionistic fuzzy set (IFS) coined by Atanassov [22] for fuzzy set generalizations has interesting and useful applications in different domains such as logic programming, decision making problems, and medical diagnostics [23–26]. This generalization presents degrees of membership and nonmembership with a degree of hesitancy. Thus, knowledge and semantic representation become more meaningful and applicable [27, 28]. Sometimes it is not appropriate to assume that membership and nonmembership degree of an object are exactly defined [29], but value ranges or value intervals can be assigned. In such cases, IFS can be generalized and interval valued intuitionistic fuzzy set (IVIFS) [29] can be defined whose components are intervals rather than exact numbers. IFSs and IVIFSs have been found to be very useful to describe and deal with vague and uncertain data [28, 30]. With this motivation, it is desirable to develop some practical approaches to clustering IFSs and IVIFSs. Intuitionistic fuzzy similarity matrix was defined by [31] and thereby intuitionistic fuzzy equivalence matrix was developed. The work in [31] gave an approach to transform intuitionistic fuzzy similarity matrices into intuitionistic fuzzy equivalence matrices, based on which a procedure for clustering intuitionistic fuzzy sets was proposed. Some methods for calculating association coefficients of IFSs or IVIFSs and corresponding clustering algorithm were introduced by [32]. The algorithm used derived association coefficients of IFSs or IVIFSs to construct an association matrix and utilized the procedure to transform it into an equivalent association matrix. Reference [33] introduced an intuitionistic fuzzy hierarchical algorithm for clustering IFSs which is based on traditional hierarchical clustering procedure and intuitionistic fuzzy aggregation operator. These algorithms cannot provide information about membership degrees of objects to each cluster.

In this work, an intuitionistic fuzzy possibilistic C means (IFPCM) algorithm to cluster IFSs is developed. IFPCM is obtained by applying IFSs to FPCM which is a known clustering method based on basic distance measures between IFSs [34, 35]. At each stage of the algorithm seeds are modified and for each IFS membership and typicality degrees to each of the clusters are estimated. The algorithm ends when all given IFSs are clustered according to estimated membership and typicality degrees. It overcomes the inherent problems encountered with information regarding membership values of objects to each cluster by generalizing membership and nonmembership with hesitancy degree. The algorithm is then extended to interval valued intuitionistic fuzzy possibilistic C means (IVIFPCM) for clustering IVIFSs. The algorithms are illustrated through conducting experiments on different datasets. The evaluation of the algorithm is performed through cluster validity measures. The clustering accuracy of the algorithm is determined by classification datasets with labeled patterns. IFPCM algorithm is simple and flexible in nature and provides information about membership and typicality degrees of samples to all clusters with low computational complexity.

This paper is organized as follows. In the next section, the concepts of IFSs and IVIFSs are defined. FPCM clustering algorithm is given in Section 3. The next section presents IFPCM clustering algorithms for IFSs and IVIFSs, respectively. The experimental results on both real world and simulated datasets are illustrated in Section 5. Finally, in Section 6 conclusions are given.

2. Intuitionistic Fuzzy Sets and Interval Valued Intuitionistic Fuzzy Sets

In this section, we present some basic definitions associated with IFSs and IVIFSs.

Definition 1. Considering X as universe of discourse [22], then IFS is defined as

$$V = \{\langle x, \mu_V(x), \nu_V(x)\rangle \mid x \in X\}. \tag{4}$$

In (4) $\mu_V(x)$ and $\nu_V(x)$ are the membership and nonmembership degrees, respectively, satisfying the following constraints:

$$\begin{aligned} &\mu_V : X \longrightarrow [0,1], \quad x \in X \longrightarrow \mu_V(x) \in [0,1], \\ &\nu_V : X \longrightarrow [0,1], \quad x \in X \longrightarrow \nu_V(x) \in [0,1]. \end{aligned} \tag{5}$$

Equation (5) is subject to the condition that $\mu_V(x)+\nu_V(x) \le 1$, $\forall x \in X$.

Definition 2. For each IFS V in X, if $\pi_V(x) = 1 - \mu_V(x) - \nu_V(x)$, then $\pi_V(x)$ is called hesitation degree (or intuitionistic index) [36] of x to V. Obviously $\pi_V(x)$ is specified in the range $0 \le \pi_V(x) \le 1$; especially if $\pi_V(x) = 0$ $\forall x \in X$, then IFS V is reduced to fuzzy set. If $\pi_V(x)$ and $\nu_V(x)$ have 0 values such that $\mu_V(x) = \nu_V(x) = 0$ $\forall x \in X$, then IFS V is completely intuitionistic.

Considering the fact that the elements x_i; $i = 1, \ldots, n$ in universe X have different importance, let us assume $w = (w_1, \ldots, w_n)$ should be the weight vector of x_i; $i = 1, \ldots, n$ with

$$w_i \ge 0, \qquad \sum_{i=1}^{n} w_i = 1. \tag{6}$$

Xu [37] defined the following weighted Euclidean distance between IFSs V and W:

$$\begin{aligned} D_\alpha(V, W) = \Bigg(\frac{1}{2} \sum_{i=1}^{n} w_i \Big(&(\mu_V(x_i) - \mu_W(x_i))^2 \\ &+ (\nu_V(x_i) - \nu_W(x_i))^2 \\ &+ (\pi_V(x_i) - \pi_W(x_i))^2 \Big) \Bigg)^{1/2}. \end{aligned} \tag{7}$$

In particular, if $w = (1/n, \ldots, 1/n)$, then (7) is reduced to normalized Euclidean distance [34] which is defined as follows:

$$\begin{aligned} D_\alpha(V, W) &= \Bigg(\frac{1}{2n} \sum_{i=1}^{n} \Big((\mu_V(x_i) - \mu_W(x_i))^2 + (\nu_V(x_i) - \nu_W(x_i))^2 \\ &\quad + (\pi_V(x_i) - \pi_W(x_i))^2 \Big) \Bigg)^{1/2}. \end{aligned} \tag{8}$$

Atanassov and Gargov [29] pointed out that sometimes it is not appropriate to assume that membership and nonmembership degrees of the element are exactly defined but value ranges or value intervals can be given. In this context, Atanassov and Gargov [29] extended IFS and introduced the concept of IVIFS, which is characterized by a membership degree and a nonmembership degree, whose values are intervals rather than exact numbers.

Definition 3. An IVIFS $\widetilde{V}$ over X is an object having the following form [29]:

$$\widetilde{V} = \{\langle x, \widetilde{\mu}_{\widetilde{V}}(x), \widetilde{\nu}_{\widetilde{V}}(x)\rangle \mid x \in X\}. \tag{9}$$

Here, $\widetilde{\mu}_{\widetilde{V}}(x) = [\widetilde{\mu}^L_{\widetilde{V}}(x), \widetilde{\mu}^U_{\widetilde{V}}(x)] \subset [0,1]$ and $\widetilde{\nu}_{\widetilde{V}}(x) = [\widetilde{\nu}^L_{\widetilde{V}}(x), \widetilde{\nu}^U_{\widetilde{V}}(x)] \subset [0,1]$ are intervals $\widetilde{\mu}^L_{\widetilde{V}}(x) = \inf \widetilde{\mu}_{\widetilde{V}}(x)$, $\widetilde{\mu}^U_{\widetilde{V}}(x) = \sup \widetilde{\mu}_{\widetilde{V}}(x)$, $\widetilde{\nu}^L_{\widetilde{V}}(x) = \inf \widetilde{\nu}_{\widetilde{V}}(x)$, $\widetilde{\nu}^U_{\widetilde{V}}(x) = \sup \widetilde{\nu}_{\widetilde{V}}(x)$, and $\widetilde{\mu}^U_{\widetilde{V}}(x) + \widetilde{\nu}^U_{\widetilde{V}}(x) \le 1, \forall x \in X$ and $\widetilde{\pi}_{\widetilde{V}}(x) = [\widetilde{\pi}^L_{\widetilde{V}}(x), \widetilde{\pi}^U_{\widetilde{V}}(x)]$, where

$\pi_{\widetilde{V}}^{L}(x) = 1-\mu_{\widetilde{V}}^{U}(x)-\nu_{\widetilde{V}}^{U}(x), \pi_{\widetilde{V}}^{U}(x) = 1-\mu_{\widetilde{V}}^{L}(x)-\nu_{\widetilde{V}}^{L}(x), \forall x \in X$. In particular, if $\mu_{\widetilde{V}}(x) = \widetilde{\mu}_{\widetilde{V}}^{L}(x) = \widetilde{\mu}_{\widetilde{V}}^{U}(x)$ and $\nu_{\widetilde{V}}(x) = \widetilde{\nu}_{\widetilde{V}}^{L}(x) = \widetilde{\nu}_{\widetilde{V}}^{U}(x)$, then $\widetilde{V}$ is reduced to an IFS.

Now we extend the weighted Euclidean distance measure given in (7) to IVIFS theory:

$$\begin{aligned} D_{\gamma}\left(\widetilde{V},\widetilde{W}\right) &= \left(\frac{1}{4}\sum_{l=1}^{n} w_l \left(\left(\widetilde{\mu}_{\widetilde{V}}^{L}(x_l) - \widetilde{\mu}_{\widetilde{W}}^{L}(x_l)\right)^2 \right.\right. \\ &\quad + \left(\widetilde{\mu}_{\widetilde{V}}^{U}(x_l) - \widetilde{\mu}_{\widetilde{W}}^{U}(x_l)\right)^2 + \left(\widetilde{\nu}_{\widetilde{V}}^{L}(x_l) - \widetilde{\nu}_{\widetilde{W}}^{L}(x_l)\right)^2 \\ &\quad + \left(\widetilde{\nu}_{\widetilde{V}}^{U}(x_l) - \widetilde{\nu}_{\widetilde{W}}^{U}(x_l)\right)^2 \\ &\quad + \left(\widetilde{\pi}_{\widetilde{V}}^{L}(x_l) - \widetilde{\pi}_{\widetilde{W}}^{L}(x_l)\right)^2 \\ &\quad \left.\left. + \left(\widetilde{\pi}_{\widetilde{V}}^{U}(x_l) - \widetilde{\pi}_{\widetilde{W}}^{U}(x_l)\right)^2 \right) \right)^{1/2}. \end{aligned} \tag{10}$$

Particularly, if $w = (1/n, \ldots, 1/n)$, then (10) is reduced to normalized Euclidean distance which is given as follows:

$$\begin{aligned} D_{\gamma}\left(\widetilde{V},\widetilde{W}\right) &= \left(\frac{1}{4n}\sum_{l=1}^{n} w_l \left(\left(\widetilde{\mu}_{\widetilde{V}}^{L}(x_l) - \widetilde{\mu}_{\widetilde{W}}^{L}(x_l)\right)^2 \right.\right. \\ &\quad + \left(\widetilde{\mu}_{\widetilde{V}}^{U}(x_l) - \widetilde{\mu}_{\widetilde{W}}^{U}(x_l)\right)^2 + \left(\widetilde{\nu}_{\widetilde{V}}^{L}(x_l) - \widetilde{\nu}_{\widetilde{W}}^{L}(x_l)\right)^2 \\ &\quad + \left(\widetilde{\nu}_{\widetilde{V}}^{U}(x_l) - \widetilde{\nu}_{\widetilde{W}}^{U}(x_l)\right)^2 + \left(\widetilde{\pi}_{\widetilde{V}}^{L}(x_l) - \widetilde{\pi}_{\widetilde{W}}^{L}(x_l)\right)^2 \\ &\quad \left.\left. + \left(\widetilde{\pi}_{\widetilde{V}}^{U}(x_l) - \widetilde{\pi}_{\widetilde{W}}^{U}(x_l)\right)^2 \right) \right)^{1/2}. \end{aligned} \tag{11}$$

3. Fuzzy Possibilistic C Means Clustering Algorithm

This section illustrates FPCM clustering algorithm proposed by Pal et al. [17] in 1997 to exploit the benefits of fuzzy and possibilistic modeling while circumventing their weaknesses. To correctly interpret the data substructure, FPCM clustering uses both memberships (relative typicality) and possibilities (absolute typicality). When we want to crisply label a data point, membership is a plausible choice as it is natural to assign a point to cluster whose prototype is closest to the point. On the other hand, while estimating the centroids, typicality is an important means for alleviating the undesirable effects of outliers. Here, the number of clusters is fixed a priori to a default value considering the dataset used in the application such that it is completely data driven. Generally it is advisable to avoid trivial clusters which may be either too large or small.

FPCM extends FCM clustering algorithm [17] by normalizing possibility values so that sum of possibilities of all data points in a cluster is 1. Although FPCM is much less prone to the problems of both FCM and PCM, the possibility values are very small when size of the dataset increases. Analogous to FCM clustering algorithm, the membership term in FPCM is a function of data point and all centroids. The typicality term in FPCM is a function of data point and cluster prototype alone. That is, the membership term is influenced by the positions of all cluster centers whereas typicality term is affected by only one. Incorporating the abovementioned facets the FPCM model is defined by the following optimization problem [17]:

$$\underbrace{\min}_{(U,T,V)} \left\{ J_{m,\eta}(U,T,V;X) = \sum_{k=1}^{p}\sum_{i=1}^{c} \left(u_{ik}^{m} + t_{ik}^{\eta}\right) D_{ik}^{2} \right\} \tag{12}$$

subject to $m > 1, \quad \eta > 1, \quad 0 < u_{ik}, \quad t_{ik} < 1$

$$\sum_{i=1}^{c} u_{ik} = 1 \quad \forall k \text{ i.e., } U \in M_{fcn} \tag{13}$$

$$\sum_{k=1}^{p} t_{ik} = 1 \quad \forall i \text{ i.e., } T^{t} \in M_{fnc}. \tag{14}$$

The transpose of admissible T's is member of set M_{fnc}. T is viewed as a typicality assignment of n objects to c clusters. The possibilistic term $\sum_{k=1}^{p}\sum_{i=1}^{c} t_{ik}^{\eta} D_{ik}^{2}$ distributes $\{t_{ik}\}$ with respect to all n data points, but not with respect to all c clusters. Under the usual conditions placed on c-means optimization problems, the first order necessary conditions for extrema of $J_{m,\eta}$ are stated in terms of the following theorem.

Theorem FPCM (see [17]). *If $D_{ik} = \|x_k - \mathbf{v}_i\| > 0$ $\forall i$ and $k, m, \eta > 1$ and X contains at least c distinct data points, then $(U, T^{t}, \mathbf{V}) \in M_{fcn} \times M_{fcn} \times \mathfrak{R}^{p}$ may minimize $J_{m,\eta}$ only if*

$$u_{ik} = \left(\sum_{j=1}^{c} \left(\frac{D_{ik}}{D_{jk}} \right)^{2/(m-1)} \right)^{-1}; \quad 1 \le i \le c, \ 1 \le k \le n, \tag{15}$$

$$t_{ik} = \left(\sum_{j=1}^{n} \left(\frac{D_{ik}}{D_{ij}} \right)^{2/(\eta-1)} \right)^{-1}; \quad 1 \le i \le c, \ 1 \le k \le n, \tag{16}$$

$$\mathbf{v}_i = \frac{\sum_{k=1}^{n} \left(u_{ik}^{m} + t_{ik}^{\eta}\right) \mathbf{x}_k}{\sum_{k=1}^{n} \left(u_{ik}^{m} + t_{ik}^{\eta}\right)}; \quad 1 \le i \le c. \tag{17}$$

The proof of the above theorem follows from [38]. FPCM has the same type of singularity as FCM. FPCM does not suffer from the sensitivity problem that PCM seems to exhibit. Unfortunately, when the number of data points is large, the typicality values will be very small. Thus, after FPCM-AO algorithm [38] for approximating solutions to (12) based on iteration through (17) terminates, the typicality values may need to be scaled up. Conceptually, this is not different than scaling typicality as is done in PCM. While scaling seems to solve the small value problem which is caused by row sum constraint on T, the scaled values do not possess any additional information about points in the data. Thus scaling $\{t_{ik}\}$ is an artificial fix for a mathematical drawback of FPCM.

4. Intuitionistic Fuzzy Possibilistic C Means Clustering Algorithms

In this section, we discuss intuitionistic fuzzy possibilistic C means clustering algorithms for IFSs and IVIFSs, respectively.

4.1. Intuitionistic Fuzzy Possibilistic C Means Algorithm for IFSs. We develop the intuitionistic fuzzy possibilistic C means (IFPCM) model and corresponding algorithm for IFSs. We take the basic distance measure D_j in (7) as proximity function of IFPCM; the objective function of IFPCM model can then be defined as follows:

$$\underbrace{\min}_{(U,T,V)} \left\{ J_{m,\eta}(U,T,V;X) = \sum_{k=1}^{p}\sum_{i=1}^{c}\left(u_{ik}^{m}+t_{ik}^{\eta}\right) D_{\alpha}^{2}(Z_k,V_i) \right\}$$
$$\text{subject to} \quad m>1, \quad \eta>1, \quad 0<u_{ik}, \quad t_{ik}<1$$
$$\sum_{i=1}^{c} u_{ik}=1 \quad \forall k \text{ i.e., } U\in M_{fcn}$$
$$\sum_{k=1}^{p} t_{ik}=1 \quad \forall i \text{ i.e., } T^{t}\in M_{fnc}. \tag{18}$$

Here $Z=\{Z_1,\ldots,Z_p\}$ are p IFSs each with n elements, c is the number of clusters ($1\le c\le p$), and $V=\{V_1,\ldots,V_c\}$ are the prototypical IFSs, that is, centroid of the clusters. The parameter m is the fuzzy factor, u_{ik} is the membership degree of jth sample Z_j to the ith cluster, $U=(u_{ik})_{c\times p}$ is matrix of order $c\times p$, parameter η is the typicality factor, t_{ik} is the typicality of jth sample Z_j to the ith cluster, and $T=(t_{ik})_{c\times p}$ is typicality matrix.

To solve the optimization problem stated in (18), we make use of Lagrange multiplier method [39], which is discussed below. Considering

$$L=\sum_{k=1}^{p}\sum_{i=1}^{c}\left(u_{ik}^{m}+t_{ik}^{\eta}\right)D_{\alpha}^{2}(Z_k,V_i)-\sum_{k=1}^{p}\lambda_k\left(\sum_{i=1}^{c}u_{ik}-1\right) - \sum_{k=1}^{p}\xi_k\left(\sum_{i=1}^{c}t_{ik}-1\right), \tag{19}$$

where,

$$D_{\alpha}^{2}(Z_k,V_i)=\frac{1}{2}\sum_{l=1}^{n}w_l\left(\left(\mu_{Z_k}(x_l)-\mu_{V_i}(x_l)\right)^2\right) + \left(\nu_{Z_k}(x_l)-\nu_{V_i}(x_l)\right)^2 + \left(\pi_{Z_k}(x_l)-\pi_{V_i}(x_l)\right)^2. \tag{20}$$

Furthermore, $\forall 1\le i\le c$, $1\le k\le p$; let

$$\frac{\partial L}{\partial u_{ik}}=0,\quad \frac{\partial L}{\partial t_{ik}}=0,\quad \frac{\partial L}{\partial \lambda_k}=0,\quad \frac{\partial L}{\partial \xi_k}=0. \tag{21}$$

From the above system of equations, we have the following expressions:

$$u_{ik}=\frac{1}{\sum_{r=1}^{c}\left(D_{\alpha}(Z_k,V_i)/D_{\alpha}(Z_k,V_r)\right)^{2/(m-1)}};\quad 1\le i\le c,\quad 1\le k\le p \tag{22}$$

$$t_{ik}=\frac{1}{\sum_{r=1}^{c}\left(D_{\alpha}(Z_k,V_i)/D_{\alpha}(Z_k,V_r)\right)^{2/(\eta-1)}};\quad 1\le i\le c,\quad 1\le k\le p. \tag{23}$$

Now we proceed to compute V_i; $i=1,\ldots,c$, the prototypical IFSs. Let us assume that

$$\frac{\partial L}{\partial \mu_{V_i}(x_l)}=\frac{\partial L}{\partial \nu_{V_i}(x_l)}=\frac{\partial L}{\partial \pi_{V_i}(x_l)}=0. \tag{24}$$

From the above expression we have

$$\mu_{V_i}(x_l)=\frac{\sum_{k=1}^{p}\left(u_{ik}^{m}+t_{ik}^{\eta}\right)\mu_{Z_k}(x_l)}{\sum_{k=1}^{p}\left(u_{ik}^{m}+t_{ik}^{\eta}\right)};\quad 1\le i\le c,\ 1\le l\le n, \tag{25}$$

$$\nu_{V_i}(x_l)=\frac{\sum_{k=1}^{p}\left(u_{ik}^{m}+t_{ik}^{\eta}\right)\nu_{Z_k}(x_l)}{\sum_{k=1}^{p}\left(u_{ik}^{m}+t_{ik}^{\eta}\right)};\quad 1\le i\le c,\ 1\le l\le n, \tag{26}$$

$$\pi_{V_i}(x_l)=\frac{\sum_{k=1}^{p}\left(u_{ik}^{m}+t_{ik}^{\eta}\right)\pi_{Z_k}(x_l)}{\sum_{k=1}^{p}\left(u_{ik}^{m}+t_{ik}^{\eta}\right)};\quad 1\le i\le c,\ 1\le l\le n. \tag{27}$$

For simplicity, we define weighted average operator for IFSs as follows.

Let $A=\{A_1,\ldots,A_p\}$ be a set of IFSs each with n elements; let $\omega=\{\omega_1,\ldots,\omega_p\}$ be a set of weights for IFSs, respectively, with $\sum_{j=1}^{p}\omega_j=1$; and then the weighted operator f is defined as

$$f(A,\omega) = \left\{\left\langle x_l, \sum_{j=1}^{p}\omega_j\mu_{A_j}(x_l), \sum_{j=1}^{p}\omega_j\nu_{A_j}(x_l)\right\rangle \mid 1\le l\le n\right\}. \tag{28}$$

According to (25) to (28), if we assume

$$\omega^{(i)} = \left\{ \frac{(u_{i1} + t_{i1})}{\sum_{k=1}^{p} (u_{ik} + t_{ik})}, \dots, \frac{(u_{ip} + t_{ip})}{\sum_{k=1}^{p} (u_{ik} + t_{ik})} \right\}; \quad 1 \le i \le c, \tag{29}$$

the prototypical IFSs $V = \{V_1, \dots, V_c\}$ of the IFPCM model can be computed as follows:

$$\begin{aligned} V_i &= f\left(Z, \omega^{(i)}\right) \\ &= \left\{ \left\langle x_l, \sum_{j=1}^{p} \omega_j^{(i)} \mu_{Z_j}(x_l), \sum_{j=1}^{p} \omega_j^{(i)} \nu_{Z_j}(x_l) \right\rangle \mid 1 \le l \le n \right\}, \\ & \qquad 1 \le i \le c. \end{aligned} \tag{30}$$

Since the above equations (22), (23), and (30) are computationally interdependent, we exploit an iterative procedure similar to the FPCM algorithm to solve these equations. The steps of algorithm are as follows.

IFPCM Algorithm

Step 1. Initialize the seed values $V(0)$; let $x = 0$ and set $\varepsilon > 0$.

Step 2(i). Calculate $U(x) = (u_{ik}(x))_{c\times p}$, where

(a) if $\forall k, r, D_\alpha(Z_k, V_r(x)) > 0$, then $u_{ik}(x) = 1/(\sum_{r=1}^{c}(D_\alpha(Z_k, V_i(x))/D_\alpha(Z_k, V_r(x)))^{2/(m-1)})$; $1 \le i \le c$, $1 \le k \le p$,

(b) if $\exists k, r$ such that $D_\alpha(Z_k, V_r(x)) = 0$, then let $u_{rk}(x) = 1$ and $u_{ik}(x) = 0\ \forall i \ne r$.

Step 2(ii). Calculate $T(x) = (t_{ik}(x))_{c\times p}$, where

(a) if $\forall k, r, D_\alpha(Z_k, V_r(x)) > 0$, then $t_{ik}(x) = 1/(\sum_{r=1}^{c}(D_\alpha(Z_k, V_i(x))/D_\alpha(Z_k, V_r(x)))^{2/(\eta-1)})$; $1 \le i \le c$, $1 \le k \le p$,

(b) if $\exists k, r$ such that $D_\alpha(Z_k, V_r(x)) = 0$, then let $t_{rk}(x) = 1$ and $t_{ik}(x) = 0\ \forall i \ne r$.

Step 3. Calculate $V(x+1) = \{V_1(x+1), \dots, V_c(x+1)\}$, where

$$V_i(x+1) = f\left(Z, \omega^{(i)}(x+1)\right), \quad 1 \le i \le c \tag{31}$$

$$\begin{aligned} &\omega^{(i)}(x+1) \\ &= \left\{ \frac{(u_{i1}(x) + t_{i1}(x))}{\sum_{k=1}^{p} (u_{ik}(x) + t_{ik}(x))}, \dots, \frac{(u_{ip}(x) + t_{ip}(x))}{\sum_{k=1}^{p} (u_{ik}(x) + t_{ik}(x))} \right\}; \\ & \qquad 1 \le i \le c. \end{aligned} \tag{32}$$

Step 4. If $\sum_{i=1}^{c}(D_\alpha(V_i(x), V_i(x+1))/c) < \varepsilon$, then go to Step 5; otherwise, let $x = x + 1$, and return to Step 2.

Step 5. End

The pseudocode of the IFPCM algorithm is given in Algorithm 1.

4.2. Interval Valued Intuitionistic Fuzzy Possibilistic C Means Algorithm for IVIFSs. If the collected data are expressed as IVIFSs, then we extend IFPCM to interval valued intuitionistic fuzzy possibilistic C means (IVIFPCM) model. We take the basic distance measure D_γ in (10) as the proximity function of the IVIFCM. The objective function of IVIFPCM model can be defined as follows:

$$\begin{aligned} &\underbrace{\min}_{(U,T,\widetilde{V})} \left\{ J_{m,\eta}\left(U, T, \widetilde{V}; X\right) = \sum_{k=1}^{p}\sum_{i=1}^{c} \left(u_{ik}^{m} + t_{ik}^{\eta}\right) D_\gamma^2\left(\widetilde{Z}_k, \widetilde{V}_i\right) \right\} \\ &\text{subject to} \quad m > 1, \quad \eta > 1, \quad 0 < u_{ik}, \quad t_{ik} < 1 \end{aligned} \tag{33}$$

$$\sum_{i=1}^{c} u_{ik} = 1 \quad \forall k \text{ i.e., } U \in M_{fcn} \tag{34}$$

$$\sum_{k=1}^{p} t_{ik} = 1 \quad \forall i \text{ i.e., } T^{t} \in M_{fnc}. \tag{35}$$

Here $\widetilde{Z} = \{\widetilde{Z}_1, \dots, \widetilde{Z}_p\}$ are p IVIFSs each with n elements, c is the number of clusters ($1 < c < p$), and $\widetilde{V} = \{\widetilde{V}_1, \dots, \widetilde{V}_c\}$ are the prototypical IVIFSs, that is, centroids of the clusters. The parameter m is the fuzzy factor, u_{ik} is the membership degree of jth sample $\widetilde{Z}_j$ to the ith cluster, $U = (u_{ik})_{c\times p}$ is matrix of order $c \times p$, parameter η is the typicality factor, t_{ik} is the typicality of jth sample $\widetilde{Z}_j$ to the ith cluster, and $T = (t_{ik})_{c\times p}$ is typicality matrix.

To solve the optimization problem stated in (30) to (35), we make use of Lagrange multiplier method [39], which is discussed below. Considering

$$\begin{aligned} L = &\sum_{k=1}^{p}\sum_{i=1}^{c} \left(u_{ik}^{m} + t_{ik}^{\eta}\right) D_\gamma^2\left(\widetilde{Z}_k, \widetilde{V}_i\right) - \sum_{k=1}^{p} \lambda_k \left(\sum_{i=1}^{c} u_{ik} - 1 \right) \\ &- \sum_{k=1}^{p} \xi_k \left(\sum_{i=1}^{c} t_{ik} - 1 \right), \end{aligned} \tag{36}$$

where,

$$\begin{aligned} D_\gamma^2\left(\widetilde{Z}_k, \widetilde{V}_i\right) = \frac{1}{4}\sum_{l=1}^{n} w_l \Bigg(&\left(\widetilde{\mu}_{\widetilde{Z}_k}^{L}(x_l) - \widetilde{\mu}_{\widetilde{V}_i}^{L}(x_l)\right)^2 \\ &+ \left(\widetilde{\mu}_{\widetilde{Z}_k}^{U}(x_l) - \widetilde{\mu}_{\widetilde{V}_i}^{U}(x_l)\right)^2 \\ &+ \left(\widetilde{\nu}_{\widetilde{Z}_k}^{L}(x_l) - \widetilde{\nu}_{\widetilde{V}_i}^{L}(x_l)\right)^2 \\ &+ \left(\widetilde{\nu}_{\widetilde{Z}_k}^{U}(x_l) - \widetilde{\nu}_{\widetilde{V}_i}^{U}(x_l)\right)^2 \\ &+ \left(\widetilde{\pi}_{\widetilde{Z}_k}^{L}(x_l) - \widetilde{\pi}_{\widetilde{V}_i}^{L}(x_l)\right)^2 \\ &+ \left(\widetilde{\pi}_{\widetilde{Z}_k}^{U}(x_l) - \widetilde{\pi}_{\widetilde{V}_i}^{U}(x_l)\right)^2 \Bigg). \end{aligned} \tag{37}$$

Given an unlabeled dataset $X = \{x_1, \ldots, x_n\}$, partition X into $1 < c < n$ clusters such that objective function $J_{m,\eta}(U, T, V; X)$ is minimized
(1) **Input:** Consider the seed values $V(0)$ and assume $x = 0$ and set $\varepsilon > 0$
(2) **Output:** Generate clusters using the IFPCM clustering algorithm for IFSs
(3) **begin procedure**
(4) **repeat**
(5) calculate $U(x) = (u_{ik}(x))_{c\times p}$
(6) **begin**
(7) **if** $(\forall k, r, D_\alpha(Z_k, V_r(x)) > 0)$ **then**
(8) $u_{ik}(x) = \dfrac{1}{\sum_{r=1}^{c}(D_\alpha(Z_k, V_i(x))/D_\alpha(Z_k, V_r(x)))^{2/(m-1)}}; \quad 1 \le i \le c, 1 \le k \le p$
(9) **end if**
(10) **if** $(\exists k, r, D_\alpha(Z_k, V_r(x)) = 0)$ **then**
(11) $u_{rk}(x) = 1$ and $u_{ik}(x) = 0\ \forall i \ne r$
(12) **end if**
(13) **end**
(14) calculate $T(x) = (t_{ik}(x))_{c\times p}$
(15) **begin**
(16) **if** $(\forall k, r, D_\alpha(Z_k, V_r(x)) > 0)$ **then**
(17) $t_{ik}(x) = \dfrac{1}{\sum_{r=1}^{c}(D_\alpha(Z_k, V_i(x))/D_\alpha(Z_k, V_r(x)))^{2/(\eta-1)}}; \quad 1 \le i \le c, 1 \le k \le p$
(18) **end if**
(19) **if** $(\exists k, r, D_\alpha(Z_k, V_r(x)) = 0)$ **then**
(20) $t_{rk}(x) = 1$ and $t_{ik}(x) = 0\ \forall i \ne r$
(21) **end if**
(22) **end**
(23) calculate $V(x+1) = \{V_1(x+1), \ldots, V_c(x+1)\}; V_i(x+1) = f(Z, \omega^{(i)}(x+1)), 1 \le i \le c; \omega^{(i)}(x+1) = \{(u_{i1}(x) + t_{i1}(x))/(\sum_{k=1}^{p}(u_{ik}(x) + t_{ik}(x))), \ldots, (u_{ip}(x) + t_{ip}(x))/(\sum_{k=1}^{p}(u_{ik}(x) + t_{ik}(x)))\}, 1 \le i \le c$
(24) $x = x + 1$
(25) **until** $(\sum_{i=1}^{c}(D_\alpha(V_i(x), V_i(x+1))/c) < \varepsilon)$
(26) **end procedure**

Algorithm 1

Similar to IFPCM model, we establish the system of partial differential functions of L as follows:

$$\begin{gathered}\frac{\partial L}{\partial u_{ik}} = 0; \quad 1 \le i \le c,\ 1 \le k \le p,\\ \frac{\partial L}{\partial t_{ik}} = 0; \quad 1 \le i \le c,\ 1 \le k \le p,\\ \frac{\partial L}{\partial \lambda_k} = 0, \quad \frac{\partial L}{\partial \xi_k} = 0; \quad 1 \le k \le p,\\ \frac{\partial L}{\partial \mu^L_{\widetilde{V}_i}(x_l)} = \frac{\partial L}{\partial \nu^L_{\widetilde{V}_i}(x_l)} = \frac{\partial L}{\partial \pi^L_{\widetilde{V}_i}(x_l)} = 0,\\ \frac{\partial L}{\partial \mu^U_{\widetilde{V}_i}(x_l)} = \frac{\partial L}{\partial \nu^U_{\widetilde{V}_i}(x_l)} = \frac{\partial L}{\partial \pi^U_{\widetilde{V}_i}(x_l)} = 0\\ 1 \le i \le c, \quad 1 \le l \le p.\end{gathered} \tag{38}$$

The solution for the above system of equations is

$$u_{ik} = \frac{1}{\sum_{r=1}^{c}\left(D_\gamma(\widetilde{Z}_k, \widetilde{V}_i)/D_\gamma(\widetilde{Z}_k, \widetilde{V}_r)\right)^{2/(m-1)}}; \quad 1 \le i \le c, \quad 1 \le k \le p \tag{39}$$

$$t_{ik} = \frac{1}{\sum_{r=1}^{c}\left(D_\gamma(\widetilde{Z}_k, \widetilde{V}_i)/D_\gamma(\widetilde{Z}_k, \widetilde{V}_r)\right)^{2/(\eta-1)}}; \quad 1 \le i \le c, \quad 1 \le k \le p, \tag{40}$$

$$\begin{aligned}\widetilde{V}_i &= \widetilde{f}\left(\widetilde{Z}, \omega^{(i)}\right)\\ &= \left\{\left\langle x_l, \left[\sum_{j=1}^{p}\omega_j^{(i)}\widetilde{\mu}^L_{\widetilde{Z}_j}(x_l), \sum_{j=1}^{p}\omega_j^{(i)}\widetilde{\mu}^U_{\widetilde{Z}_j}(x_l)\right],\right.\right.\\ &\qquad \left.\left.\left[\sum_{j=1}^{p}\omega_j^{(i)}\widetilde{\nu}^L_{\widetilde{Z}_j}(x_l), \sum_{j=1}^{p}\omega_j^{(i)}\widetilde{\nu}^U_{\widetilde{Z}_j}(x_l)\right]\right\rangle \mid 1 \le l \le n\right\},\\ &\qquad 1 \le i \le c,\end{aligned} \tag{41}$$

where

$$\omega^{(i)} = \left\{\frac{(u_{i1} + t_{i1})}{\sum_{k=1}^{p}(u_{ik} + t_{ik})}, \ldots, \frac{(u_{ip} + t_{ip})}{\sum_{k=1}^{p}(u_{ik} + t_{ik})}\right\}; \quad 1 \le i \le c. \tag{42}$$

Because (41) and (42) are computationally interdependent, we exploit a similar iteration procedure as follows.

IVIFPCM Algorithm

Step 1. Initialize the seed values $\widetilde{V}(0)$; let $x = 0$ and set $\varepsilon > 0$.

Step 2(i). Calculate $U(x) = (u_{ik}(x))_{c\times p}$, where

(c) if $\forall k, r, D_\gamma(\widetilde{Z}_k, \widetilde{V}_r(x)) > 0$, then $u_{ik}(x) = 1/(\sum_{r=1}^{c}(D_\gamma(\widetilde{Z}_k, \widetilde{V}_i(x))/D_\gamma(\widetilde{Z}_k, \widetilde{V}_r(x)))^{2/(m-1)})$; $1 \le i \le c$, $1 \le k \le p$.

(d) if $\exists k, r$ such that $D_\gamma(\widetilde{Z}_k, \widetilde{V}_r(x)) = 0$, then let $u_{rk}(x) = 1$ and $u_{ik}(x) = 0\ \forall i \ne r$.

Step 2(ii). Calculate $T(x) = (t_{ik}(x))_{c\times p}$, where

(c) if $\forall k, r, D_\gamma(\widetilde{Z}_k, \widetilde{V}_r(x)) > 0$, then $t_{ik}(x) = 1/(\sum_{r=1}^{c}(D_\gamma(\widetilde{Z}_k, \widetilde{V}_i(x))/D_\gamma(\widetilde{Z}_k, \widetilde{V}_r(x)))^{2/(\eta-1)})$; $1 \le i \le c$, $1 \le k \le p$,

(d) if $\exists k, r$ such that $D_\gamma(\widetilde{Z}_k, \widetilde{V}_r(x)) = 0$, then let $t_{rk}(x) = 1$ and $t_{ik}(x) = 0\ \forall i \ne r$.

Step 3. Calculate $\widetilde{V}(x+1) = \{\widetilde{V}_1(x+1), \ldots, \widetilde{V}_c(x+1)\}$, where

$$\widetilde{V}_i(x+1) = \widetilde{f}\left(\widetilde{Z}, \omega^{(i)}(x+1)\right), \quad 1 \le i \le c$$

$$\omega^{(i)}(x+1) = \left\{\frac{(u_{i1}(x) + t_{i1}(x))}{\sum_{k=1}^{p}(u_{ik}(x) + t_{ik}(x))}, \ldots, \frac{\left(u_{ip}(x) + t_{ip}(x)\right)}{\sum_{k=1}^{p}(u_{ik}(x) + t_{ik}(x))}\right\}; \quad 1 \le i \le c. \tag{43}$$

Step 4. If $\sum_{i=1}^{c}(D_\gamma(\widetilde{V}_i(x), \widetilde{V}_i(x+1))/c) < \varepsilon$, then go to Step 5; otherwise let $x = x + 1$, and return to Step 2.

Step 5. End

The pseudocode of the IVIFPCM algorithm is given in Algorithm 2.

5. Experimental Results

In this section, we enumerate the results of experiments performed on both real world and simulated datasets [32] in order to demonstrate the effectiveness of IFPCM clustering algorithm. IFPCM algorithm is implemented through MATLAB. We first explain the steps of the algorithm by the use of some experimental data which is evaluated through cluster validity measures. Next, the algorithm is applied to some classification datasets, that is, data with labeled patterns in order to examine its clustering accuracy.

5.1. Application of IFPCM Algorithm on Experimental Data. The parameters set in IFPCM algorithm are shown in Table 1. It is to be noted that if $\pi(x) = 0\ \forall x \in X$, then IFPCM is reduced to FPCM algorithm. Hence, we present a comparative performance of both the algorithms.

The experimental data used here is investment portfolio dataset which contains information regarding ten investments at the disposal of the investor to invest some money to be classified in ICICI prudential financial services, India. Let I_i; $i = 1, \ldots, 10$ be the investments described by six attributes, namely a_1: investment price; a_2: advance mobilization; a_3: time period; a_4: return on investment; a_5: risk factor; a_6: security factor. The weight vector of these attributes is $w = (0.20, 0.10, 0.30, 0.15, 0.10, 0.15)$. The characteristics of ten investments under six attributes are represented by IFSs in Table 2. Simulated datasets are used for comparing with the experimental data. We assume that there are three classes in the simulated dataset, C_i; $i = 1, 2, 3$. The number of IFSs in each class is considered as 300. The different classes have different IFSs which are characterized as follows:

Table 1: IFPCM parameters.

Parameter	Description
f	Name of input file
c	Number of clusters (default value = 3)
m	Fuzzy factor (default value = 2)
η	Typicality factor (default value = 2)
w	Type of sample weights (default value = 0 (equal); user specified value = 1)
s	Type of initial centroids (default value = 0 (random); user specified value = 1)
i	Maximum number of iterations till convergence (default value = 100)
t	Threshold for iterations stoppage (default value = 0.001)

(a) IFSs in C_1 have relatively high and positive scores,

(b) IFSs in C_2 have relatively high and uncertain scores,

(c) IFSs in C_3 have relatively high and negative scores.

Considering this, we generate simulated dataset as follows:

(a) $\mu(x) \sim U(0.7, 1) \wedge \nu(x) + \pi(x) \sim U(0, 1 - \mu(x))\ \forall x \in C_1$,

(b) $\nu(x) \sim U(0.7, 1) \wedge \mu(x) + \pi(x) \sim U(0, 1 - \nu(x))\ \forall x \in C_2$,

(c) $\pi(x) \sim U(0.7, 1) \wedge \mu(x) + \nu(x) \sim U(0, 1 - \pi(x))\ \forall x \in C_3$.

Here, $U(a, b)$ is the uniform distribution on the interval $[a, b]$. We thus generate a simulated dataset which consists of 3 classes comprising 900 IFSs.

5.1.1. Cluster Validity Measures. In IFPCM algorithm big challenge lies in setting the parameter c, that is, the number of clusters. To resolve this, we use two relative measures for fuzzy cluster validity mentioned in [40], namely, partition coefficient (PC) and classification entropy (CE). The descriptions of these two measures are given in Table 3. In PC and CE p is number of samples in the dataset.

5.1.2. IFPCM Algorithm on Investment Portfolio Dataset. IFPCM algorithm is used to cluster ten investments I_i; $i = 1, \ldots, 10$, involving the following steps.

Step 1. Let $c = 3$ and $\varepsilon = 0.005$. Now randomly select initial centroids $V(0)$ from the dataset:

$$V(0) = \begin{bmatrix} I_9 \\ I_{10} \\ I_7 \end{bmatrix}. \tag{44}$$

Given an unlabeled dataset $X = \{x_1, \ldots, x_n\}$, partition X into $1 < c < n$ clusters such that objective function $J_{m,\eta}(U, T, V; X)$ is minimized

(1) **Input:** Consider the seed values $\widetilde{V}(0)$ and assume $x = 0$ and set $\varepsilon > 0$

(2) **Output:** Generate clusters using the IFPCM clustering algorithm for IVIFSs

(3) **begin procedure**

(4) **repeat**

(5) calculate $U(x) = (u_{ik}(x))_{c\times p}$

(6) **begin**

(7) **if** $(\forall k, r, D_\gamma(\widetilde{Z}_k, \widetilde{V}_r(x)) > 0)$ **then**

(8) $u_{ik}(x) = \dfrac{1}{\sum_{r=1}^{c}\left(D_\gamma\left(\widetilde{Z}_k, \widetilde{V}_i(x)\right)/D_\gamma\left(\widetilde{Z}_k, \widetilde{V}_r(x)\right)\right)^{2/(m-1)}}; \quad 1 \le i \le c, 1 \le k \le p$

(9) **end if**

(10) **if** $(\exists\, k, r, D_\gamma(\widetilde{Z}_k, \widetilde{V}_r(x)) = 0)$ **then**

(11) $u_{rk}(x) = 1$ and $u_{ik}(x) = 0\ \forall i \ne r$

(12) **end if**

(13) **end**

(14) calculate $T(x) = (t_{ik}(x))_{c\times p}$

(15) **begin**

(16) **if** $(\forall k, r, D_\gamma(\widetilde{Z}_k, \widetilde{V}_r(x)) > 0)$ **then**

(17) $t_{ik}(x) = \dfrac{1}{\sum_{r=1}^{c}\left(D_\gamma\left(\widetilde{Z}_k, \widetilde{V}_i(x)\right)/D_\gamma\left(\widetilde{Z}_k, \widetilde{V}_r(x)\right)\right)^{2/(\eta-1)}}; \quad 1 \le i \le c, 1 \le k \le p$

(18) **end if**

(19) **if** $(D_\gamma(\widetilde{Z}_k, \widetilde{V}_r(x)) = 0)$ **then**

(20) $t_{rk}(x) = 1$ and $t_{ik}(x) = 0\ \forall i \ne r$

(21) **end if**

(22) **end**

(23) calculate $\widetilde{V}(x+1) = \left\{\widetilde{V}_1(x+1), \ldots, \widetilde{V}_c(x+1)\right\}; \widetilde{V}_i(x+1) = \widetilde{f}\left(\widetilde{Z}, \omega^{(i)}(x+1)\right), 1 \le i \le c; \omega^{(i)}(x+1) = \left\{(u_{i1}(x) + t_{i1}(x))/\left(\sum_{k=1}^{p}(u_{ik}(x) + t_{ik}(x))\right), \ldots, \left(u_{ip}(x) + t_{ip}(x)\right)/\left(\sum_{k=1}^{p}(u_{ik}(x) + t_{ik}(x))\right)\right\}, 1 \le i \le c$

(24) $x = x + 1$

(25) **until** $\left(\sum_{i=1}^{c}\left(D_\gamma\left(\widetilde{V}_i(x), \widetilde{V}_i(x+1)\right)/c\right) < \varepsilon\right)$

(26) **end procedure**

Algorithm 2

Step 2(i). Calculate the membership degrees and centroids iteratively. According to (15), we have

$$U(0) = \begin{bmatrix} 0.4012 & 0.3179 & 0.2812 \\ 0.2155 & 0.2527 & 0.5336 \\ 0.2896 & 0.2312 & 0.4806 \\ 0.8969 & 0.0546 & 0.0512 \\ 0.1666 & 0.6313 & 0.2033 \\ 0.3196 & 0.3906 & 0.2914 \\ 0.1796 & 0.2139 & 0.6079 \\ 0.0000 & 0.0000 & 1.0000 \\ 1.0000 & 0.0000 & 0.0000 \\ 0.0000 & 1.0000 & 0.0000 \end{bmatrix}. \quad (45)$$

Step 2(ii). Calculate the typicality degrees and centroids iteratively. According to (16), we have

$$T(0) = \begin{bmatrix} 0.4215 & 0.3286 & 0.2816 \\ 0.2269 & 0.2626 & 0.5439 \\ 0.2899 & 0.2415 & 0.4809 \\ 0.8996 & 0.0545 & 0.0412 \\ 0.1569 & 0.6416 & 0.2133 \\ 0.3369 & 0.3909 & 0.2813 \\ 0.1696 & 0.2239 & 0.6286 \\ 0.0000 & 1.0000 & 1.0000 \\ 1.0000 & 0.0000 & 1.0000 \\ 1.0000 & 1.0000 & 0.0000 \end{bmatrix}. \quad (46)$$

Step 3. According to (17), we update the centroids as follows:

$$V(1) = \begin{bmatrix} \langle 0.366, 0.386\rangle & \langle 0.839, 0.086\rangle & \langle 0.786, 0.156\rangle & \langle 0.626, 0.186\rangle & \langle 0.209, 0.707\rangle & \langle 0.196, 0.739\rangle \\ \langle 0.760, 0.152\rangle & \langle 0.679, 0.139\rangle & \langle 0.586, 0.269\rangle & \langle 0.496, 0.206\rangle & \langle 0.709, 0.226\rangle & \langle 0.509, 0.466\rangle \\ \langle 0.679, 0.212\rangle & \langle 0.575, 0.209\rangle & \langle 0.669, 0.169\rangle & \langle 0.366, 0.520\rangle & \langle 0.386, 0.569\rangle & \langle 0.669, 0.139\rangle \end{bmatrix}. \quad (47)$$

Step 4. We made a check whether to stop the iterations:

$$\sum_{i=1}^{3} \frac{D_\alpha\left(V_i(0), V_i(1)\right)}{3} = 0.086 > 0.005. \quad (48)$$

Since this value exceeds the chosen threshold value, we continue with the next iteration.

When $x = 1$,

$$U(1) = \begin{bmatrix} 0.3866 & 0.3333 & 0.2755 \\ 0.0866 & 0.1022 & 0.8077 \\ 0.1339 & 0.1239 & 0.7333 \\ 0.8386 & 0.0333 & 0.0355 \\ 0.1027 & 0.7525 & 0.2222 \\ 0.2899 & 0.3786 & 0.2779 \\ 0.0655 & 0.8002 & 0.7530 \\ 0.1000 & 0.1955 & 0.6336 \\ 0.7933 & 0.0550 & 0.0500 \\ 0.0720 & 0.7120 & 0.1012 \end{bmatrix},$$

$$T(1) = \begin{bmatrix} 0.3716 & 0.3065 & 0.2515 \\ 0.2024 & 0.2562 & 0.5233 \\ 0.2698 & 0.2210 & 0.4620 \\ 0.8696 & 0.0525 & 0.0312 \\ 0.1455 & 0.6045 & 0.2030 \\ 0.3036 & 0.3799 & 0.2715 \\ 0.1595 & 0.2138 & 0.6185 \\ 0.1000 & 0.8888 & 0.7775 \\ 0.6867 & 0.0444 & 0.0505 \\ 0.0577 & 0.6600 & 0.0005 \end{bmatrix}, \quad (49)$$

$$V(2) = \begin{bmatrix} \langle 0.354, 0.385\rangle & \langle 0.838, 0.075\rangle & \langle 0.750, 0.145\rangle & \langle 0.616, 0.139\rangle & \langle 0.185, 0.706\rangle & \langle 0.170, 0.725\rangle \\ \langle 0.755, 0.142\rangle & \langle 0.635, 0.174\rangle & \langle 0.533, 0.222\rangle & \langle 0.396, 0.160\rangle & \langle 0.625, 0.222\rangle & \langle 0.472, 0.399\rangle \\ \langle 0.536, 0.176\rangle & \langle 0.512, 0.170\rangle & \langle 0.650, 0.142\rangle & \langle 0.279, 0.515\rangle & \langle 0.327, 0.525\rangle & \langle 0.596, 0.120\rangle \end{bmatrix},$$

$$\sum_{i=1}^{3} \frac{D_\alpha\left(V_i(1), V_i(2)\right)}{3} = 0.009 > 0.005.$$

Since this value exceeds the chosen threshold value, we continue with the next iteration.

When $x = 2$,

$$U(2) = \begin{bmatrix} 0.3864 & 0.3332 & 0.2752 \\ 0.0865 & 0.1021 & 0.8072 \\ 0.1337 & 0.1236 & 0.7331 \\ 0.8385 & 0.0330 & 0.0354 \\ 0.1026 & 0.7524 & 0.2221 \\ 0.2898 & 0.3784 & 0.2778 \\ 0.0653 & 0.8000 & 0.7529 \\ 0.1000 & 0.1954 & 0.6333 \\ 0.7932 & 0.0550 & 0.0499 \\ 0.0716 & 0.7118 & 0.1010 \end{bmatrix},$$

$$T(2) = \begin{bmatrix} 0.3715 & 0.3064 & 0.2512 \\ 0.2022 & 0.2560 & 0.5232 \\ 0.2697 & 0.2209 & 0.4619 \\ 0.8693 & 0.0524 & 0.0311 \\ 0.1454 & 0.6042 & 0.2028 \\ 0.3035 & 0.3798 & 0.2714 \\ 0.1593 & 0.2137 & 0.6184 \\ 0.9999 & 0.8886 & 0.7772 \\ 0.6862 & 0.0443 & 0.0504 \\ 0.0575 & 0.6599 & 0.0004 \end{bmatrix},$$

$$V(3)=\begin{bmatrix}\langle 0.352,0.380\rangle & \langle 0.837,0.077\rangle & \langle 0.752,0.144\rangle & \langle 0.614,0.133\rangle & \langle 0.180,0.705\rangle & \langle 0.172,0.722\rangle\\ \langle 0.765,0.140\rangle & \langle 0.632,0.172\rangle & \langle 0.532,0.220\rangle & \langle 0.389,0.156\rangle & \langle 0.623,0.222\rangle & \langle 0.470,0.396\rangle\\ \langle 0.528,0.169\rangle & \langle 0.510,0.168\rangle & \langle 0.647,0.145\rangle & \langle 0.276,0.512\rangle & \langle 0.326,0.523\rangle & \langle 0.586,0.115\rangle\end{bmatrix},$$

$$\sum_{i=1}^{3}\frac{D_\alpha\left(V_i(2),V_i(3)\right)}{3}=0.003<0.005. \tag{50}$$

Since this value is less than the threshold value, we stop the iterations and calculate the values of $U(3)$ and $T(3)$ when $x=3$:

$$U(3)=\begin{bmatrix}0.3862 & 0.3330 & 0.2750\\ 0.0665 & 0.1018 & 0.8069\\ 0.1335 & 0.1236 & 0.7330\\ 0.8380 & 0.0325 & 0.0352\\ 0.1023 & 0.7516 & 0.2220\\ 0.2886 & 0.3765 & 0.2769\\ 0.0650 & 0.8002 & 0.7527\\ 0.9999 & 0.1953 & 0.6331\\ 0.7922 & 0.0545 & 0.0495\\ 0.0712 & 0.7116 & 0.1009\end{bmatrix},$$

$$T(3)=\begin{bmatrix}0.3712 & 0.3050 & 0.2510\\ 0.2020 & 0.2555 & 0.5226\\ 0.2695 & 0.2205 & 0.4520\\ 0.8690 & 0.0520 & 0.0310\\ 0.1450 & 0.6040 & 0.2025\\ 0.3033 & 0.3796 & 0.2710\\ 0.1590 & 0.2135 & 0.6175\\ 0.9666 & 0.8880 & 0.7769\\ 0.6860 & 0.0442 & 0.0502\\ 0.0572 & 0.6596 & 0.0002\end{bmatrix}. \tag{51}$$

According to $U(4)$ and $T(4)$, cluster validation measures V_{PC} and V_{CE} are calculated as

$$V_{\mathrm{PC}}=\frac{1}{3}\sum_{i=1}^{3}\sum_{k=1}^{10}\left(u_{ik}^2+t_{ik}^2\right)=0.633,$$
$$V_{\mathrm{CE}}=-\frac{1}{10}\sum_{i=1}^{3}\sum_{k=1}^{10}\left(u_{ik}\log u_{ik}+t_{ik}\log t_{ik}\right)=0.865. \tag{52}$$

If we further assume that $u_{ik}\ge 0.75$, $t_{ik}\ge 0.65 \Rightarrow I_j \in C_i$ $(1\le j\le 10, 1\le i\le 3)$, where C_i denotes cluster i, then we have clusters as shown in Table 4.

5.1.3. Convergence of IFPCM Algorithm. Now, we proceed to investigate the convergence of IFPCM algorithm on investment portfolio dataset. The movements of objective function values $J_{m,\eta}(U,T,V;X)$ are shown in Figure 1 along the iterations. As evident from Figure 1, the IFPCM algorithm decreases the objective function value continuously by iterating two phases, namely, updating the membership and typicality degrees in (15) and (16) and updating prototypical IFSs in (17). The IFPCM algorithm has lower computational complexity as compared to other clustering algorithms [1–3, 5–7, 9]. The space and time complexities of IFPCM algorithm are $O(p(n+c)+cn)$ and $O(Icpn)$, where p is the number of samples, n is the number of IFSs in sample, c is the number of clusters, and I is the maximum number of iterations preset for optimal value search process. Some advantages of IFPCM algorithm include simplicity and flexibility, information about the membership, and typicality degrees of samples to all clusters and relatively low computational complexity.

5.1.4. Comparative Performance of IFPCM and FPCM on Investment Portfolio Dataset. In this subsection, we present a comparative performance of IFPCM and FPCM algorithms. We first experiment IFPCM algorithm on the simulated dataset. Here, we set a series of c values in the range of 2 to 10 and compute V_{PC} and V_{CE} measures for each clustering result. The results are given in Table 5, where *Obj* is the objective function value after convergence of IFPCM algorithm. The optimal values of cluster validity measures are highlighted.

As evident from Table 5, when $c = 4$ V_{PC} reaches its optimal value 0.9664 (maximum) and V_{CE} also reaches its optimal value 0.1866 (minimum), this implies that both V_{PC} and V_{CE} are capable of finding optimal number of clusters, that is, c. However, this is not the case for objective function value. From Figure 2, as the number of clusters increases, *Obj* decreases continuously and finally reaches 1.1969 when $c=4$. Hence, the usage of V_{PC} and V_{CE} is justified in the evaluation of clustering results produced by IFPCM algorithm. Next, we experiment FPCM algorithm on the simulated dataset for comparison purpose. The results are given in Table 6. The optimal values of cluster validity measures are highlighted. As indicated by V_{PC} and V_{CE} values in Table 6, FPCM algorithm prefers to cluster the modified simulated datasets into three clusters which are actually away from four true clusters in the data. In other words, FPCM algorithm cannot identify all four classes precisely. This further signifies the importance of uncertainty information in IFSs.

5.2. Examining Clustering Accuracy of IFPCM Algorithm. To assess the ability of IFPCM algorithm to explore natural clusters in real world data, 11 classification and two clustering datasets with numerical attributes are chosen from the University of California at the Irvine Machine Learning Repository [41] and Knowledge Extraction based on Evolutionary Learning Repository [42]. In Table 7, these datasets are summarized.

The accuracy of IFPCM algorithm is adhered by removing the class labels of data before applying the algorithm.

Table 2: Investment portfolio dataset.

	a_1		a_2		a_3		a_4		a_5		a_6	
	$\mu_{I_i}(a_1)$	$\nu_{I_i}(a_1)$	$\mu_{I_i}(a_2)$	$\nu_{I_i}(a_2)$	$\mu_{I_i}(a_3)$	$\nu_{I_i}(a_3)$	$\mu_{I_i}(a_4)$	$\nu_{I_i}(a_4)$	$\mu_{I_i}(a_5)$	$\nu_{I_i}(a_5)$	$\mu_{I_i}(a_6)$	$\nu_{I_i}(a_6)$
I_1	0.30	0.40	0.20	0.70	0.50	0.50	0.80	0.10	0.40	0.50	0.20	0.70
I_2	0.40	0.30	0.50	0.10	0.60	0.20	0.20	0.70	0.30	0.60	0.70	0.20
I_3	0.40	0.20	0.60	0.10	0.80	0.10	0.20	0.60	0.30	0.70	0.50	0.20
I_4	0.30	0.50	0.90	0.00	0.90	0.10	0.70	0.20	0.20	0.90	0.20	0.90
I_5	0.90	0.20	0.70	0.20	0.70	0.10	0.50	0.10	0.90	0.20	0.50	0.70
I_6	0.40	0.30	0.30	0.60	0.20	0.60	0.70	0.10	0.50	0.40	0.30	0.60
I_7	0.50	0.40	0.50	0.20	0.70	0.20	0.30	0.70	0.30	0.70	0.60	0.10
I_8	0.90	0.10	0.70	0.20	0.70	0.10	0.40	0.50	0.40	0.60	0.80	0.00
I_9	0.40	0.40	1.00	0.20	0.90	0.20	0.70	0.20	0.20	0.70	0.20	0.80
I_{10}	0.90	0.10	0.90	0.00	0.60	0.30	0.50	0.20	0.80	0.10	0.60	0.40

Table 3: Description of two cluster validity criteria.

Validity criteria	Functional description	Optimal cluster number
PC	$V_{\mathrm{PC}} = \frac{1}{p}\sum_{i=1}^{c}\sum_{k=1}^{p}\left(u_{ik}^2 + t_{ik}^2\right)$	$\underbrace{\operatorname{argmax}}_{c}(V_{\mathrm{PC}}, U, T, c)$
CE	$V_{\mathrm{CE}} = -\frac{1}{p}\sum_{i=1}^{c}\sum_{k=1}^{p}\left(u_{ik}\log u_{ik} + t_{ik}\log t_{ik}\right)$	$\underbrace{\operatorname{argmin}}_{c}(V_{\mathrm{CE}}, U, T, c)$

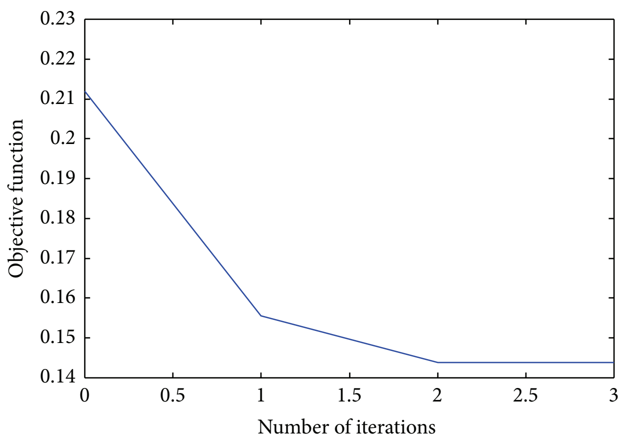

Figure 1: The convergence of IFPCM algorithm on investment portfolio dataset.

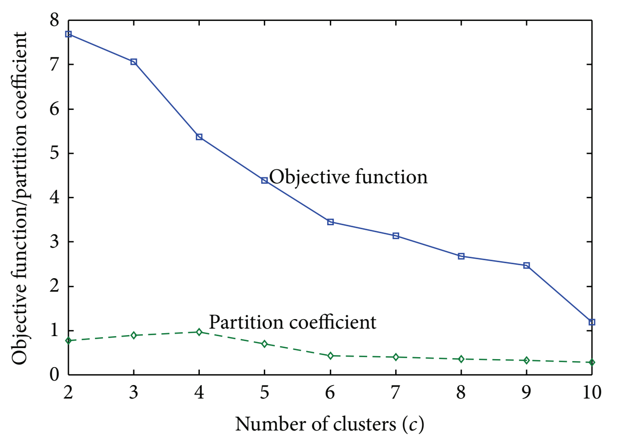

Figure 2: Comparative performance of Obj and V_{PC} given different c values.

Table 4: Clustering result of the investment portfolio dataset by IFPCM.

Instance	Cluster ID
I_6, I_7	1
I_4, I_8	2
I_2, I_5, I_9	3
I_1, I_3, I_{10}	4

Each attribute value of all datasets is rescaled to a unit interval [0, 1] via linear transformation. The clustering results of the application of IFPCM algorithm on 11 classification datasets are shown in Table 8, where FPCM algorithm results as a benchmark fuzzy clustering method are also provided. The threshold t for effectiveness measure is set to 0.1 for all the datasets, provided that at least two clusters are explored. For fairness of comparison between IFPCM and FPCM algorithms, the number of clusters that are needed by FPCM as a parameter for each dataset is set to the number of clusters that are explored by IFPCM. In this table, two super cells for each dataset are confusion matrices [43], which represent the clustering accuracy of IFPCM and FPCM algorithms on that data. In confusion matrix, cell c_{ij} contains the number of patterns with class label i which are grouped by cluster j. Accordingly, the cells in each row of actual class label are summed up to the number of patterns in that class. In addition, the summation of each column's cells represents the number of patterns in that cluster. Ideally, optimal clusters are

Table 5: IFPCM algorithm with different cluster numbers on investment portfolio dataset.

	c								
	2	3	4	5	6	7	8	9	10
Obj	7.6896	7.0678	5.3655	4.3867	3.4464	3.1393	2.6695	2.4755	1.1969
V_{PC}	0.7666	0.8986	**0.9664**	0.6999	0.4365	0.4050	0.3555	0.3333	0.2757
V_{CE}	0.5766	0.4672	**0.1866**	0.7679	1.0009	1.3339	1.4755	1.6462	1.8689

Table 6: FPCM algorithm with different cluster numbers on investment portfolio dataset.

	c								
	2	3	4	5	6	7	8	9	10
V_{PC}	**0.9866**	0.6987	0.5366	0.4999	0.3566	0.2999	0.2555	0.2222	0.2009
V_{CE}	**0.0509**	0.7000	1.1569	1.2767	1.5509	1.8750	2.0215	2.2554	2.2869

Table 7: Datasets used for evaluating IFPCM algorithm.

Dataset	Number of attributes	Number of classes	Number of samples
Iris	4	3	150
Thyroid	5	3	215
Ecoli	7	8	336
Cancer (Breast)	9	2	684
Glass	9	6	214
Vowel	10	11	990
Wine	13	3	178
Vehicle	18	4	846
WDBC	30	2	569
Ionosphere	33	2	351
Sonar	60	2	208

achieved when patterns of each class are covered by only one cluster and each cluster just contains patterns of one class. Such a case occurred for the first class of the Iris dataset in both clustering methods. As evident from results in Table 8, the performance of IFPCM is better as compared to FPCM algorithm in all datasets.

Although confusion matrices for Vehicle and WDBC datasets show almost identical overall performance, the clustering accuracy of IFPCM algorithm for Iris, Thyroid, Cancer, Glass, and Sonar datasets is comparatively better. On the other hand, FPCM algorithm obtains better performance for Ecoli, Vowel, Wine, and Ionosphere datasets. IFPCM algorithm explores potential clusters that are embedded in datasets and needs only a distinguishing threshold t for effectiveness measure while the number of clusters in FPCM algorithm is provided in advance. The clusters obtained by FPCM algorithm convey no specific cognitive interpretation while those clusters explored by IFPCM algorithm are identified by intuitionistic measure. This intuitionistic interpretability which represents the clusters justifies the claim that IFPCM algorithm is more suitable for knowledge discovery in datasets. The IFPCM algorithm is more robust to outliers and noise in data. Moreover, the computational cost of IFPCM algorithm is higher than that of FPCM algorithm as given in Table 9. Although some of the datasets which are used in the experiments are high dimensional, they are not too large. The application of IFPCM algorithm on large datasets consumes greater CPU time. Since, threshold t for effectiveness measure is set to 0.1 for all data given in Table 8, the number of clusters that are explored for multiclass datasets is less than their classes. Consequently, IFPCM algorithm needs a lower threshold to explore more clusters. Table 10 illustrates clustering results obtained for these datasets when threshold t is set to 0.01.

To compare the effectiveness of IFPCM algorithm with other fuzzy clustering methods, some recently developed algorithms have been considered and their results on some real world datasets are presented in Table 11. The performance of these methods is expressed in terms of pureness ratio which is the average pureness of clusters after cluster labeling that is based on maximum number of sample classes in each cluster. Along with FPCM algorithms, some other clustering algorithms that run on these datasets are FCM, PCM [8], α-cut FCM (AFCM) [18], entropy based fuzzy clustering (EFC) [19], fuzzy mixture of Student's t factor analyzers (FMSFA) [20], and fuzzy principal component analysis guided robust k-means (FPRk) [21]. The IFPCM algorithm maintains appreciable performance compared to

Table 8: Clustering results of IFPCM and FPCM algorithms.

Dataset	Actual class label	IFPCM number of cluster			FPCM number of cluster		
		1	2	3	1	2	3
Iris	1	50			50		
	2		50			47	3
	3		17	33		12	38
Thyroid	1	150			145	5	
	2	12	23			35	
	3	30			30		
Ecoli	1	139	4		139	4	
	2	17	60		15	62	
	3		2			2	
	4		2		1	1	
	5	2	33		3	32	
	6	3	17		5	15	
	7		5			5	
	8	13	39		25	27	
Cancer	1	20	219		22	217	
	2	439	6		438	7	
Glass	1	39	31		70		
	2	39	37		66	10	
	3	9	8		17		
	4	9	4		3	10	
	5	9			5	4	
	6	30			5	25	
Vowel	1	17	22	51		39	51
	2	12	5	73		42	48
	3	2	22	66		51	39
	4	2	55	33	4	65	21
	5	4	44	42	9	51	30
	6	20	42	28		66	24
	7	72	12	6	16	65	9
	8	69	16	5	2	58	30
	9	82	3	5	5	35	50
	10	12	45	33	10	50	30
	11	25	15	50		25	65
Wine	1	59			59		
	2	12	55	4	4	60	7
	3			48			48
Vehicle	1	196	3		196	3	
	2	102	115		98	119	
	3	162	56		162	56	
	4	96	116		95	117	
WDBC	1	46	166		26	186	
	2	350	7		345	12	
Ionosphere	1	116	109		156	69	
	2	30	96		36	90	
Sonar	1	50	47		39	58	
	2	54	57		50	61	

Table 9: Computational costs of IFPCM and FPCM algorithms.

Dataset	CPU time (in seconds)	
	IFPCM	FPCM
Iris	0.86	0.15
Thyroid	0.16	0.09
Ecoli	1.00	0.09
Cancer	5.50	0.10
Glass	0.99	0.09
Vowel	21.00	0.12
Wine	0.99	0.09
Vehicle	5.00	0.12
WDBC	2.10	0.13
Ionosphere	2.50	0.09
Sonar	1.60	0.16

other methods in terms of pureness ratio although this is not true for clustering accuracy. The specification of threshold t in IFPCM algorithm for effectiveness measure is more intuitionistic and less data dependent in nature.

6. Conclusion

In this paper, we have proposed IFPCM and IVIFPCM algorithms to cluster IFSs and IVIFSs, respectively. Both the algorithms are developed by integrating concepts of FPCM, IFSs, IVIFSs, and basic distance measures. In interval valued intuitionistic fuzzy environments, the clustering algorithm has membership and nonmembership degrees as intervals rather than exact numbers. The algorithms overcome problems involved with membership values of objects to each cluster by generalizing degrees of membership of objects to each cluster. This is achieved by extending membership and nonmembership degrees with hesitancy degree. The algorithms also provide information about membership and typicality degrees of samples to all clusters. Experiments on both real world and simulated datasets show that IFPCM has some notable advantages over FPCM. IFPCM algorithm is simple and flexible. It generates valuable information and produces overlapped clusters where instances have different membership degrees in accordance with different real world applications. The algorithm has relatively lower computational complexity. It also takes into account inherent uncertainty in information captured by IFSs which becomes crucial for success of some clustering tasks. The evaluation of the algorithm is performed through cluster validity measures. The clustering accuracy of the algorithm is determined by classification datasets with labeled patterns. IFPCM maintains appreciable performance compared to other methods in terms of pureness ratio although this is not true for clustering accuracy. For multiclass datasets there is a chance for exploring fewer clusters than classes. This is handled by

Table 10: Clustering results of IFPCM algorithm on multiclass datasets.

Dataset	Actual class label	Number of cluster							
		1	2	3	4	5	6	7	8
Thyroid	1	150							
	2	12	23						
	3	9		21					
Ecoli	1	21	20	102					
	2	59	9	9					
	3	1		1					
	4	1		1					
	5	30	2	3					
	6	2	2	16					
	7	2		3					
	8	9	4	39					
Glass	1	46	24						
	2	39	25	12					
	3	12	5						
	4	2	3	8					
	5	5		4					
	6	5		24					
Vowel	1					54	9	27	
	2		24			45	9	12	
	3		16	33		35			6
	4		55	3	20				12
	5		33	14	17	5			21
	6	5	26	2	35		12		10
	7	25	5		55		5		
	8	39	2	3	25	3	3	3	12
	9	46			9	12	5	12	6
	10	15	36	4	12	4		3	16
	11					46	11	33	
Vehicle	1	115	2	82					
	2	64	114	39					
	3	72	60	86					
	4	60	119	33					

decreasing value of threshold for effectiveness measure. The specification of threshold is more intuitionistic and less data dependent in nature. For an unknown dataset, IFPCM must compute cluster accuracy measure for all potential clusters. A sudden drop in values should be considered as stopping criterion whereby the number of clusters is determined which can be explored. The different drawbacks of FPCM are

Table 11: Comparative analysis of IFPCM algorithm with other Fuzzy clustering methods in terms of Pureness ratio.

Fuzzy clustering method	Parameters	Datasets				
		Iris	Wine	Thyroid	Cancer	Sonar
IFPCM	$t = 0.1$	93.62	93.86	**98.55**	**98.16**	55.39
FPCM	$t = 0.1$	91.60	93.09	96.46	96.28	53.36
FCM [7]	$c = 3, m = 2$	89.93	90.46	96.08	95.83	53.61
PCM [8]	$c = 3, m = 2$ $\varepsilon = 0.01$	80.80	91.42	87.78	77.17	52.72
AFCM [18]	$c = 3, m = 2$ $\varepsilon = 0.01$	89.79	94.68	69.46	73.90	**56.47**
EFC [19]	$\beta = 0.5 - 0.7$ $\gamma = 0 - 0.05m_I$	96.68	87.15	45.15	95.60	—
FMSFA [20]	$c = 3, q = 1$ $\lambda = 0.7 - 0.9$	**98.72**	96.26	91.44	94.96	—
FPRk [21]	$k = 3, \lambda = 0.1$ $\beta = 1.5$	79.60	96.50	96.18	—	—

effectively handled by possibilistic fuzzy C means (PFCM) model proposed by Pal et al. in 2005. Our future work entails development of IFSs framework for PFCM.

Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

References

[1] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, John Wiley & Sons, New York, NY, USA, 2nd edition, 2000.

[2] A. K. Jain and R. C. Dubes, *Algorithms for Clustering Data*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1988.

[3] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: a review," *ACM Computing Surveys*, vol. 31, no. 3, pp. 264–323, 1999.

[4] H. Frigui, "Simultaneous clustering and feature discrimination with applications," in *Advances in Fuzzy Clustering and Feature Discrimination with Applications*, pp. 285–312, John Wiley & Sons, New York, NY, USA, 2007.

[5] B. S. Everitt, S. Landau, and M. Leese, *Cluster Analysis*, Oxford University Press, Oxford, UK, 2001.

[6] W. Pedrycz, *Knowledge Based Clustering*, John Wiley & Sons, Hoboken, NJ, USA, 2005.

[7] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York, NY, USA, 1981.

[8] R. Krishnapuram and J. M. Keller, "A possibilistic approach to clustering," *IEEE Transactions on Fuzzy Systems*, vol. 1, no. 2, pp. 98–110, 1993.

[9] M. R. Anderberg, *Cluster Analysis for Applications*, Academic Press, New York, NY, USA, 1972.

[10] F. D. A. T. de Carvalho, "Fuzzy c-means clustering methods for symbolic interval data," *Pattern Recognition Letters*, vol. 28, no. 4, pp. 423–437, 2007.

[11] G. Beliakov and M. King, "Density based fuzzy c-means clustering of non-convex patterns," *European Journal of Operational Research*, vol. 173, no. 3, pp. 717–728, 2006.

[12] M. Barni, V. Cappellini, and A. Mecocci, "Comments on "a possibilistic approach to clustering"," *IEEE Transactions on Fuzzy Systems*, vol. 4, no. 3, pp. 393–396, 1996.

[13] H. Timm, C. Borgelt, C. Doring, and R. Kruse, "Fuzzy cluster analysis with cluster repulsion," in *Proceedings of the European Symposium in Intelligent Technologies*, Tenerife, Spain, 2001.

[14] H. Timm and R. Kruse, "A modification to improve possibilistic fuzzy cluster analysis," in *Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE '02)*, pp. 1460–1465, Honolulu, Hawaii, USA, May 2002.

[15] H. Timm, C. Borgelt, C. Doring, and R. Kruse, "An extension to possibilistic fuzzy cluster analysis," *Fuzzy Sets and Systems*, vol. 147, no. 1, pp. 3–16, 2004.

[16] D. E. Gustafson and W. C. Kessel, "Fuzzy clustering with a fuzzy covariance matrix," in *Proceedings of the IEEE Conference on Decision and Control including the 17th Symposium on Adaptive Processes*, pp. 761–766, San Diego, Calif, USA, January 1979.

[17] N. R. Pal, K. Pal, and J. C. Bezdek, "A mixed c-means clustering model," in *Proceedings of the 6th IEEE International Conference on Fuzzy Systems*, vol. 1, pp. 11–21, Barcelona, Spain, July 1997.

[18] M.-S. Yang, K.-L. Wu, J.-N. Hsieh, and J. Y. Hsieh, "Alpha-cut implemented fuzzy clustering algorithms and switching regressions," *IEEE Transactions on Systems, Man, and Cybernetics Part B: Cybernetics*, vol. 38, no. 3, pp. 588–603, 2008.

[19] J. Yao, M. Dash, S. T. Tan, and H. Liu, "Entropy-based fuzzy clustering and fuzzy modeling," *Fuzzy Sets and Systems*, vol. 113, no. 3, pp. 381–388, 2000.

[20] S. Chatzis and T. Varvarigou, "Factor analysis latent subspace modeling and robust fuzzy clustering using t-distributions," *IEEE Transactions on Fuzzy Systems*, vol. 17, no. 3, pp. 505–517, 2009.

[21] K. Honda, A. Notsu, and H. Ichihashi, "Fuzzy PCA-guided robust k-means clustering," *IEEE Transactions on Fuzzy Systems*, vol. 18, no. 1, pp. 67–79, 2010.

[22] K. T. Atanassov, "Intuitionistic fuzzy sets," *Fuzzy Sets and Systems*, vol. 20, no. 1, pp. 87–96, 1986.

[23] K. T. Atanassov and G. K. Gargov, "Intuitionistic fuzzy logic," *Computing Research Academy of Bulgarian Sciences*, vol. 43, no. 3, pp. 9–12, 1990.

[24] K. Atanassov and C. Georgiev, "Intuitionistic fuzzy prolog," *Fuzzy Sets and Systems*, vol. 53, no. 2, pp. 121–128, 1993.

[25] E. Szmidt and J. Kacprzyk, "Intuitionistic fuzzy sets in group decision making," *Notes on Intuitionistic Fuzzy Sets*, vol. 2, no. 1, pp. 15–32, 1996.

[26] S. K. De, R. Biswas, and A. R. Roy, "An application of intuitionistic fuzzy sets in medical diagnosis," *Fuzzy Sets and Systems*, vol. 117, no. 2, pp. 209–213, 2001.

[27] K. T. Atanassov, "New operations defined over the intuitionistic fuzzy sets," *Fuzzy Sets and Systems*, vol. 61, no. 2, pp. 137–142, 1994.

[28] K. T. Atanassov, *Intuitionistic Fuzzy Sets: Theory and Applications*, vol. 35 of *Studies in Fuzziness and Soft Computing*, Physica, Heidelberg, Germany, 1999.

[29] K. Atanassov and G. Gargov, "Interval valued intuitionistic fuzzy sets," *Fuzzy Sets and Systems*, vol. 31, no. 3, pp. 343–349, 1989.

[30] H. Bustince, F. Herrera, and J. Montero, *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models*, Physica, Heidelberg, Germany, 2007.

[31] H. M. Zhang, Z. S. Xu, and Q. Chen, "Clustering approach to intuitionistic fuzzy sets," *Control and Decision*, vol. 22, no. 8, pp. 882–888, 2007.

[32] Z. S. Xu, J. Chen, and J. J. Wu, "Clustering algorithm for intuitionistic fuzzy sets," *Information Sciences*, vol. 178, no. 19, pp. 3775–3790, 2008.

[33] Z. Xu, "Intuitionistic fuzzy hierarchical clustering algorithms," *Journal of Systems Engineering and Electronics*, vol. 20, no. 1, pp. 90–97, 2009.

[34] E. Szmidt and J. Kacprzyk, "Distances between intuitionistic fuzzy sets," *Fuzzy Sets and Systems*, vol. 114, no. 3, pp. 505–518, 2000.

[35] Z. Xu, "Some similarity measures of intuitionistic fuzzy sets and their applications to multiple attribute decision making," *Fuzzy Optimization and Decision Making*, vol. 6, no. 2, pp. 109–121, 2007.

[36] P. Burillo and H. Bustince, "Entropy on intuitionistic fuzzy sets and on interval-valued fuzzy sets," *Fuzzy Sets and Systems*, vol. 78, no. 3, pp. 305–316, 1996.

[37] Z. S. Xu, "Some similarity measures of intuitionistic fuzzy sets and their applications to multiple attribute decision making," *Fuzzy Optimization and Decision Making*, vol. 6, no. 2, pp. 109–121, 2007.

[38] N. R. Pal, K. Pal, J. M. Keller, and J. C. Bezdek, "A possibilistic fuzzy c-means clustering algorithm," *IEEE Transactions on Fuzzy Systems*, vol. 13, no. 4, pp. 517–530, 2005.

[39] K. Ito and K. Kunisch, *Lagrange Multiplier Approach to Variational Problems and Applications*, SIAM Advances in Design and Control, SIAM, Philadelphia, Pa, USA, 2008.

[40] E. N. Nasibov and G. Ulutagay, "A new unsupervised approach for fuzzy clustering," *Fuzzy Sets and Systems*, vol. 158, no. 19, pp. 2118–2133, 2007.

[41] A. Asuncion and D. J. Newman, *UCI Machine Learning Repository*, Department of Information and Computer Science, University of California, Irvine, Calif, USA, 2007.

[42] J. Alcalá-Fdez, A. Fernández, J. Luengo et al., "KEEL datamining software tool: data set repository, integration of algorithms and experimental analysis framework," *Journal of Multiple-Valued Logic and Soft Computing*, vol. 17, no. 2-3, pp. 255–287, 2011.

[43] R. Kohavi and F. Provost, "Glossary of terms," *Machine Learning*, vol. 30, pp. 271–274, 1998.

Hindawi Publishing Corporation
Advances in Fuzzy Systems
Volume 2015, Article ID 247069, 6 pages
http://dx.doi.org/10.1155/2015/247069

Research Article

Provenance Study of the Terracotta Army of Qin Shihuang's Mausoleum by Fuzzy Cluster Analysis

Rongwu Li[1] and Guoxia Li[2]

[1]*Department of Physics, Beijing Normal University, Beijing 100875, China*
[2]*Faculty of Physical Engineering, Zhengzhou University, Zhengzhou 450052, China*

Correspondence should be addressed to Rongwu Li; rongwuli@bnu.edu.cn

Received 29 August 2014; Revised 8 October 2014; Accepted 9 October 2014

Academic Editor: Ferdinando Di Martino

20 samples and 44 samples of terracotta warriors and horses from the 1st and 3rd pits of Qin Shihuang's Mausoleum, 20 samples of clay near Qin's Mausoleum, and 2 samples of Yaozhou porcelain bodies are obtained to determine the contents of 32 elements in each of them by neutron activation analysis (NAA). The NAA data are further analyzed using fuzzy cluster analysis to obtain the fuzzy cluster trend diagram. The analysis shows that the origins of the raw material of the terracotta warriors and horses from 1st and 3rd pits are not exactly the same but are closely related to the loam soil layer near Qin's Mausoleum while distant from the loess layers in the same area and remotely related to the Yaozhou porcelain bodies. It can be concluded that the raw material of the terracotta warriors and horses was taken from certain loam layer near Qin's Mausoleum and the kiln sites might be located nearby.

1. Introduction

The companion burial pits of terracotta army of Qin Shi Huang (the first emperor of China), which was begun to be built in 247 BC and expanded nearly 40 years, were discovered in 1970s. Ever since then, they have been well known all over the world because of their long-standing history and their gigantic scale. There are over 7,000 life size terracotta warriors, chariots, and horses in these pits to guard Qin's Mausoleum. They are arranged in certain rules in three pits (numbered 1st, 2nd, and 3rd, resp.) to the east of Qin's Mausoleum. Experts name the 1st pit as Zhen (line-up), meaning troops charging forward. Experts name the 3rd pit as Mu (curtain), meaning commanding post [1]. But where is the raw material of the terracotta warriors and horses taken from? Where are the kiln sites in which the terracotta warriors and horses were fired located? In order to solve these problems long puzzling archaeologists and scholars [2, 3], we use nuclear analysis technologies to study the origin of the raw material of the terracotta warriors and horses in Qin's Mausoleum in this paper.

The contents of microelements in the clay in an area are generally constant over extended period of time, and they are generally not affected by pottery production process. Therefore, the contents of microelement can be used as origin indicator for raw material of potteries. Many techniques, such as proton-induced X-ray emission (PIXE) or NAA, can be used to determine contents of microelements. These techniques have been used to study the origins of large amount ancient ceramics [4–8]. NAA technique has the advantage that it can determine contents of more than 30 elements of any given samples simultaneously. In this paper, we use NAA to determine the contents of 32 elements in each sample. The data are further analyzed using fuzzy cluster analysis. Such analysis reveals critical information about the raw material origins of Qin's terracotta warriors and horses.

2. Measuring Methods and Results

2.1. Samples. As listed in Table 1, nineteen terracotta warrior shards and one terracotta horse shard from the 1st pit and nine terracotta warrior shards and thirty-five terracotta horse shards from the 3rd pit are selected as samples.

In order to study the relationship between Qin's terracotta warriors and the clay near Qin's Mausoleum, twenty soil samples are taken from different areas at different layers.

Table 1: Samples of terracotta warriors and horses from the 1st and 3rd pits of Qin's Mausoleum.

Site	Sample code	Sample name
1st pit	Q101 Q102 Q103 Q104 Q105 Q106 Q107 Q108	Terracotta warrior shard
	Q109	Terracotta horse shard
	Q110 Q111 Q112 Q113 Q114 Q115 Q116 Q117 Q118 Q119 Q120	Terracotta warrior shard
3rd pit	Q302 Q304 Q306 Q311 Q313 Q315 Q316 Q317 Q320	Terracotta warrior shard
	Q325 Q326 Q327 Q328 Q329 Q330 Q331 Q332 Q333 Q334 Q335 Q336	Terracotta horse shard
	Q337 Q338 Q339 Q340 Q341 Q342 Q343 Q344 Q345 Q346 Q347 Q348	Terracotta horse shard
	Q350 Q358 Q362 Q364 Q368 Q370 Q371 Q373 Q374 Q377 Q379	Terracotta horse shard

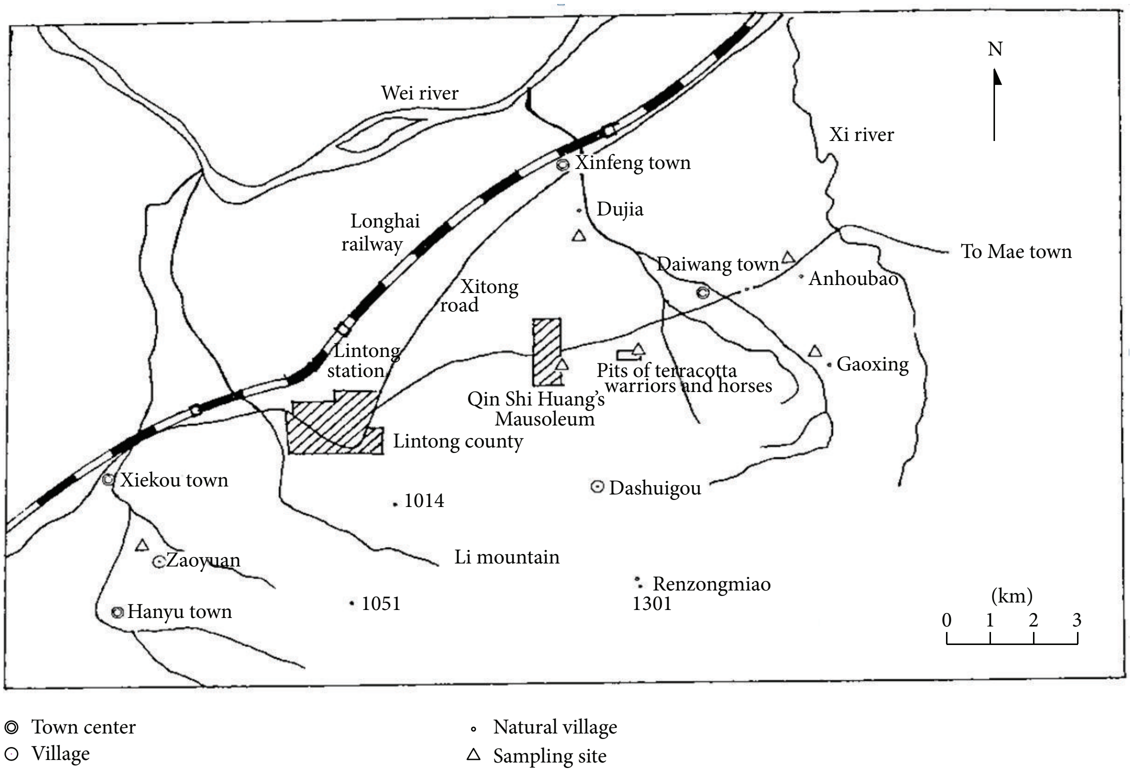

Figure 1: The map of sampling sites near Qin's Mausoleum.

The sampling areas are to the north of Lishan Hill and to the south of Wei River, within 10 KM range of Qin's Mausoleum. As shown in Figure 1, the sampling sites are located to the west, east, and north of Qin's Mausoleum. The other clay samples are taken from the sealing earth of Qin's Mausoleum and 2nd pit. Table 2 lists all the samples taken. In addition, two samples (Y3b and Y6b) of Yaozhou porcelain bodies are also included as contrasting samples.

2.2. NAA Experiments and Results. In these experiments, the samples of terracotta warriors and the clays near Qin's Mausoleum are cut and grounded into fine powder and then oven-dried at 80°C for 8 hours. Each sample weighs 30 mg and is wrapped with double-layered highly purified aluminum foil. In the meantime, 2 standard reference matter samples (code GBW07104 rock and code GBW07406 soil) are taken, 20 mg each. All samples and standard reference matter are put into a radiation jar and irradiated for 8 hours in the heavy water reactor in the Institute of Chinese Atomic Energy Science. The neutron radiant flux is about (3~7) × 10^{13} n·cm^{-2} s^{-1}. The samples are measured for the γ-ray intensity for the first time by the highly pure germanium γ-ray spectrum instrument at the Institute of High Energy Physics of Chinese Academy of Science after being cooled for 7 days. They are measured again for the second time after being cooled for 15 days. 32 elements are identified in the samples and the content of each element is measured after comparing with that of the standard reference matters using neutron activation analysis program. The confidence level of the NAA data is 90% with its unit being μg/g. Among these elements, 9 are rare earth elements, namely, La, Ce, Nd, Sm, Eu, Tb, Ho, Yb, and Lu, while the remaining elements are Na,

Table 2: The samples information of clay near Qin's Mausoleum.

Sample code	Sample name	Sampling site	Location and direction with respect to Qin's Mausoleum (KM)	Depth beneath earth (M)	Characteristics of the soil
QL01	Sealing earth of Qin's Mausoleum	Qin's Mausoleum	Eastern part of Qin's mausoleum	4.0~5.0	Light earth yellow
LZ01	Loess	Zaoyuan, Lintong	West south, 9.5	5	Light earth yellow
LZ02	Black loam	Zaoyuan, Lintong	West south, 9.5	6	Dark yellow
LZ03	Red loam	Zaoyuan, Lintong	West south, 9.5	7	Dark red
LZ04	Black loam	Zaoyuan, Lintong	West south, 9.5	8	Same as LZ02
LZ05	Loess	Zaoyuan, Lintong	West south, 9.5	10	Light earth yellow
LB01	Black loam	Gaoxing, Lintong	East, 5.5	1	Dark yellow
LB02	Loess	Gaoxing, Lintong	East, 5.5	2	Light earth yellow
LB03	Black loam	Gaoxing, Lintong	East, 5.5	6	Same as LZ02
LB04	Red loam	Gaoxing, Lintong	East, 5.5	6.5	Dark red
LB05	Loess	Gaoxing, Lintong	East, 5.5	10	Light earth yellow
LX01	Black loam	Dujia, Lintong	North, 2.5	0.5	Dark yellow
LA01	Black loam	Anhoubao, Lintong	East north, 5.0	1.5	Dark yellow
LA02	Black loam	Anhoubao, Lintong	East north, 5.0	1.9	Dark yellow
LA03	Black loam	Anhoubao, Lintong	East north, 5.0	2.5	Dark yellow
LA04	Black loam	Anhoubao, Lintong	East north, 5.0	3.3	Dark yellow
QK21	Backfill	2nd pit	East, 1.5		
QK22	Tamping	2nd pit	East, 1.5		
QK23	Black loam	2nd pit	East, 1.5		
QK24	Loess	2nd pit	East, 1.5		Light earth yellow

K, Ca, Sc, Cr, Fe, Co, Ni, Zn, As, Se, Rb, Sr, Zr, Mo, Sb, Cs, Ba, Hf, Ta, W, Th, and U. Of all 32 elements, Na, K, Ca, and Fe are common, and the remaining 28 are microelements.

2.3. Fuzzy Cluster Analysis Results. Cluster analysis is a generic term for a wide range of numerical methods for examining multivariate data with a view to uncovering or discovering groups or clusters of homogeneous observations. Clustering techniques have been employed in a remarkable number of different disciplines. In archaeology clustering has been used to investigate the relationship between various types of archaeological samples. In fuzzy clustering, objects are not assigned to a particular cluster: they possess a membership function indicating the strength of membership in all or some of the clusters. Memberships can be scaled to lie between 0 and 1 and can then be interpreted as probabilities. The fuzzy cluster analysis is based on fuzzy mathematics. It uses fuzzy matrix to define concepts, to discover rules, and to establish models. The NAA data of the 32 elements in the 86 samples are analyzed by fuzzy cluster analysis. Figure 2 is the trend fuzzy cluster analysis diagram. In this diagram, each sample belongs to a single cluster, and the complete set of clusters contains all samples. In some circumstances, however, overlapping clusters may provide a more acceptable solution. It should be noted that one acceptable answer from a cluster analysis is that no grouping of the data is justified. Based on the value of confidence level λ, the classification of the samples can be different. Nevertheless, each different classification based on different λ is useful in the origin study. There is an optimum confidence level in every classification.

The basic data for cluster analysis is the usual $n \times p$ multivariate (two-mode) data matrix, X, containing the variable values describing each object to be clustered; that is,

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & \cdots & \cdots & x_{np} \end{bmatrix}. \tag{1}$$

Entry x_{ij} in X gives the value of the jth variable on object i.

Of central importance in attempting to identify clusters of observations which may be present in data is knowledge of how "close" individuals are to each other or how far apart they are. Two individuals are "close" when their dissimilarity or distance is small or their similarity is large. Proximities can be determined either directly or indirectly, and the latter is more common in most applications.

Indirect proximities are usually derived from the $n \times p$ multivariate (two-mode) matrix, X. There are a vast range of possible proximity measures.

In this study, when λ is set to 0.886, all samples are classified into 5 categories according to the trend fuzzy cluster analysis diagram as follows.

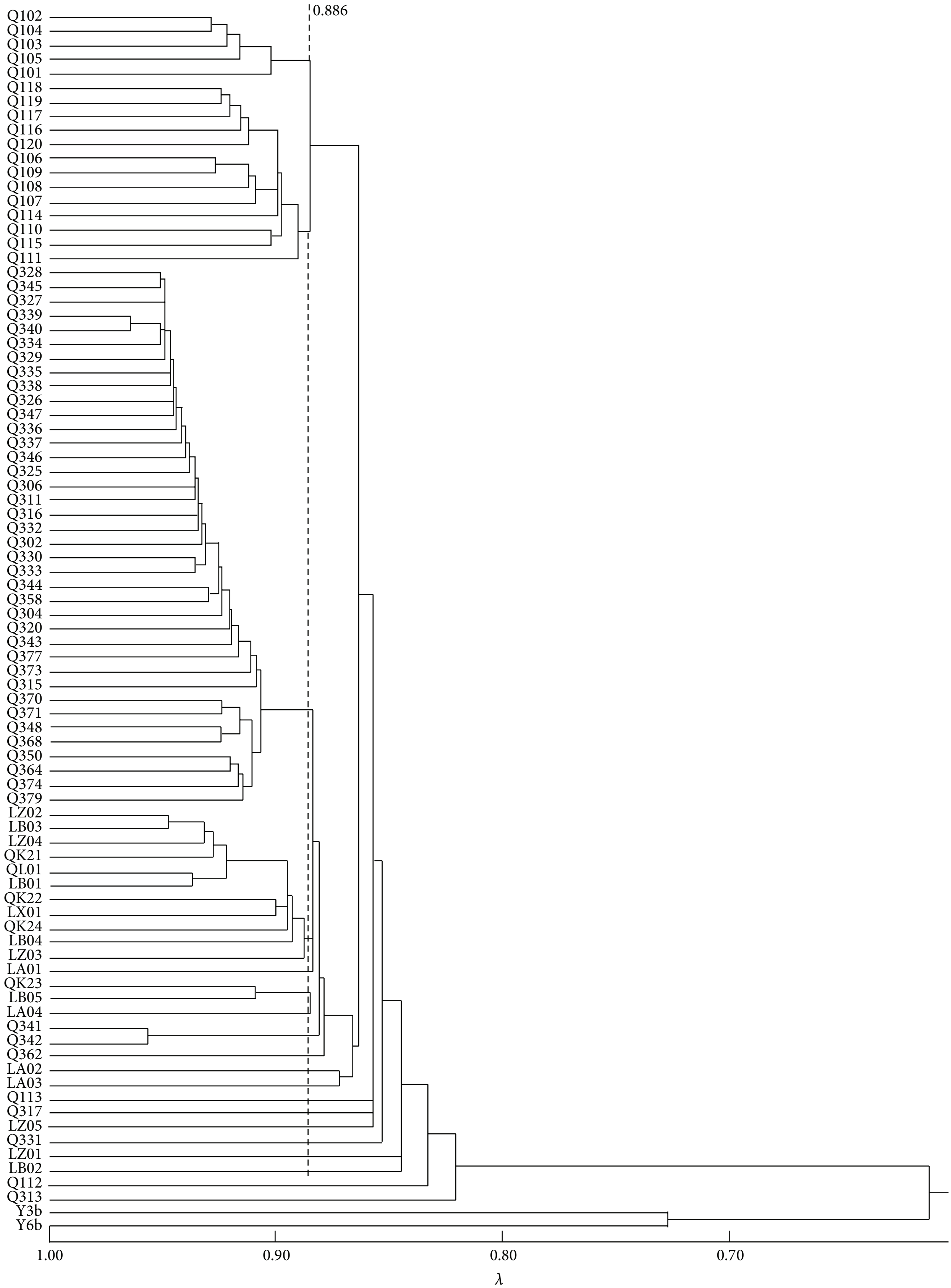

FIGURE 2: Trend fuzzy cluster analysis diagram for samples of terracotta warriors in Qin's Mausoleum 1st and 3rd pits and nearby clay.

As shown in Figure 2, the 1st category (Q102 to Q111) includes 18 samples from 1st pit. Among them, Q102, Q104, Q103, and Q105 are more closely related as they can be classified into one category when $\lambda = 0.915$. They are mostly shards of terracotta warrior's robes. Q118, Q119, Q117, Q116, and Q120 are more closely related as they can be classified into the same category when $\lambda = 0.913$. They are all shards of terracotta warriors. The remaining 9 samples are not so closely related to them. This demonstrates that the raw material origins in the 1st pit are more diversified and the kilns for firing the terracotta warriors of the 1st pit might also be quite scattered around.

The 2nd category (Q328 to Q379) includes 38 samples. They include most samples from the 3rd pit. These samples are closely related which shows the origins of the raw material for the 3rd pit are quite concentrated and are quite independent of the 1st pit.

The 3rd category (LZ02 to LA01) includes 12 clay samples near Qin's Mausoleum. Among these samples, LZ02, LZ04 (the black loam 6 M and 8 M beneath the earth in Zaoyuan and Lingtong, resp.), LB01, LB03 (the black loam 1 M and 6 M beneath the earth in Gaoxing and Lingtong, resp.), QK21 (the backfill soil from the 2nd pit), and QL01 (the sealing earth of Qin's Mausoleum) are very closely related. These six samples can be classified into one category when $\lambda = 0.922$. QL01, QK21, and QK22 (tamping earth) are regarded as the soil samples of the Qin or near Qin Dynasty. The samples that are closely related to them are black loam at various depths at Zaoyuan, Gaoxing, and Dujia in Lintong. This means the sealing earth of Qin's Mausoleum, the backfill earth, and tamping earth must be taken from these places or other places near Qin's Mausoleum where there are similar loam layers. In this category, QK24 is the only loess sample. It might have been contaminated either by nature causes or by human.

When $\lambda = 0.884$, the 2nd and 3rd categories above can be merged into one category. This means most samples from the 3rd pit are closely related to the loams from different depths at Zaoyuan, Gaoxing, and Dujia and the soil layer 1.5 meter under the earth at Anhoubao.

The 4th category (QK23 to Q313) includes 16 samples. When $\lambda = 0.818$, the 16 samples fall into the same category. These samples include the loess of different depths from Zaoyuan and Gaoxing, the loam of different depths from Anhoubao, and some samples from pits number 1 and number 3. This category is comparatively complicated and it virtually consists of 16 samples which are not closely related to one another. The loess layers could be regarded as the representative of this category.

All the samples above merged into one category when λ was set to 0.821.

Fifth category (Y3b, Y6b) includes 2 samples. They are all Yaozhou porcelain body samples. They are quite distant because they come from different kilns and have different mineral raw material origins. They are merged into one category when $\lambda = 0.727$. They have no relation with the terracotta warriors and horses and the clays near Qin's Mausoleum.

3. Conclusion

We use fuzzy cluster analysis to analyze the NAA data for the 1st and 3rd pits of Qin's Mausoleum and nearby clay samples. Our study shows the following.

(i) The raw materials of the 1st pit are quite diversified. Therefore, the kilns for firing the terracotta warriors are quite scattered around.

(ii) The raw material origins of most 3rd pit terracotta warrior samples are very concentrated. The kilns for firing these terracotta warriors are limited to a small number of concentrated ones.

(iii) Samples from the 1st pit are relatively independent of those from the 3rd pit. The origins of their raw material are not exactly the same.

(iv) The samples from the 1st and 3rd pits are closely related to the loam samples near Qin's Mausoleum, while quite distant to loess samples near Qin's Mausoleum and very distant to Yaozhou porcelain body.

(v) The raw material for making terracotta warriors of the 1st and 3rd pits might be taken from loam layer at different depths at Zaoyuan, Gaoxing, Dujia, and Anhoubao in Lintong or another loam layer that has similar soil properties near Qin's Mausoleum. Their raw material origins should be places near Qin's Mausoleum. Therefore, the kilns for firing the terracotta warriors of the 1st and 3rd pits should be located near Qin's Mausoleum.

4. Discussion

NAA technology can be used to measure tens of elements in given samples. Fuzzy cluster analysis on NAA data can give clear, objective, and comprehensive results. As compared to other technologies, combing these two methods has unique advantage in studying raw material origins of ancient ceramics.

As reported by archeologists of Qin's Mausoleum terracotta warriors and horses, most of the parts of the terracotta warriors and horses were made from mold first and then glued together with clay paste, which was made of water-washed clay and mixed with silver sand [9]. Nevertheless, the microelements contents of the terracotta warriors still preserve information about their origin. However, the samples in this study are taken from the shards of Qin terracotta warriors and horses. There is no information about which terracotta warriors and horses or which parts these samples belong to and the sample number is very limited. More samples are needed to enhance the confidence level of such analysis. It is very useful to set up an NAA database of Qin's terracotta warriors and nearby clay, with samples including different coded terracotta warriors and horses, as well as clay at different depths and places near Qin's Mausoleum. Such database can greatly enhance the research on the raw material origins of Qin's terracotta warriors, their firing kiln sites, and their craftsmanship.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work is supported by National Natural Science Foundation of China (51172212), Fundamental and Frontier Technology Research Projects in Henan province (102300410168), and Nuclear Analysis Technology Open Laboratory of Chinese Academy of Science (B0901).

References

[1] X. L. Wang, *Focus Study on Qin's Mausoleum*, San Qin Press, 1994, (Chinese).

[2] G. Q. Jin, X. J. Pan, and Z. T. Sun, *Physical Archaeology*, Shanghai Science & Technology Press, 1989, (Chinese).

[3] J. Z. Li, *Science and Technology History, Ceramics Volume*, Beijing Science & Technology Press, 1998 (Chinese).

[4] Z. Q. Zhang, H. S. Cheng, H. N. Xia, J. C. Jiang, M. H. Gao, and F. J. Yang, "Trace elements measurement by PIXE in the appraisal of the ancient potteries," *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, vol. 190, no. 1–4, pp. 497–500, 2002.

[5] J. D. Robertson, H. Neff, and B. Higgins, "Microanalysis of ceramics with PIXE and LA-ICP-MS," *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, vol. 189, no. 1–4, pp. 378–381, 2002.

[6] T. Calligaro, A. Mossmann, J.-P. Poirot, and G. Querré, "Provenance study of rubies from a Parthian statuette by PIXE analysis," *Nuclear Instruments and Methods in Physics Research Section B*, vol. 136–138, pp. 846–850, 1998.

[7] Z. Gao, J. Wang, S. Chen et al., "Provenance investigation on ancient Chinese Ru porcelains by neutron activation analysis," *Atomic Energy Science and Technology*, vol. 31, no. 4, pp. 360–364, 1997 (Chinese).

[8] W. J. Zhao, Z. Y. Gao, and S. H. Chen, "Neutron activation analysis of the sources of raw material of porcelain made in official kilns in the Southern Song Dynasty," *Archaeology*, vol. 7, pp. 80–85, 1998 (Chinese).

[9] "The excavation team of the terracotta army of Qin's Mausoleum," *Archaeo and Cultural Relics*, vol. 3, pp. 108–119, 1980.

Hindawi Publishing Corporation
Advances in Fuzzy Systems
Volume 2015, Article ID 841485, 14 pages
http://dx.doi.org/10.1155/2015/841485

Research Article

A Fuzzy Supplier Selection Application Using Large Survey Datasets of Delivery Performance

Jonathan Davis, Margaret F. Shipley, and Gary Stading

The University of Houston-Downtown, 326 N Main Street Houston, TX 77002, USA

Correspondence should be addressed to Jonathan Davis; davisg@uhd.edu

Received 27 August 2014; Revised 10 October 2014; Accepted 13 October 2014

Academic Editor: Ferdinando Di Martino

A model is developed using fuzzy probability to screen survey data across relevant criteria for selecting suppliers based on fuzzy expected values. The values are derived from qualitative variables and expert opinion of membership in these variables found in industry survey data. The application is made to a supply chain management decision of supplier selection based upon delivery performance which is further divided into attributes that comprise this criterion. The algorithm allows multiple criteria to be considered for each decision parameter. Large sets of survey data regarding six suppliers in the electronic parts industry are gathered from over 150 purchasers and are analyzed through spreadsheet modeling of the fuzzy algorithm. The resulting decision support system allows supply chain managers to improve supplier selection decisions by applying fuzzy measures of criteria and associated beliefs across the dataset. The proposed model and method are highly adaptable to existing survey datasets, including datasets that have incomplete data, and can be implemented in organizations with low decision support resources, such as small and medium sized organizations.

1. Introduction

Selecting suppliers has become an area of increasing study due to its importance for establishing long-term channel relationships. Firms have become highly selective of their suppliers but suffer in this process because existing evaluative measures often do not rank suppliers on a relative basis. Thus, of importance is that the firm be able to identify suppliers through an effective evaluation process [1]. Yet this supplier selection evaluation process is a complex decision space with overlapping, complementary, and often contradictory selection criteria. Coupled with the ambiguity in selection criteria, uncertainty in such a decision exists since all relevant attributes may not be identified or even identifiable.

Many factors contribute to the difficulty of rank ordering in the supplier evaluation process. Selection criteria and scoring can occur at various levels of the organization, which leads to conflicts in scoring, particularly in systems built to reflect qualitative criteria. Assessors operating in these systems often experience inherent uncertainty regarding supplier performance due to lack of a proper anchoring or definition of numerical scores. This ambiguity means that a supplier's score on a criterion may not be determinant. Assessors rely on statistical analysis that provides the highest probability of achieving a qualitatively defined membership, yet membership of a supplier in the category of "good," for example, may vary in opinion over time and by different assessors.

Scoring models have been used in various aspects of supply chain management for many years because supplier selection lends itself to models incorporating management opinion [2, 3]. However, decision support systems can benefit from improvements that bring clarity to otherwise vague scoring criteria [4–6], because such improvements enlarge the available sets of useful data these systems can use. Given the benefit of using expert judgment from management and professionals, a fuzzy set based model that has a proven record of handling uncertainty of the decision maker and ambiguity of data while allowing these experts to assess membership across several criteria can be an effective evaluation tool.

Indeed, extant methods that use classical logic or statistics have been shown to be inadequate for effectively dealing with situations where limited information is available. Behavioral and expected utility theorists have argued that probability theory and other traditional quantitative techniques are not equipped to consider the uncertainty that exists in most personal judgments [7, 8]. Among alternative techniques, the Dempster-Shafer theory of evidence [9–12] provides useful measures for the evaluation of subjective certainty. Fuzzy set theory is among the most powerful tools for dealing with uncertainty where ambiguous terms are present [13–16]. Of particular interest to the development of the algorithm and the ensuing application are fuzzy probability distributions [17] and the concept of fuzzy expected values [18].

This research furthers the use of fuzzy logic models in supplier selection decisions. Background information is presented relative to traditional supplier selection methods and the general advancement of fuzzy logic-based models into supply chain management with more recent models targeting supplier selection decisions. Additional background literature is presented that identifies the traditional criteria for supplier selection. Next, fuzzy logic notation and model development lead to a spreadsheet application based on a survey of responses from a sample of professional buyers. Guidelines are introduced to resolve problems with the dataset regarding data insufficiency and/or incomplete responses, and the dataset is filtered to create a subset of the original dataset that is sufficiently populated with data. Results are then presented and analyzed.

2. Materials and Methods

As noted in the introduction, supplier selection decisions are receiving increasing focus in industry, and the associated academic papers have no shortage of ideas about how selection decisions can be made. Alternatives range from simple expert opinion methods that are not reliably repeatable to quantitative methods that may not account for various criteria associated with standards for supplier performance.

2.1. Supplier Selection Models. Recent supplier selection models have focused on using the analytic hierarchy process (AHP) or providing case study illustrations of decision making processes utilizing expert opinions [19–21]. Measuring the performance of suppliers [1] and stressing the importance of supplier selection criteria [2] for small firms in the United States [22] or large firms in Japan [23] reinforce that this initiative is not restricted to large firms and is global. Typical criteria used are price, delivery, and quality (PDQ) [1, 23] but Park and Krishnan [22] add several managerial criteria beyond PDQ including managerial forecasts, trust level, and organizational structure. Supplier selection research using decision support systems also extends supplier selection criteria in other dimensions including environmental considerations [6]. Weber et al. [24] considered other factors like facility location, capacity, and financial position. Ellram [25] discussed the importance of incorporating these and other decision criteria in the process of selecting suppliers, and Simpson et al. [1] provide an extensive study of criteria perceived as most crucial to the assessment.

Using the traditional PDQ criteria, Verma and Pullman [26] provide an extensive study of how purchasing managers evaluate trade-offs among the criteria. They point out that while the PDQ criteria are generally accepted in industry, delivery and quality lend themselves to decision support criteria models because the complexity and multiple dimensions of delivery and quality can confound a decision maker. Of these criteria, quality tends to be extensively studied by researchers [21], while supplier selection and evaluation are most often based entirely on price [27, 28]. Thus, by elimination, delivery does not have nearly the research support.

In response to this dearth of research and in recognition of the global importance of supplier selection, four separate focus groups representing international and national geographical areas were used to identify attributes of delivery that would confound a purchasing decision. While many attributes were identified by the focus groups, those used for the "delivery" example in this study include the following: inventory availability, on-time deliveries, shipping accuracy, and a supplier's receptiveness to authorizing material returns (RMAs), as supported by Simpson et al. [1]. Given the previous justification of delivery as the focus criterion for this project with support by Vahdani and Zandieh [29] as to the strength of fuzzy set theory and outranking methods relative to the other decision support methodologies, fuzzy logic appeared appropriate for supplier selection decisions.

2.2. Fuzzy Logic in Supply Chain Management. As early as 1999, Liu [30] proposed a fuzzy model for partial backordering models. Little was done with inventory considerations until fully five years later when inventory discounting considered the buyer-seller relationships [31], and location aspects for inventory control became fuzzy considerations [32]. Supply chain decisions for integrated just-in-time inventory systems recognized the fuzzy nature of annual demand and production rates as being no longer statistically based. The supposition of known annual demand was considered by the authors to be unrealistic such that the proposed model included fuzzy annual demand and/or production rate, employing the signed distance, a ranking method for fuzzy numbers, to estimate fuzzy total cost of the JIT production in today's supply chain environment. A fuzzy-set based method derived the optimal buyer's quantity and number of lots from the vendor [33].

Some fuzzy set-based decision-making models appeared when fuzzy programming was used for the following: optimal product mix based on ABC analysis [34]; fuzzy multiobjective linear programming minimized total production and transportation costs; the number of rejected items and total delivery time as related to labor and budget constraints [35]; and fuzzy goal programming considered supply chain management from the perspective of activity-based costing with mathematically derived optimization for evaluating performance of the value-chain relationship [36]. Manufacturing processes as related to business logistics looked at the data itself as fuzzy in Quality Function Deployment's relationship to customer service [37]. The attainment of goals such as quality further led to attempts to balance production processes of assembly lines. Fuzzy goals were used as an instrument and product for measuring, displaying, and controlling industrial process variables [34].

Considering different quality standards in a supply chain network a fuzzy neural approach was utilized to suggest adjustments of product quantity from various suppliers [38]. The Fuzzy Suitability Index (FSI) aggregated rankings and multiplied, by weight, each criterion [39]. With the same goal of ranking suppliers according to performance, a method was proposed whereby n decision makers evaluated the performance of m suppliers in k criteria, rating the importance of the k criteria in linguistic terms. Aggregation of the fuzzy expressions for importance weights and a fuzzy preference index led to rank ordering of the suppliers [40].

Supplier selection was developed from a rule-based perspective. The approach selected was fuzzy associated rule mining from the database for supplier assessment [14]. Sevkli [41] in a comparison of a recognized crisp ELECTRE model versus a fuzzy ELECTRE model reached the conclusion that using fuzzy sets for multicriteria supplier selection decisions is superior.

2.3. Theory. The model developed herein uses fuzzy set theory and extension principles with fuzzy probabilities that determine a belief-weighted score for each attribute of each criterion. A revision to an existing algorithmic process [12] is defined with presentation and justification specific to this supplier selection decision-making problem based upon delivery attributes.

2.3.1. Fuzzy Logic Approach. Fuzzy logic addresses the ambiguity of data and uncertainty in this decision making situation, where a fuzzy subset A of a set X is a function of X into $[0, 1]$. For a brief foundation in the basics, see Zadeh [42], Bellman and Zadeh [13], Dubois and Prade [43], and Freeling [44]. While a new class of implication operators has been proposed [45], the more traditionally utilized fuzzy operations are used in this research. A and B denote two fuzzy sets, so the intersection, union, and complement are defined by

$$\begin{aligned} A \cap B &= \sum \frac{\gamma_i}{x_i}, \quad \text{where } \gamma_i = \operatorname{Min}\{\alpha_i, \beta_i\}; \\ A \cup B &= \sum \frac{\gamma_i}{x_i}, \quad \text{where } \gamma_i = \operatorname{Max}\{\alpha_i, \beta_i\}; \\ \neg A &= \sum \frac{\gamma_i}{x_i}, \quad \text{where } \gamma_i = 1 - \alpha_i; \end{aligned} \tag{1}$$

and it is assumed that $B = \sum \beta_i / x_i$ [46–49].

Extension principles [18, 43, 50] often guide the computations when dealing with fuzzy sets. Letting f be a function from X into Y, with Y as any set and A as above, then f can be extended to fuzzy subsets of X by

$$f(A) = \sum_y \frac{u_{f(A)}(y)}{y}, \quad \text{where } u_{f(A)}(y) = \operatorname*{Max}_{x \varepsilon f^{-1}(y)} A(x). \tag{2}$$

Thus, $f(A)$ is a fuzzy subset of Y. In particular, if f is a mapping from a Cartesian product such as $X \times Y$ to any set, Z, then f can be extended to objects of the form (A, B) where A and B are fuzzy subsets of X and Y by

$$\begin{gathered} f(A, B) = \sum \frac{u_{f(A,B)}(z)}{z}, \\ \text{where } u_{f(A,B)}(z) = \operatorname*{Max}_{(x,y)\varepsilon f^{-1}(z)} \operatorname{Min}\{A(x), B(x)\}. \end{gathered} \tag{3}$$

A fuzzy set P whose elements all lie on the interval $[0, 1]$ can be expressed as a fuzzy probability.

2.3.2. Fuzzy Probability Distributions. Consider a set of n fuzzy probabilities each having r elements,

$$a_i = \sum_{j=1}^{r} \frac{\alpha_{ij}}{a_{ij}} \quad \text{for } i = 1, 2, \ldots, n, \tag{4}$$

where α_{ij} denotes the degree of belief that a possible value of a_i is a_{ij}. Then $(a_1, a_2, \ldots, a_n)$ constitutes a finite fuzzy probability distribution if and only if there are n-tuples a_i, $i = 1, 2, \ldots, n$, such that $\sum_{i=1}^{n} a_i = 1$.

To qualify as a finite fuzzy probability distribution, each fuzzy probability in the distribution must have the same number of elements (some of the a's may be zero), and these elements should be ordered in the sense that the sum of the elements in each specific position must equal one. So the n-tuples (a_{ij}), $i = 1, 2, \ldots, n$, form probability distributions in the crisp sense. This type of probability distribution can be transformed such that the resulting distribution has entropy at least as great as the original [17].

2.3.3. Fuzzy Expected Value. A version of fuzzy expected values was first used when Zebda [18] defined $Q_{ijk} = \sum \alpha_{ijk} / a_k$ as the fuzzy probability that, from State i and making Decision j, reaches State k. Associated with this are fuzzy benefits B_{ijk}, where $B_{ijk} = \sum \beta_{ijk} / b'_k$.

Then the averaged benefit is defined by $E(B_{ijk}) = \sum c_{ij\ell} / b_\ell$ where

$$\begin{gathered} c_{ij\ell} = \operatorname*{Max}_{(a_1,\ldots,a_p,b'_1,\ldots b'_p)\varepsilon f^{-1} b_\ell k} \operatorname{Min}\left(\alpha_{ijk}, \beta_{ijk}\right) \\ \text{for } b_\ell = \sum_k a_x b_x \text{ if } \sum_k a_x = 1 \text{ and 0 otherwise.} \\ \text{Here, } f\left(a_1, \ldots, a_p, b'_1, \ldots, b'_p\right) = \sum a_x b'_x. \end{gathered} \tag{5}$$

2.3.4. Fuzzy Supplier Selection. In an earlier pilot study application [51], a fuzzy probability-based algorithm demonstrated a decision model based on a single rating for each supplier. But suppliers are usually not scored by only one kind of criterion, the attributes of which may vary in importance from one customer to another. Also, the application abbreviated the decision space considerably by disallowing suppliers that scored Below Average on an attribute, which works well for a single-attribute example but breaks down when applied across many attributes or criteria in a large dataset.

The proposed fuzzy algorithm recognizes that a supplier could score Below Average on an attribute that is relatively unimportant to the customer but still be rated highest overall if the supplier's other scores are sufficiently high in areas of greater importance to the customer. Since the many respondents are not equally confident in the scores they give, the algorithm makes use of the value of fuzzy probability with respect to each supplier measure and limits the discarding of a great deal of valuable data.

2.4. Algorithm. The algorithm preserves information during the process of computing and evaluating fuzzy probabilities until a final weighted model collapses the results into an objective score.

(0) Randomly partition the criteria dataset into ℓ subsets of equal size.

(1) For each attribute ϕ of each supplier ν, subjectively assign scores $s_{\phi k\nu}$. The supplier rating ($s_{\phi k\nu}$) is then given by the equation $s_{\phi k\nu} = \sum \tau_{\phi k}/s_{\phi k}$ for all ν where $\tau_{\phi k} = 1$ ($\nu = 1, 2, \ldots, m$; $k = 1, 2, \ldots, n$; and $1 < \phi < x$).

(2) Define the fuzzy expected value, $Q_{\phi k\nu}$, for each attribute ϕ of each ν in terms of each $s_{\phi k\nu}$ as $Q_{\phi kj\nu} = \sum \alpha_{\phi kj\nu}/a_{\phi kj\nu}$ for all $s_{\phi kj\nu}$, where each $\alpha_{\phi kj\nu}$ represents belief in the probability $a_{\phi kj\nu}$ that ν will be scored $s_{\phi kj\nu}$ ($\nu = 1, 2, \ldots, m$; $k = 1, 2, \ldots, n$; $1 < \phi < x$ and $j = 1, 2, \ldots, \ell$).

(3) Group the probabilities $a_{\phi kj\nu}$ into combinations $\varphi_{\phi\nu}$ such that $\sum a_{\phi kj\nu} = 1$ for some set H of k's. $a_{\phi kj\nu} = 0$ for $k \notin H$.

(4) Across all partitions ℓ, compute $b_{\phi\nu} = \sum a_{\phi kj\nu} s_{\phi kj\nu}$ if $\sum a_{\phi kj\nu} = 1$, otherwise 0 ($k = 1, 2, \ldots, r$; $j = 1, 2, \ldots, \ell$ and p = the distinct number of $\sum a_{\phi kj\nu} = 1$; $1 < \ell \leq p$).

(5) For all $\alpha_{\phi kj\nu} \neq 0$ find $c_{\phi\nu} = \mathrm{Min}\{\tau_{\phi kj\nu}, \alpha_{\phi kj\nu}\}$, where $c_{\phi\nu}$ is the degree of belief that the expected value is $b_{\phi\nu}$.

(6) Defuzzify the expected value for each attribute ϕ to find $E(s_{\phi\nu}) = \sum c_{\phi\nu} b_{\phi\nu} / \sum c_{\phi\nu}$.

An illustration of this supplier selection modeling algorithm is presented next.

2.5. Algorithm Methods and Illustration. The application presents a real-world supplier selection decision-making problem based upon (1) generation of data from a survey of purchasing professionals and (2) partitioning of the resulting data to fit the algorithm detailed above.

2.5.1. Example Data. The algorithm example uses results from a survey instrument built with input from industry expert focus groups. The subsequent survey measures customer ratings of a group of suppliers for various variables including delivery attributes. The survey was distributed to about 3,000 companies that purchase semiconductors, passives, RF/microwaves, connectors and interconnects, and electromechanical devices from a small set of dominant suppliers. Representative industries included automotive, communications, contract design/engineering, power/electrical, medical/dental, computer, manufacturing, and military/aerospace. The survey queried each customer's number of years of activity in the industry in designated ranges from less than two to 21 or more. Customers dealt with multiple suppliers and specified their firm's annual sales revenue as under \$5,000,000 to over \$2,000,000,000. With 406 surveys received, the response rate was slightly under 15%.

The survey was scored on a 0–5 Likert scale for seven suppliers: Arrow, Avnet, Future, Insight, Kent, Pioneer, and TTI. Respondents used the same scale for satisfaction with each supplier individually on the service performance attributes.

For model application purposes, the survey provided performance measurements on each supplier, as well as measures of the importance of each criterion to the customer and the customer's level of belief explicitly tied to the company's annual amount of business conducted with the targeted group of suppliers. Survey questions relating directly to the importance of this fuzzy supplier selection application included a query of the amount of money the customer spends on electronic components in a year: <\$100,000; \$100,000–\$499,999; \$500,000–\$999,999; \$1,000,000–\$9,999,999; \$10,000,000–\$24,999,999; and >\$25,000,000. These ranges were used to identify a firm's level of activity with the suppliers in question and, therefore, its expected level of confidence (interpreted as belief) in its assessments.

Given that real-world collection of survey data is, to varying degrees, imperfect, a threshold (10 data points) was established beyond which a respondent's data cluster would be considered too incomplete and removed. Furthermore, one supplier, Kent, was removed from the set due to (a) low survey responses compared to the other suppliers and (b) no longer existing as an independent company, having been acquired by Avnet after the survey was conducted. The resulting dataset left a pool of 150 responses considered "useable" to be applied to the fuzzy algorithm. These responses each had between 0 and 10 missing data points, which correspond to individual survey questions that went unanswered by the respondent. It is necessary to accept and adapt to this imperfect data in order to have a sufficiently large dataset to draw useful conclusions, given the nature of survey data.

The 150 acceptable survey responses resulting from the filtering of the data from the original 406 surveys were randomly partitioned into two sets of 75 responses each in accordance with Step 0 of the model algorithm. These respondents evaluated suppliers on the delivery-specific attributes: on-time performance, availability of inventory, shipping accuracy, and RMAs.

2.5.2. Fuzzy Preference Algorithm. Using the partitioned datasets provides two inputs into the defined fuzzy probability distributions.

By Step 1 of the algorithm, $\phi = 1, 2, 3, 4$ attributes as defined above. Each of the four attributes is subjectively assigned a score by the respondent for each of the six suppliers ($m = 6$), equating to Poor, Below Average, Average, Above Average, and Excellent ($n = 5$). Supplier rating $s_{\phi\nu}$ is then given by the equation $s_{\phi\nu} = \sum \tau_{\phi k\nu}/s_{\phi k\nu}$ for each supplier, ν, and, by Step 2, the fuzzy probability $Q_{\phi kj\nu}$ for each attribute of ν in terms of $s_{\phi kj\nu}$ is $Q_{\phi kj\nu} = \sum \alpha_{\phi kj\nu}/a_{\phi kj\nu}$ for all $s_{\phi\nu}$. Each $\alpha_{\phi kj\nu}$ represents belief in the probability $a_{\phi kj\nu}$ that ν will perform to the level of the assigned score $s_{\phi\nu}$ ($k = 1, 2, \ldots, 5$; $\nu = 1, 2, \ldots, 6$; $\phi = 1, 2, 3, 4$; and $j = 1, 2$).

The belief functions were populated based on a survey question indicating the amount of annual spending done by the respondent. Table 1 describes the scoring of respondent belief as proportional to total possible spending (conservatively assumed to be the low end of the top category, \$25,000,000) as shown in Table 1.

Beliefs were assigned for each respondent according to Table 1 and averaged across all respondents for each rating. After the assignment of belief, the algorithmic process for supplier ν shows four significant digits to make the method clear, although subsequent suppliers are rounded to two significant digits for readability and brevity. $Q_{\phi k\nu}$ is as follows:

$$\begin{aligned} Q_{111} &= 0.0000/0.0000 + 0.2500/0.0400 \\ &\qquad \text{for } s_{\phi k\nu} = s_{111} = \text{Poor}\,(1) \\ Q_{121} &= 0.1100/0.0533 + 0.3400/0.0800 \\ &\qquad \text{for } s_{121} = \text{Below Average}\,(2) \\ Q_{131} &= 0.3900/0.2267 + 0.3800/0.2133 \\ &\qquad \text{for } s_{131} = \text{Average}\,(3) \\ Q_{141} &= 0.3000/0.6133 + 0.3500/0.3600 \\ &\qquad \text{for } s_{141} = \text{Above Average}\,(4) \\ Q_{151} &= 0.2200/0.1067 + 0.1900/0.3067 \\ &\qquad \text{for } s_{151} = \text{Excellent}\,(5)\,. \end{aligned} \tag{6}$$

For Arrow, $s_{\phi\nu} = 1.0/1 + 1.0/2 + 1.0/3 + 1.0/4 + 1.0/5$.

In the case of 0.0 beliefs, the estimation of no likelihood (0.0) that the supplier's on-time delivery will rate Poor is because no respondents in partition one scored this supplier as Poor. Since one respondent in partition two did rate the company's delivery performance as Poor, and this respondent's sales revenue volume was in the middle range, there is a 0.51 belief that there is a 0.04 probability that Arrow's on-time delivery will be Poor. The highest beliefs (0.39 and 0.38) are for low probabilities (0.2267 and 0.2133) that the supplier has Average on-time performance. The highest probabilities (0.6133 and 0.3600) for Above Average have among the highest beliefs (0.3 and 0.35, respectively). While one group of respondents considered a 0.3067 probability of occurrence with 0.19 beliefs, the other group held an even higher belief for a low probability (0.1067) of Excellence in on-time delivery by Arrow.

Beliefs ($\alpha_{\phi kj\nu}$) and corresponding probabilities ($a_{\phi kj\nu}$) are then defined as

$$\begin{aligned} \alpha_{1111} &= 0.0000, \\ \alpha_{1211} &= 0.1100, \\ \alpha_{1311} &= 0.3000, \\ \alpha_{1411} &= 0.3000, \\ \alpha_{1511} &= 0.2200, \\ \alpha_{1121} &= 0.2500, \\ \alpha_{1221} &= 0.3400, \\ \alpha_{1321} &= 0.3800, \\ \alpha_{1421} &= 0.3500, \\ \alpha_{1521} &= 0.1900, \\ a_{1111} &= 0.0000, \\ a_{1211} &= 0.0533, \\ a_{1311} &= 0.2267, \\ a_{1411} &= 0.6133, \\ a_{1511} &= 0.1067, \\ a_{1121} &= 0.0400, \\ a_{1221} &= 0.0800, \\ a_{1321} &= 0.2133, \\ a_{1421} &= 0.3600, \\ a_{1521} &= 0.3067. \end{aligned} \tag{7}$$

Table 1: Respondent belief associated with spending.

Spending	Degree of belief
<\$100,000	0.0020
<\$500,000	0.0100
<\$1,000,000	0.0300
<\$10,000,000	0.2000
<\$25,000,000	0.7000
>\$25,000,000	1.0000

According to Step 3 of the algorithm, all combinations $\varphi_{\phi\nu}$ of the five scores across both data partitions are considered for each outcome that sums one, which in this example yields

$$\begin{aligned} \varphi_{\phi\nu 1} &= a_{1111} + a_{1211} + a_{1311} + a_{1411} + a_{1511} = 1.0, \\ \varphi_{\phi\nu 2} &= a_{1121} + a_{1221} + a_{1321} + a_{1421} + a_{1521} = 1.0. \end{aligned} \tag{8}$$

Possible n-tuples are (0.0000, 0.0533, 0.2267, 0.6133, 0.1067) and (0.0400, 0.0800, 0.2133, 0.3600, 0.3067). Following Step 4, a "weighted average" probability $b_{\phi\nu}$ for all $\varphi_{\phi\nu}$ is derived

$$\begin{aligned} b_{\phi\nu 1} &= (0.0000)(1) + (0.0533)(2) + (0.2267)(3) \\ &\quad + (0.6133)(4) + (0.1067)(5) = 3.7733, \\ b_{\phi\nu 2} &= (0.0400)(1) + (0.0800)(2) + (0.2133)(3) \\ &\quad + (0.3600)(4) + (0.3067)(5) = 3.8133. \end{aligned} \tag{9}$$

The minimum degree of belief in the on-time delivery then assessed according to Step 5 considers only the cases where belief is greater than zero:

$$c_{\phi\nu 1} = \text{Min}\{0.11, 1; 0.39, 1; 0.30, 1; 0.22, 1\} = 0.11,$$
$$c_{\phi\nu 2} = \text{Min}\{0.25, 1; 0.34, 1; 0.38, 1; 0.35, 1; 0.19, 1\} = 0.19. \quad (10)$$

Step 6 defuzzifies the expected score such that ν_1's expected fuzzy score for on-time delivery performance is

$$E(s_{11}) = \frac{[(0.1100)(3.7733) + (0.1900)(3.8133)]}{[0.1100 + 0.1900]} = 3.7987. \quad (11)$$

Applying the algorithm (with significant digit rounding) to the second supplier ν_2 (Avnet) yields

$$\begin{aligned}
Q_{112} &= 0.2/0.0 + 0.4/0.0 \quad \text{for } s_{112} = \text{Poor}\,(1)\\
Q_{122} &= 0.4/0.1 + 0.3/0.1 \quad \text{for } s_{122} = \text{Below Average}\,(2)\\
Q_{132} &= 0.4/0.3 + 0.4/0.3 \quad \text{for } s_{132} = \text{Average}\,(3)\\
Q_{142} &= 0.3/0.6 + 0.3/0.4 \quad \text{for } s_{142} = \text{Above Average}\,(4)\\
Q_{152} &= 0.2/0.1 + 0.1/0.2 \quad \text{for } s_{142} = \text{Excellent}\,(5).
\end{aligned} \quad (12)$$

Again, according to Step 3 of the algorithm, all combinations are considered for each outcome that sums one, which in this example yields $a_{\phi kj\nu}$ combinations:

$$\begin{aligned}
\varphi_{\phi\nu 1} &= a_{1111} + a_{1211} + a_{1311} + a_{1411} + a_{1511} = 1.0,\\
\varphi_{\phi\nu 2} &= a_{1121} + a_{1221} + a_{1321} + a_{1421} + a_{1521} = 1.0
\end{aligned} \quad (13)$$

as for supplier ν_1 (Arrow) but also

$$\begin{aligned}
\varphi_{\phi\nu 3} &= a_{1111} + a_{1221} + a_{1311} + a_{1411} + a_{1511} = 1.0,\\
\varphi_{\phi\nu 4} &= a_{1111} + a_{1211} + a_{1321} + a_{1411} + a_{1511} = 1.0,\\
\varphi_{\phi\nu 5} &= a_{1121} + a_{1221} + a_{1321} + a_{1421} + a_{1521} = 1.0,\\
\varphi_{\phi\nu 6} &= a_{1121} + a_{1211} + a_{1321} + a_{1421} + a_{1521} = 1.0,\\
\varphi_{\phi\nu 7} &= a_{1121} + a_{1221} + a_{1311} + a_{1421} + a_{1521} = 1.0.
\end{aligned} \quad (14)$$

At this point, the spreadsheet organizes and solves the equations. The expected values are calculated based upon the algorithmic steps from the following respondent-provided fuzzy probabilities.

2.6. Illustrative Spreadsheet Example. While the algorithmic process is not overly complex, the transitioning of the algorithm into a realistic decision support system (DSS) capable of handling an extensive set of data has many challenges. Due to spreadsheet ease of use and availability to businesses, including small to mid-size enterprises, the focus of the research was to provide a fully functioning spreadsheet-based DSS. The illustration provided shows the resulting complexity of interpreting the algorithm with a standard spreadsheet approach.

An organized layout of the intermediate results of algorithm Steps 1 and 2 was created, as shown in Table 2, where 1*s*–5*s* represent the qualitative variables defined as Poor, Below Average, Average, Above Average, and Excellent, respectively. Question 20 referred to the on-time delivery performance of the suppliers denoted as "a" through "f" for Arrow, Avnet, Future, Insight, Pioneer, and TTI, respectively.

The probability calculation is a simple counting of the number of responses of, for example, "2," divided by a count of the number of responses in the appropriate partition. The function used is as follows:

```
=COUNTIF([cell range of responses],"=2")/
COUNTIF([cell range of responses],">0")
```

The belief calculation relied upon the assumptions stated above and given in Table 1. Accordingly, the belief value for each score (1–5) is calculated using the following formula:

```
=SUMIF([cell range of scores],"=2",
[cell range of beliefs])/COUNTIF([cell
range of scores],"=2")
```

Rounding leads to the two-significant-digit 5-tuples shown in Table 3.

Organizing the 5-tuples into sums of fuzzy probabilities is the next step, according to Step 3 of the algorithm. This was accomplished by creating a row/column spot for every possible combination of every element of every question, as shown in Table 3.

The numbers in Table 3 represent the possible combinations of beliefs summed for Question 20 for suppliers ν_1 through ν_6; 16 for each data partition per supplier. The shaded cells represent combinations that have fuzzy probabilities that sum 1.0 or that sum very close to 1.0, within the established threshold. Note that the top and bottom combinations always sum exactly to one since all possibilities are represented by the five scores (1–5) in any given partition.

With the 5-tuples identified that have fuzzy probabilities that sum 1, the resulting score is calculated for each of those 5-tuples by multiplying each fuzzy probability by its corresponding score. As an example, the overall score of the first 5-tuple is found using the following equation (rounded to four significant digits):

$$0 \times 1 + 0.5333 \times 2 + 0.2267 \times 3 + 0.6133 \times 4 + 0.1067 \times 5 = 3.77. \quad (15)$$

The scores for all these first 5-tuples are shown in Table 4.

Step 5 of the algorithm requires that the minimum belief for each usable 5-tuple is identified, discarding any case where the belief was zero.

Finally, each value in Table 4 is multiplied by each value in Table 5 and then summed down each column to form an expected score for each supplier on each question.

Table 2: Scores for all suppliers on all attributes of the delivery criterion.

Supplier	On-Time delivery	Availability of inventory	Shipping accuracy	Return authorization	Overall average
v_1: Arrow	3.80	**3.77**	3.93	**3.80**	**3.82**
v_2: Avnet	3.69	3.63	3.90	3.68	3.72
v_3: Future	3.71	3.60	3.81	3.30	3.61
v_4: Insight	3.65	3.33	3.83	3.66	3.62
v_5: Pioneer	3.57	3.27	3.70	3.61	3.54
v_6: TTI	**3.89**	3.71	**4.06**	3.68	**3.83**

Table 3: Steps 1 and 2 of the algorithm applied to survey question 20.

				20a	b	c	d	e	f
$a11$	First partitioned data	Probability of score	$1s$	0.0000	0.0133	0.0405	0.0278	0.0676	0.0000
$a21$			$2s$	0.0533	0.0533	0.0270	0.0556	0.0676	0.0448
$a31$			$3s$	0.2267	0.2533	0.2568	0.2917	0.2703	0.2388
$a41$			$4s$	0.6133	0.6133	0.5811	0.4722	0.4730	0.4627
$a51$			$5s$	0.1067	0.0667	0.0946	0.1528	0.1216	0.2537
$\alpha12$		Weight of score (belief)	$1s$	0.0000	0.2000	0.6300	0.4500	0.2600	0.0000
$\alpha22$			$2s$	0.1100	0.4000	0.6000	0.4300	0.0900	0.2400
$\alpha32$			$3s$	0.3900	0.3800	0.3400	0.1900	0.3900	0.3300
$\alpha42$			$4s$	0.3000	0.2800	0.2400	0.3700	0.2700	0.2500
$\alpha52$			$5s$	0.2200	0.1600	0.3300	0.3000	0.3900	0.2600
$a11$	Second partitioned data	Probability of score	$1s$	0.0400	0.0267	0.0411	0.0143	0.0811	0.0282
$a21$			$2s$	0.0800	0.0800	0.0685	0.1286	0.0270	0.0563
$a31$			$3s$	0.2133	0.2667	0.1644	0.3000	0.2703	0.2676
$a41$			$4s$	0.3600	0.4000	0.5205	0.3143	0.4324	0.3239
$a51$			$5s$	0.3067	0.2267	0.2055	0.2429	0.1892	0.3239
$\alpha12$		Weight of score (belief)	$1s$	0.2500	0.3600	0.7300	0.2000	0.3100	0.7000
$\alpha22$			$2s$	0.3400	0.3400	0.3300	0.4400	0.1200	0.6500
$\alpha32$			$3s$	0.3800	0.4000	0.2100	0.2800	0.3100	0.2600
$\alpha42$			$4s$	0.3500	0.3400	0.3500	0.3900	0.3700	0.2900
$\alpha52$			$5s$	0.1900	0.1000	0.1900	0.1900	0.1600	0.2500

Using the spreadsheet results, the first delivery attribute assessed (on-time delivery) is calculated from the fuzzy probability distributions for the remaining suppliers ($Q_{\phi k v}$) as

$$
\begin{aligned}
& \qquad\ \ \underline{j=1} \quad \underline{j=2} \\
Q_{113} &= 0.63/0.04 + 0.73/0.04 \quad \text{for } s_{113} = \text{Poor (1)} \\
Q_{123} &= 0.60/0.03 + 0.33/0.07 \quad \text{for } s_{123} = \text{Below Average (2)} \\
Q_{133} &= 0.34/0.26 + 0.21/0.16 \quad \text{for } s_{133} = \text{Average (3)} \\
Q_{143} &= 0.24/0.58 + 0.35/0.52 \quad \text{for } s_{143} = \text{Above Average (4)} \\
Q_{153} &= 0.33/0.09 + 0.19/0.21 \quad \text{for } s_{153} = \text{Excellent (5)} \\
Q_{114} &= 0.45/0.03 + 0.20/0.01 \quad \text{for } s_{114} = \text{Poor (1)} \\
Q_{124} &= 0.43/0.06 + 0.44/0.13 \quad \text{for } s_{124} = \text{Below Average (2)} \\
Q_{134} &= 0.19/0.29 + 0.28/0.27 \quad \text{for } s_{134} = \text{Average (3)} \\
Q_{144} &= 0.37/0.47 + 0.35/0.52 \quad \text{for } s_{144} = \text{Above Average (4)} \\
Q_{154} &= 0.30/0.15 + 0.19/0.24 \quad \text{for } s_{154} = \text{Excellent (5)} \\
Q_{115} &= 0.26/0.07 + 0.31/0.08 \quad \text{for } s_{115} = \text{Poor (1)} \\
Q_{125} &= 0.09/0.07 + 0.12/0.03 \quad \text{for } s_{125} = \text{Below Average (2)} \\
Q_{135} &= 0.39/0.27 + 0.31/0.27 \quad \text{for } s_{135} = \text{Average (3)} \\
Q_{145} &= 0.27/0.47 + 0.37/0.43 \quad \text{for } s_{145} = \text{Above Average (4)} \\
Q_{155} &= 0.39/0.12 + 0.16/0.19 \quad \text{for } s_{155} = \text{Excellent (5)} \\
Q_{116} &= 0.00/0.00 + 0.70/0.03 \quad \text{for } s_{116} = \text{Poor (1)} \\
Q_{126} &= 0.24/0.04 + 0.65/0.06 \quad \text{for } s_{126} = \text{Below Average (2)} \\
Q_{136} &= 0.33/0.24 + 0.26/0.27 \quad \text{for } s_{136} = \text{Average (3)} \\
Q_{146} &= 0.25/0.46 + 0.29/0.32 \quad \text{for } s_{146} = \text{Above Average (4)} \\
Q_{156} &= 0.26/0.25 + 0.25/0.32 \quad \text{for } s_{156} = \text{Excellent (5)}.
\end{aligned} \tag{16}
$$

Table 4: 5-tuple sums of fuzzy probability combinations for Question 20.

		20a	b	c	d	e	f
	a11a21a31a41a51	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
	a11a21a31a41a52	1.20	1.16	1.11	1.09	1.07	1.07
	a11a21a31a42a51	0.75	0.79	0.94	0.84	0.96	0.86
	a11a21a31a42a52	0.95	0.95	1.05	0.93	1.03	0.93
	a11a21a32a41a51	0.99	1.01	0.91	**1.01**	**1.00**	1.03
	a11a21a32a41a52	1.19	1.17	1.02	1.10	1.07	1.10
	a11a21a32a42a51	0.73	0.80	0.85	0.85	0.96	0.89
	a11a21a32a42a52	0.97	0.97	1.09	**1.01**	0.99	0.94
	a11a22a31a41a51	1.03	1.03	1.04	1.07	0.96	1.01
	a11a22a31a41a52	1.23	1.19	1.15	1.16	1.03	1.08
	a11a22a31a42a51	0.77	0.81	0.98	0.92	0.92	0.87
	a11a22a31a42a52	0.97	0.97	1.09	**1.01**	0.99	0.94
	a11a22a32a41a51	1.01	1.04	0.95	1.08	0.96	1.04
	a11a22a32a41a52	1.21	1.20	1.06	1.17	1.03	1.11
	a11a22a32a42a51	0.76	0.83	0.89	0.92	0.92	0.90
Raw sums of n-tuples	a11a22a32a42a52	0.96	0.99	**1.00**	1.01	0.99	0.97
	a12a21a31a41a51	1.04	1.01	**1.00**	0.99	1.01	1.03
	a12a21a31a41a52	1.24	1.17	1.11	1.08	1.08	1.10
	a12a21a31a42a51	0.79	0.80	0.94	0.83	0.97	0.89
	a12a21a31a42a52	0.99	0.96	1.05	0.92	1.04	0.96
	a12a21a32a41a51	1.03	1.03	0.91	**0.99**	1.01	1.06
	a12a21a32a41a52	1.23	1.19	1.02	1.08	1.08	1.13
	a12a21a32a42a51	0.77	0.81	0.85	0.84	0.97	0.92
	a12a21a32a42a52	1.01	0.99	1.09	**0.99**	**1.00**	0.97
	a12a22a31a41a51	1.07	1.04	1.04	1.06	0.97	1.04
	a12a22a31a41a52	1.27	1.20	1.15	1.15	1.04	1.11
	a12a22a31a42a51	0.81	0.83	0.98	0.90	0.93	0.90
	a12a22a31a42a52	1.01	0.99	1.09	**0.99**	**1.00**	0.97
	a12a22a32a41a51	1.05	1.05	0.95	1.07	0.97	1.07
	a12a22a32a41a52	1.25	1.21	1.06	1.16	1.04	1.14
	a12a22a32a42a51	0.80	0.84	0.89	0.91	0.93	0.93
	a12a22a32a42a52	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**

3. Results and Discussion

3.1. Results. The fuzzy probabilities from the respondents for the six suppliers are found by Step 3 of the algorithm and result in four fuzzy probability (summation to 1.0) combinations for v_3; 8 combinations for v_4; 5 combinations for v_5; and 2 combinations for v_6. The results for the six suppliers in this spreadsheet example are as shown in Table 6.

The question then becomes whether this fuzzy set-based model is useful in identifying supplier performance and indeed is a better supplier selection approach. One measure for this is the mean absolute percent deviation of these values from a simple averaging of the 150 example responses, which amounts to 0.35 percent in this example. This indicates that the algorithm as illustrated with survey data is statistically robust.

The expected values derived from the fuzzy probability scores and the simple average scores are consistent when rounded to two significant digits. Thus, the data selection and partitioning process from Step 0 of the algorithm did not perturb the numerical results unduly.

The fuzzy probability approach shows v_6 as the best for on-time delivery and accuracy of the shipment with v_1 second best, but v_1 is best for inventory availability and authorization of returns with v_6 second best. With an overall average score for these suppliers differing by only 0.01, a decision maker might be inclined to rate both Arrow and TTI as preferred suppliers.

As another basic comparison, a simple averaging of all respondent input is considered as shown in Table 7.

In contrast to results from both the fuzzy set-based model and the simple average of the partitioned dataset, when using a simple average method across all data, the rankings are fairly inconsistent except for v_6 which scored the highest on every delivery attribute, thus ranking as #1. The rankings for the other suppliers are not obvious.

Table 5: Fuzzy probabilities for Question 20 multiplied by associated scores (1–5) and summed for each usable 5-tuple.

		20a	b	c	d	e	f
Fuzzy probabilities times values 1 through 5	$a11a21a31a41a51$	3.7733	3.6667	3.6622	3.6667	3.5135	3.9254
	$a11a21a31a41a52$						
	$a11a21a31a42a51$						
	$a11a21a31a42a52$						
	$a11a21a32a41a51$				3.6917	3.5135	
	$a11a21a32a41a52$						
	$a11a21a32a42a51$						
	$a11a21a32a42a52$				3.6313		
	$a11a22a31a41a51$						
	$a11a22a31a41a52$						
	$a11a22a31a42a51$						
	$a11a22a31a42a52$				3.6313		
	$a11a22a32a41a51$						
	$a11a22a32a41a52$						
	$a11a22a32a42a51$						
	$a11a22a32a42a52$			3.7803			
	$a12a21a31a41a51$			3.6627			
	$a12a21a31a41a52$						
	$a12a21a31a42a51$						
	$a12a21a31a42a52$						
	$a12a21a32a41a51$				3.6782		
	$a12a21a32a41a52$						
	$a12a21a32a42a51$						
	$a12a21a32a42a52$				3.6179	3.6216	
	$a12a22a31a41a51$						
	$a12a22a31a41a52$						
	$a12a22a31a42a51$						
	$a12a22a31a42a52$				3.6179	3.6216	
	$a12a22a32a41a51$						
	$a12a22a32a41a52$						
	$a12a22a32a42a51$						
	$a12a22a32a42a52$	3.8133	3.7200	3.7808	3.6429	3.6216	3.8592

For example, v_3 was second best for on-time delivery and inventory availability, but weakest of all suppliers with respect to return authorization. The v_2 criterion was second for accuracy of shipping and return authorization, but otherwise inconsistent. The v_5 criterion was the least competitive for on-time delivery, availability of inventory, and shipping accuracy but was competitive for return authorization. The v_1 criterion ranked fourth in on-time performance and managed only a third place on the other attributes.

Therefore, using all available respondent data for every supplier does not yield as clear a set of results as does the fuzzy algorithm. The fuzzy set-based model resolves these types of ambiguity derived from customer perceptions and close scores within [1, 5].

3.2. Discussion. The fuzzy set-based approach has an advantage not observed in simple averages or a typical scoring approach. Each customer describes the importance to them of each of the various supplier attributes, which can be compared against the cumulative fuzzy probabilities of each supplier to perform to expectations, often revealing an apparent contradiction when compared with each customer's stated preferred supplier. For example, customer "B2573" states a preference for supplier v_2 when supplier v_6 should better meet this customer's needs along the lines of delivery performance. This customer ranks supplier v_6 5th out of 6, effectively excluding supplier v_6 from its primary business. As observed in the literature, fuzzy set-based models can handle this type of decision maker uncertainty and the ambiguity of data provided. The algorithm presented herein with illustrative application adds further support to this conclusion.

The algorithm incorporates beliefs into the determination of the scores, so by the approach detailed, the weight is inherent to the score and avoids the potential bias of weighting

Table 6: Minimum beliefs for Question 20 for each usable 5-tuple.

		20a	b	c	d	e	f
Fuzzy probabilities times values 1 through 5	$a11a21a31a41a51$	0.1100	0.1600	0.2400	0.1900	0.0900	0.2400
	$a11a21a31a41a52$						
	$a11a21a31a42a51$						
	$a11a21a31a42a52$						
	$a11a21a32a41a51$				0.2800	0.0900	
	$a11a21a32a41a52$						
	$a11a21a32a42a51$						
	$a11a21a32a42a52$				0.1900		
	$a11a22a31a41a51$						
	$a11a22a31a41a52$						
	$a11a22a31a42a51$						
	$a11a22a31a42a52$				0.1900		
	$a11a22a32a41a51$						
	$a11a22a32a41a52$						
	$a11a22a32a42a51$						
	$a11a22a32a42a52$			0.1900			
	$a12a21a31a41a51$			0.2400			
	$a12a21a31a41a52$						
	$a12a21a31a42a51$						
	$a12a21a31a42a52$						
	$a12a21a32a41a51$				0.2000		
	$a12a21a32a41a52$						
	$a12a21a32a42a51$						
	$a12a21a32a42a52$				0.1900	0.1200	
	$a12a22a31a41a51$						
	$a12a22a31a41a52$						
	$a12a22a31a42a51$						
	$a12a22a31a42a52$				0.1900	0.1200	
	$a12a22a32a41a51$						
	$a12a22a32a41a52$						
	$a12a22a32a42a51$						
	$a12a22a32a42a52$	0.1900	0.1000	0.1900	0.1900	0.1200	0.2500

and determining a score after seeing the average score for each supplier. It is this type of ambiguity and potential bias in scoring models that the fuzzy probability model eliminates.

The algorithm utilizes a partitioning of the dataset prior to execution of the fuzzy set machinations. The partitioning for ease of presentation and spreadsheet application was confined to two equal sets. This study was restricted to five scores and two datasets partly to limit complexity to 32 unique combinations per supplier per attribute, a problem that could be hard-coded into a spreadsheet fairly easily. The algorithm is not restrictive and can handle any number of criteria, attributes, decision maker input, and partitioning without loss of generality.

The spreadsheet itself is a unique contribution. While the algorithm has been accepted in a more generalized form, no further modeling was conducted on large databases due to limitations caused by the size of any realistic supplier selection decision making process. The spreadsheet provided easily interpreted input into the algorithmic format. Also, the use of the survey instrument allowed representative input from an extensive number (150 viable respondents) with whom a one-to-one assessment and measure of beliefs would have been difficult to manage, especially as the survey respondents were based internationally.

It is worthwhile to consider the fuzzy application as applied to the largest available dataset. Whereas the dataset in use was pared down to 150 total respondents (75 per partition), the original dataset includes some 406 individual responses, although they vary greatly in the percentage of the survey questions completed. Table 8 shows the results of the fuzzy method applied to this approximately 2.7 times larger dataset.

Table 7: Averaged operational parameter data for the six suppliers.

Supplier	n	On-Time delivery	n	Inventory availability	n	Shipping accuracy	n	RMA	Average
ν_1: Arrow	361	3.6953	362	3.6353	359	3.8719	334	**3.6377**	3.71
ν_2: Avnet	**375**	3.6773	**377**	3.6220	**374**	3.9251	**356**	**3.6404**	3.72
ν_3: Future	347	3.7666	347	3.6455	344	3.8517	310	3.3129	3.64
ν_4: Insight	282	3.6631	287	3.3171	284	3.8627	245	3.5918	3.61
ν_5: Pioneer	298	3.5872	300	3.0333	295	3.7492	265	3.5509	3.48
ν_6: TTI	281	**3.9288**	291	**3.7251**	283	**4.0212**	256	**3.6523**	**3.83**

Table 8: Fuzzy scores for all suppliers on all attributes of the delivery criterion, dataset of all 406 respondents.

Supplier	On-Time delivery	Availability of inventory	Shipping accuracy	Return authorization	Overall average
ν_1: Arrow	3.70	3.64	3.88	**3.63**	3.71
ν_2: Avnet	3.68	3.59	3.91	3.55	3.68
ν_3: Future	3.77	3.66	3.86	3.31	3.65
ν_4: Insight	3.66	3.33	3.86	3.57	3.61
ν_5: Pioneer	3.59	3.30	3.75	3.53	3.54
ν_6: TTI	**3.93**	**3.73**	**4.02**	**3.65**	**3.83**

Of note is that this result from the data superset conforms even more closely to the average operational parameter data from Table 7, with resulting rankings substantially similar, except for the RMA category. However, when averaged into overall supplier ranking, a different set of preferences emerges from the subtle differences between simple average scores and fuzzy criteria scores. The same supplier, TTI, is chosen first, but this time Arrow outranks Avnet for second place, supporting the decision made when using the fuzzy algorithm on the smaller data subset and reinforcing the notion that a fuzzy approach yields different results that consistently capture and use belief to handle ambiguity in the data. Removing the incomplete responses to create the aforementioned data subset (as was used to create Table 7) merely served to enhance this effect, to the extent that Arrow challenges TTI for first place and becomes a plausible option as a supplier of first choice.

A major benefit of using survey data for this decision system is the ease of data accumulation and modeling. The application shows that a typical industry survey can be adapted, which allows a company to employ the fuzzy set-based model whenever canned data are available. This aspect alone could allow companies to engage in supplier selection evaluations and potentially improve the number that evaluate suppliers from the approximately 50 percent that have been measured to do so [1].

The closeness of the survey data when statistically analyzed is not a surprise given the use of the survey and adjustments for belief functions. Of support to the fuzzy set-based approach is that the results were similar, while an opportunity for a company following this approach would be to adjust the design of the survey instrument to more directly obtain respondent beliefs, rather than rely on the assumption that total spending corresponds directly to belief.

Finally, some consideration of complexity is in order. The fuzzy algorithm presented is not, itself, a significant source of computational complexity and should operate with suitably low complexity on any large dataset that can nevertheless fit in a modern spreadsheet file. However, as the dataset increases in size, so does the size of the data partitions, until there begin to be algorithm advantages to splitting the dataset into more than two partitions. However, every increase in the number of data partitions causes a geometric increase in the complexity, not of the algorithm, but of the organization of the dataset. In this study, a brute force method of dataset organization was applied that manually broke the dataset into two partitions. Because the data in the questions consisted of responses on a Likert scale, this study resulted in an organizational complexity best illustrated by Table 4. Every possible combination of responses within a question/partition combination must be calculated and considered, so any increase in the number of partitions (or in the size of the question scale; from a 5-point Likert scale to a 7-point Likert scale, for example) adds a huge amount of additional set-up complexity to the process. This added complexity amounts to a "price" paid for additional fuzzy performance, because a dataset of arbitrarily large size could still be addressed using only the two-partition approach shown in this paper's example. However, additional performance from the fuzzy algorithm can be achieved in larger datasets by paying the complexity cost of increased partitions.

At some point of increased complexity (partition and/or question scale), the organization of the dataset must be handled programmatically, because it becomes untenable to manually organize the evaluation of the question/partition n-tuples as was done in this study. It will be a topic of further study to develop methods of scaling up this complexity smoothly to match the opportunities presented by larger datasets. As it is, arbitrarily large datasets can be handled effectively using the presented method.

4. Conclusions

The fuzzy set-based method for ranking suppliers is shown to be more robust than other methods. Traditional scoring methods fail to adequately consider the ambiguity in data belief among opinion holders. Fuzzy set theory is equipped to handle these complexities. The spreadsheet application is a considerable benefit of this model, especially for small to mid-size companies (SMOs). SMOs lack the resources for expensive data analysis systems or consultants. SMOs can readily gather or repurpose large sets of data using a survey similar to the one used in this application and can apply this algorithm to obtain the benefit of using fuzzy logic to evaluate their suppliers. Furthermore, this fuzzy set-based model utilizes basic computational tools of a spreadsheet that any organization can apply with little or no additional cost.

The fact that the algorithmic process can be adapted to use such a database is a benefit to any company. All combinations must be considered where fuzzy probabilities for the outcomes (Excellent, Above Average, etc.) sum 100%. The use of standard statistical probability by dividing the quantity seen of each score by the total number of responses always provides one "default" case for each set of partitioned data. Then, the spreadsheet calculates all possible combinations of the fuzzy probabilities, per the algorithm. A limitation to be addressed is redesign of the survey instrument to allow for directly obtaining belief functions for supplier membership in qualitatively defined variables such as excellent, good, fair, and so forth and allowing membership not to be crisply defined as 1.0.

As was discussed in Section 2.6, the fuzzy algorithm, in theory, is not overly complex, but in practice the resulting decision support system can become complex in proportion to the dataset size. There are two primary factors in this. The first is that, for the example spreadsheet illustration, only two partitions were created (per Step 0 of the algorithm). For 150 sets of respondent data, this puts a relatively small amount of data in each partition but more importantly creates a relatively small number of possible fuzzy combinations of scores and partitions. Even when all 406 responses are evenly partitioned, the result is manageable datasets. But, increasing to significantly large datasets would make using only two partitions suboptimal. Increasing the number of data partitions addresses this but increases the combinatorial testing for fuzzy probabilities summing 1.0 (Steps 3 and 4 of the algorithm). When dealing with a survey dataset where at least one combination is forced into a sum of 1.0 (the default case described above), this becomes less of a consideration than when fuzzy beliefs are provided resulting in potential inconsistencies of judgments by the decision makers. Still, the spreadsheet model can address numerous partitions without loss of functionality. The second problem is that (at present) the combinations are not programmatically generated but are instead the result of visual inspection (or manually created automation for each different dataset) for all fuzzy probabilities. This brute forcing of the data is being addressed through development of a network-type of interface by which the spreadsheet will be populated directly from the dataset into combinations, regardless of the number of partitioned sets. Thus, while larger datasets increase the complexity of the spreadsheet-based DSS, the algorithm itself is fully able to handle datasets of any size with negligible increases in computational power.

Future research on this topic will seek to integrate other criteria and available data—not just delivery-specific data. As in this study, by using other survey responses involving criteria beyond price, delivery, and quality, it becomes possible to make specific recommendations by evaluating suppliers using criteria not usually considered. Also, an aforementioned improvement in the ease by which users may scale up complexity for ever larger datasets is an area for further research and development.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] P. M. Simpson, J. A. Siguaw, and S. C. White, "Measuring the performance of suppliers: an analysis of evaluation processes," *Journal of Supply Chain Management*, vol. 38, no. 1, pp. 29–41, 2002.

[2] L. Vijayvagy, "Decision framework for supplier selection through multi criteria evaluation models in supply chain," *International Journal of Management & Innovation*, vol. 4, no. 2, pp. 16–28, 2012.

[3] R. M. Monczka, K. J. Petersen, R. B. Handfield, and G. L. Ragatz, "Success factors in strategic supplier alliances: the buying company perspective," *Decision Sciences*, vol. 29, no. 3, pp. 553–573, 1998.

[4] G. Wang, S. H. Huang, and J. P. Dismukes, "Product-driven supply chain selection using integrated multi-criteria decision-making methodology," *International Journal of Production Economics*, vol. 91, no. 1, pp. 1–15, 2004.

[5] K. S. Bhutta and F. Huq, "Supplier selection problem: a comparison of the total cost of ownership and analytic hierarchy process approaches," *Supply Chain Management*, vol. 7, no. 3, pp. 126–135, 2002.

[6] R. Handfield, S. V. Walton, R. Sroufe, and S. A. Melnyk, "Applying environmental criteria to supplier assessment: a study in the application of the Analytical Hierarchy Process," *European Journal of Operational Research*, vol. 141, no. 1, pp. 70–87, 2002.

[7] H. J. Einhorn and R. M. Hogarth, "Ambiguity and uncertainty in probabilistic inference," *Psychological Review*, vol. 92, no. 4, pp. 433–461, 1985.

[8] D. Kahneman and A. Tversky, "Prospect theory: an analysis of decision under risk," *Econometrica*, vol. 47, no. 2, pp. 263–292, 1979.

[9] G. Shafer, *A Mathematical Theory of Evidence*, vol. 1, Princeton University Press, Princeton, NJ, USA, 1976.

[10] J. Yen, "Generalizing the Dempster-Shafer theory to fuzzy sets," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 20, no. 3, pp. 559–570, 1990.

[11] L. A. Zadeh, "A simple view of the Dempster-Shafer theory of evidence and its implication for the rule of combination," *AI Magazine*, vol. 7, no. 2, pp. 85–90, 1986.

[12] A. D. Korvin and M. F. Shipley, "Dempster-Shafer-based approach to compromise decision making with multiattributes applied to product selection," *IEEE Transactions on Engineering Management*, vol. 40, no. 1, pp. 60–67, 1993.

[13] R. E. Bellman and L. A. Zadeh, "Decision-making in a fuzzy environment," *Management Science*, vol. 17, no. 4, pp. B141–B164, 1970.

[14] V. Jain, S. Wadhwa, and S. G. Deshmukh, "Supplier selection using fuzzy association rules mining approach," *International Journal of Production Research*, vol. 45, no. 6, pp. 1323–1353, 2007.

[15] R. R. Yager, "Toward a general theory of reasoning with uncertainty. I: nonspecificity and fuzziness," *International Journal of Intelligent Systems*, vol. 1, no. 1, pp. 45–67, 1986.

[16] G. J. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic*, Prentice Hall, Upper Saddle River, NJ, USA, 1995.

[17] R. R. Yager and V. Kreinovich, "Entropy conserving probability transforms and the entailment principle," *Fuzzy Sets and Systems*, vol. 158, no. 12, pp. 1397–1405, 2007.

[18] A. Zebda, "The investigation of cost variances: a fuzzy set theory approach," *Decision Sciences*, vol. 15, no. 3, pp. 359–388, 1984.

[19] A. Hadi-Vencheh and M. Niazi-Motlagh, "An improved voting analytic hierarchy process-data envelopment analysis methodology for suppliers selection," *International Journal of Computer Integrated Manufacturing*, vol. 24, no. 3, pp. 189–197, 2011.

[20] A. Guneri and A. Kuzu, "Supplier selection by using a fuzzy approach in just-in-time: a case study," *International Journal of Computer Integrated Manufacturing*, vol. 22, no. 8, pp. 774–783, 2009.

[21] F.-H. F. Liu and H. L. Hai, "The voting analytic hierarchy process method for selecting supplier," *International Journal of Production Economics*, vol. 97, no. 3, pp. 308–317, 2005.

[22] D. Park and H. A. Krishnan, "Supplier selection practices among small firms in the United States: testing three models," *Journal of Small Business Management*, vol. 39, no. 3, pp. 259–271, 2001.

[23] N. Hirakubo and M. Kublin, "The relative importance of supplier selection criteria: the case of electronic components procurement in Japan," *Journal of Supply Chain Management*, vol. 34, no. 2, pp. 19–24, 1998.

[24] C. A. Weber, J. R. Current, and W. C. Benton, "Vendor selection criteria and methods," *European Journal of Operational Research*, vol. 50, no. 1, pp. 2–18, 1991.

[25] L. M. Ellram, "The supplier selection decision in strategic," *Journal of Purchasing and Materials Management*, vol. 26, no. 4, pp. 8–14, 1990.

[26] R. Verma and M. E. Pullman, "An analysis of the supplier selection process," *Omega*, vol. 26, no. 6, pp. 739–750, 1998.

[27] Z. Degraeve and F. Roodhooft, "A smarter way to buy," *Harvard Business Review*, vol. 79, no. 6, pp. 22–145, 2001.

[28] Z. Degraeve and F. Roodhooft, "Effectively selecting suppliers using total cost of ownership," *Journal of Supply Chain Management*, vol. 35, no. 1, pp. 5–10, 1999.

[29] B. Vahdani and M. Zandieh, "Selecting suppliers using a new fuzzy multiple criteria decision model: the fuzzy balancing and ranking method," *International Journal of Production Research*, vol. 48, no. 18, pp. 5307–5326, 2010.

[30] B. Liu, "Fuzzy criterion models for inventory systems with partial backorders," *Annals of Operations Research*, vol. 87, pp. 117–126, 1999.

[31] K. Das, T. K. Roy, and M. Maiti, "Buyer-seller fuzzy inventory model for a deteriorating item with discount," *International Journal of Systems Science: Principles and Applications of Systems and Integration*, vol. 35, no. 8, pp. 457–466, 2004.

[32] J. Usenik and M. Bogataj, "A fuzzy set approach for a location-inventory model," *Transportation Planning and Technology*, vol. 28, no. 6, pp. 447–464, 2005.

[33] J. C.-H. Pan and M.-F. Yang, "Integrated inventory models with fuzzy annual demand and fuzzy production rate in a supply chain," *International Journal of Production Research*, vol. 46, no. 3, pp. 753–770, 2008.

[34] Y. Kara, H. Gökçen, and Y. Atasagun, "Balancing parallel assembly lines with precise and fuzzy goals," *International Journal of Production Research*, vol. 48, no. 6, pp. 1685–1703, 2010.

[35] T.-F. Liang, "Integrating production-transportation planning decision with fuzzy multiple goals in supply chains," *International Journal of Production Research*, vol. 46, no. 6, pp. 1477–1494, 2008.

[36] W.-H. Tsai and S.-J. Hung, "A fuzzy goal programming approach for green supply chain optimisation under activity-based costing and performance evaluation with a value-chain structure," *International Journal of Production Research*, vol. 47, no. 18, pp. 4991–5017, 2009.

[37] M.-H. Shu and H.-C. Wu, "Measuring the manufacturing process yield based on fuzzy data," *International Journal of Production Research*, vol. 48, no. 6, pp. 1627–1638, 2010.

[38] F. T. S. Chan, N. Kumar, M. K. Tiwari, H. C. W. Lau, and K. L. Choy, "Global supplier selection: a fuzzy-AHP approach," *International Journal of Production Research*, vol. 46, no. 14, pp. 3825–3857, 2008.

[39] M. Bevilacqua and A. Petroni, "From traditional purchasing to supplier management: a fuzzy logic-based approach to supplier selection," *International Journal of Logistics*, vol. 5, no. 3, pp. 235–255, 2002.

[40] M. Y. Bayrak, N. Çelebi, and H. Takin, "A fuzzy approach method for supplier selection," *Production Planning and Control*, vol. 18, no. 1, pp. 54–63, 2007.

[41] M. Sevkli, "An application of the fuzzy ELECTRE method for supplier selection," *International Journal of Production Research*, vol. 48, no. 12, pp. 3393–3405, 2010.

[42] L. A. Zadeh, "Generalized theory of uncertainty (GTU)—principal concepts and ideas," *Computational Statistics & Data Analysis*, vol. 51, no. 1, pp. 15–46, 2006.

[43] D. Dubois and H. Prade, "Systems of linear fuzzy constraints," *Fuzzy Sets and Systems*, vol. 3, no. 1, pp. 37–48, 1980.

[44] A. N. S. Freeling, "Fuzzy sets and decision analysis," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 10, no. 7, pp. 341–354, 1980.

[45] R. R. Yager, "On some new classes of implication operators and their role in approximate reasoning," *Information Sciences*, vol. 167, no. 1–4, pp. 193–216, 2004.

[46] A. Kaufmann and M. Gupta, *An Introduction to Fuzzy Sets Arithmetic*, Nosfrand Reinhold, New York, NY, USA, 1985.

[47] G. J. Klir and T. A. Folger, *Fuzzy Sets, Uncertainty, and Information*, 1988.

[48] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, pp. 338–353, 1965.

[49] L. A. Zadeh, "Fuzzy logic and approximate reasoning," *Synthese*, vol. 30, no. 3-4, pp. 407–428, 1975.

[50] M. M. Gupta, R. K. Ragade, and R. R. Yager, *Advances in Fuzzy Set Theory and Applications*, North-Holland, New York, NY, USA, 1979.

[51] M. F. Shipley and G. L. Stading, "Supplier selection decisions: a fuzzy logic model based on quality aspects of delivery," in *Advances in Computational Intelligence*, vol. 300 of *Communications in Computer and Information Science*, pp. 1–9, Springer, Berlin, Germany, 2012.

Hindawi Publishing Corporation
Advances in Fuzzy Systems
Volume 2014, Article ID 197876, 7 pages
http://dx.doi.org/10.1155/2014/197876

Research Article

Interest Measures for Fuzzy Association Rules Based on Expectations of Independence

Michal Burda

Institute for Research and Applications of Fuzzy Modeling, Centre of Excellence IT4Innovations, University of Ostrava, 30. dubna 22, 701 03 Ostrava, Czech Republic

Correspondence should be addressed to Michal Burda; michal.burda@osu.cz

Received 25 August 2014; Accepted 9 September 2014; Published 7 October 2014

Academic Editor: Salvatore Sessa

Lift, leverage, and conviction are three of the best commonly known interest measures for crisp association rules. All of them are based on a comparison of observed support and the support that is expected if the antecedent and consequent part of the rule were stochastically independent. The aim of this paper is to provide a correct definition of lift, leverage, and conviction measures for fuzzy association rules and to study some of their interesting mathematical properties.

1. Introduction

Searching for association rules is a broadly discussed, developed, and accepted data mining technique [1, 2]. An association rule is an expression $X \rightharpoonup Y$, where antecedent X and consequent Y are conditions, the former usually in the form of elementary conjunction and the latter being usually atomic. Such rules are usually interpreted as the following implicational statement: "if X is satisfied then Y is true very often too."

Naturally, analysts are interested only in such rules that are somehow interesting, unusual, or exceptional. To assess rule interestingness objectively, there have been developed many measures of rule interest or intensity. Among the most essential, *support* and *confidence* are traditionally considered. An objective of association rules mining is to find rules with *support* and *confidence* above some user-defined thresholds.

Searching for association rules fits particularly well on binary or categorical data and many have been written on that topic [1–4]. For association analysis on numeric data, a prior discretization is proposed, for example, by Srikant and Agrawal [5]. Another alternative is to take an advantage of fuzzy logic [6].

The use of fuzzy logic in connection with association rules has been motivated by many authors (see e.g., [7] for recent overview). Fuzzy association rules are appealing also because of the use of vague linguistic terms such as "small" and "very big" [8–11].

In this paper, we focus on three measures of rule intensity that are all based on comparison between the observed support and the support that is expected under the assumption of independence of the rule's antecedent and consequent. These measures are *lift*, *leverage*, and *conviction*. All of them were initially developed for nonfuzzy (i.e., "crisp") association rules.

Lift was firstly described in [12] under its original name "interest." It was well studied for association rules on binary data in [13, 14]. Lift is defined as a ratio of observed support $X \wedge Y$ to the support that is expected under the assumption of independence of X and Y.

On the other hand, *leverage* [15] measures the difference of observed and expected supports. (Hence, it is very similar to lift.)

Conviction [12] is defined as a ratio of expected and observed support of $X \wedge \neg Y$. Although it is very similar to lift, its properties are more similar to confidence. Lallich et al. [16] provide a nice overview of many other crisp rule measures.

As quite many have been written about these measures for crisp association rules, not so much has been done with respect to fuzzy rules. Unfortunately, simplicity of definitions for crisp rules sometimes led to oversimplified definitions for fuzzy rules. Some authors believe the generalization of those crisp measures for fuzzy data is as trivial as substituting crisp terms with analogous fuzzy terminology inside of crisp-case definitions; see, for example, [17, 18]. Unfortunately, as discussed in this paper, such oversimplification may lead to erroneous outputs. In order to preserve some nice mathematical properties, one must take care of the type of the t-norms being used.

In Section 2, a brief theoretical background for both binary and fuzzy association rules is provided. Section 2.3 discusses shortly naive definitions of lift, leverage, and conviction and shows a simple example where the rational interpretations of these measures become broken. Before providing corrected definitions in Section 4, an essential notion of *expected support* is analysed in Section 3. Finally, Section 5 concludes the paper with summarization of the achieved goals and drawings of the possible directions of future research.

2. Theoretical Background

2.1. Binary Association Rules. Let $\mathcal{O} := \{o_1, o_2, \dots, o_n\}$, $n > 0$, be a finite set of objects and let $\mathcal{A} := \{a_1, a_2, \dots, a_m\}$, $m > 0$, be a finite set of attributes (features). Each attribute can be considered as a logical predicate: $a_i(o_j)$ is true (or false) accordingly to whether the ith attribute applies (or does not apply) to object o_j. For a subset $X \subseteq \mathcal{A}$ of attributes, let us define a new predicate of a logical conjunction of the attributes contained in X as follows:

$$X(o_j) :\equiv \forall a_i \in X : a_i(o_j). \tag{1}$$

Moreover, let us define a negated predicate $\neg X(o_j)$ as follows:

$$\neg X(o_j) :\equiv \exists a_i \in X : \neg a_i(o_j). \tag{2}$$

An association rule is a formula $X \rightharpoonup Y$, where $X \subset \mathcal{A}$ is an *antecedent*, $Y \subset \mathcal{A}$ is a *consequent,* and $X \cap Y = \emptyset$. (Typically, $|X| \geq 1$ and $|Y| = 1$.) Both X and Y are sometimes called *itemsets*. Please consider the following rule as an example:

$$\{\text{tequila}, \text{salt}\} \longrightarrow \{\text{lemon}\}. \tag{3}$$

The *support* and *confidence* are defined as follows [2, 19]:

$$\operatorname{supp}(X) := \frac{|\{o \in \mathcal{O} \mid X(o)\}|}{n}, \tag{4}$$

$$\operatorname{supp}(X \rightharpoonup Y) := \frac{|\{o \in \mathcal{O} \mid X(o) \wedge Y(o)\}|}{n}, \tag{5}$$

$$\operatorname{conf}(X \rightharpoonup Y) := \frac{\operatorname{supp}(X \rightharpoonup Y)}{\operatorname{supp}(X)}, \tag{6}$$

where $n = |\mathcal{O}|$. Evidently, $\operatorname{supp}(\neg X) = 1 - \operatorname{supp}(X)$.

If $\mathcal{O}$ is a random sample, then observing $o \in \mathcal{O}$ is a random event. Then also a random event X may be defined on the basis of the truth value of the predicate $X(o)$ and support $\operatorname{supp}(X)$ becomes an estimate of a probability $P(\mathsf{X})$. Then also $\operatorname{supp}(\neg X)$ would correspond to the probability $P(\overline{\mathsf{X}})$, where $\overline{\mathsf{X}}$ is complementary event to X. Confidence $\operatorname{conf}(X \rightharpoonup Y)$ would then be an estimate of conditional probability $P(\mathsf{Y} \mid \mathsf{X})$.

Lift [12], *leverage* [15], and *conviction* [12] for binary data are defined as follows:

$$\operatorname{lift}(X \rightharpoonup Y) := \frac{\operatorname{supp}(X \rightharpoonup Y)}{\operatorname{supp}(X) \cdot \operatorname{supp}(Y)} = \frac{\operatorname{conf}(X \rightharpoonup Y)}{\operatorname{supp}(Y)}; \tag{7}$$

$$\operatorname{lever}(X \rightharpoonup Y) := \operatorname{supp}(X \rightharpoonup Y) - \operatorname{supp}(X) \cdot \operatorname{supp}(Y); \tag{8}$$

$$\operatorname{conv}(X \rightharpoonup Y) := \frac{\operatorname{supp}(X) \cdot \operatorname{supp}(\neg Y)}{\operatorname{supp}(X \rightharpoonup \neg Y)} = \frac{1 - \operatorname{supp}(Y)}{1 - \operatorname{conf}(X \rightharpoonup Y)}. \tag{9}$$

If X and Y are stochastically independent, $P(\mathsf{X} \wedge \mathsf{Y}) = P(\mathsf{X}) \cdot P(\mathsf{Y})$; that is, the expression $\operatorname{supp}(X) \cdot \operatorname{supp}(Y)$ is an estimation of support $\operatorname{supp}(X \rightharpoonup Y)$ under the assumption of X and Y being independent.

Hence, lift $\operatorname{lift}(X \rightharpoonup Y)$ is a ratio of observed support to the support that is expected under the assumption of independence, leverage $\operatorname{lever}(X \rightharpoonup Y)$ is a difference of observed and expected support, and conviction is a ratio of the expected support of X appearing without Y to the observed support $\operatorname{supp}(X \rightharpoonup \neg Y)$.

If X and Y are independent, $\operatorname{lift}(X \rightharpoonup Y) \simeq 1$, $\operatorname{conv}(X \rightharpoonup Y) \simeq 1$, and $\operatorname{lever}(X \rightharpoonup Y) \simeq 0$. The values of $\operatorname{lift}(X \rightharpoonup Y) > 1$, $\operatorname{conv}(X \rightharpoonup Y) > 1$, and $\operatorname{lever}(X \rightharpoonup Y) > 0$ indicate positive relationship, while $\operatorname{lift}(X \rightharpoonup Y) < 1$, $\operatorname{conv}(X \rightharpoonup Y) < 1$, and $\operatorname{lever}(X \rightharpoonup Y) < 0$ indicate negative relationship.

2.2. Fuzzy Association Rules. For fuzzy association rules, domain of each fuzzy attribute $a \in \mathcal{A}$ is not binary (or "crisp") $\{0, 1\}$ but graded (or "fuzzy"), that is, interval $[0, 1]$. That is, for each $a \in \mathcal{A}$ and $o \in \mathcal{O}$, $a(o) \in [0, 1]$. For a subset $X \subseteq \mathcal{A}$ of fuzzy attributes, we define a new predicate of a logical conjunction (similarly to binary case (1)) by using a t-norm $\otimes$ as

$$X(o_j) := \bigotimes_{a \in X} a(o_j). \tag{10}$$

T-norm $\otimes$ is a generalized logical conjunction, that is, a function $[0, 1] \times [0, 1] \rightarrow [0, 1]$ which is associative, commutative, and monotone increasing (in both places) and which satisfies the boundary conditions $\alpha \otimes 0 = 0$ and $\alpha \otimes 1 = \alpha$ for each $\alpha \in [0, 1]$. Some well-known examples of t-norms are as follows:

(i) product t-norm: $\otimes_{\text{prod}}(\alpha, \beta) = \alpha\beta$;
(ii) minimum t-norm: $\otimes_{\min}(\alpha, \beta) = \min(\alpha, \beta)$;
(iii) Łukasiewicz t-norm: $\otimes_{\text{Łuk}}(\alpha, \beta) = \max(0, \alpha + \beta - 1)$.

Negation of fuzzy predicate a is often defined for any $o \in \mathcal{O}$ as $\neg a(o) := 1 - a(o)$. Then, we can define (similarly to (2)) a negated predicate $\neg X$ as follows:

$$\neg X(o_j) := 1 - X(o_j). \tag{11}$$

Let $a \in \mathscr{A}$, $o \in \mathscr{O}$, $n = |\mathscr{O}|$, $n > 0$, $X, Y \subset \mathscr{A}$, $X \neq \emptyset$, $Y \neq \emptyset$, and $X \cap Y = \emptyset$. Several intensity measures may be defined as follows:

$$\begin{aligned} \operatorname{fsupp}(X) &:= \frac{\sum_{o \in \mathscr{O}} X(o)}{n}, \\ \operatorname{fsupp}(X \rightharpoonup Y) &:= \frac{\sum_{o \in \mathscr{O}} (X(o) \otimes Y(o))}{n}, \\ \operatorname{fconf}(X \rightharpoonup Y) &:= \frac{\operatorname{fsupp}(X \rightharpoonup Y)}{\operatorname{fsupp}(X)}. \end{aligned} \tag{12}$$

Convention 1. Throughout this text, we assume $\otimes$ is an arbitrary (but fixed) t-norm. Where it is important to explicitly specify a concrete t-norm, say $\otimes_{\text{prod}}$, we put $\otimes_{\text{prod}}$ in subscript and write, for example, $\operatorname{fsupp}_{\otimes_{\text{prod}}}(X \rightharpoonup Y)$ instead of $\operatorname{fsupp}(X \rightharpoonup Y)$ or $X_{\otimes_{\text{prod}}}(o)$ instead of $X(o)$.

Convention 2. For the sake of simplicity, we will sometimes express a fuzzy attribute as a vector of membership degrees. For instance, suppose $a \in \mathscr{A}$ is a fuzzy attribute on a set of objects $\mathscr{O} = \{o_1, o_2, o_3\}$. Then, instead of writing $a(o_1) := 0.3$, $a(o_2) := 0.9$, and $a(o_3) := 1$, we will use a much concise form: $a := (0.3, 0.9, 1)$.

Example 1. Let $a := (a_1, a_2, a_3)$, $b := (b_1, b_2, b_3)$, and $c := (c_1, c_2, c_3)$ be fuzzy attributes. Then,

$$\operatorname{fsupp}_{\otimes_{\text{prod}}}(\{a, b\} \rightharpoonup \{c\}) = \frac{1}{3} \sum_{i=1}^{3} \left(a_i \otimes_{\text{prod}} b_i \otimes_{\text{prod}} c_i \right). \tag{13}$$

2.3. Naive Definition of Lift, Leverage, and Conviction for Fuzzy Association Rules. A naive approach for introducing lift, leverage, and conviction into the fuzzy association rules framework is to use simply their definitions (7), (8), and (9) for binary rules and replace binary support (4) and (5) and confidence (6) with their fuzzy alternatives (12) as, for example, in [17, 18]. Unfortunately, that approach works well only for $\otimes$ being the product t-norm. As indicated in the following experiment, using minimum or Łukasiewicz t-norms may lead to erroneous interpretations.

Experiment 1. Two vectors X and Y of size $n = 2000$ were randomly generated from the uniform distribution on the interval $[0, 1]$ so that they are stochastically independent. Above-described naive versions of lift, leverage, and conviction of a rule $X \rightharpoonup Y$ were computed with using minimum, product, and Łukasiewicz t-norms as $\otimes$; see Table 1 for results.

As discussed in Section 2.1, stochastically independent data are expected to result in lift and conviction being close to 1 while leverage is expected to be close to 0. As can be seen, this is the case only for product t-norm.

The values of naive lift and naive leverage wrongly indicate positive (resp., negative) relationship, if the minimum (resp., Łukasiewicz) t-norm is used. Paradoxically, naive conviction indicates opposite sign of relationship.

Table 1: Comparison of lift, leverage, and conviction computed with different t-norms on stochastically independent data generated randomly from uniform distribution.

	Naive lift	Naive leverage	Naive conviction
Łukasiewicz t-norm	0.675	−0.081	1.544
Product t-norm	1.012	0.003	1.012
Minimum t-norm	1.353	0.088	0.755

3. Expected Support of a Conjunction of Fuzzy Attributes under the Assumption of Independence

As indicated in Experiment 1 presented in Section 2.3 above, naive lift, leverage, and conviction no more behave like their alternatives for crisp data: they no more represent a ratio of what is observed to what is expected under the assumption of independence. To recover their definitions, independency of fuzzy attributes must be treated correctly. Only then a proper definitions of lift, leverage, and conviction can be formulated.

Given sets X and Y of fuzzy attributes, what support of $X \wedge Y$ is expected if X and Y are independent? Moreover, what does independency of fuzzy attributes mean?

For the sake of simplicity, let us assume X and Y are sets containing a single fuzzy attribute; that is, $|X| = |Y| = 1$. For more complex cases, a new attribute can be created from the set of fuzzy attributes by using (10).

If $\mathscr{O}$ is a set of randomly selected objects, one can consider membership values $X(o)$ and $Y(o)$ as random variables X and Y and treat the independence of fuzzy attributes as stochastic independence of random variables X and Y.

Two random variables X, Y are stochastically independent, if the combined random variable (X, Y) has a joint probability density as

$$f_{\mathsf{X},\mathsf{Y}}(x, y) = f_{\mathsf{X}}(x) f_{\mathsf{Y}}(y). \tag{14}$$

If X and Y are two independent random variables from interval $[0, 1]$, then

$$\sigma(x, y) := \frac{x \otimes y}{n} \tag{15}$$

is a random variable with probability density function $f_\sigma(x, y) = f_{\mathsf{X},\mathsf{Y}}(x, y)$.

By definition, *expected value* $E[\mathsf{Z}]$ of a random variable Z is a weighted average of all possible values. More formally,

$$E[\mathsf{Z}] = \int_{-\infty}^{\infty} z f_{\mathsf{Z}}(z)\, \mathrm{d}z, \tag{16}$$

where f_{Z} is a probability density function of random variable Z.

Similarly, an expected value $E[\sigma(x, y)]$ is a weighted average of all possible (x, y) pairs, namely,

$$E[\sigma(x, y)] = \iint_0^1 \sigma(x, y) f_\sigma(x, y)\, \mathrm{d}x\, \mathrm{d}y. \tag{17}$$

In real setting, $f_\sigma(x, y)$ is unknown but we can estimate its values from data (i.e., from objects $\mathscr{O}$ and their fuzzy attributes $\mathscr{A}$) by using the assumption of independence (14) as follows:

$$f_\sigma(x,y) = f_{\mathsf{X}}(x)\, f_{\mathsf{Y}}(y) \approx \frac{\mathrm{count}_X(x)}{n} \cdot \frac{\mathrm{count}_Y(y)}{n}, \quad (18)$$

where $\mathrm{count}_A(a)$ is the number of objects from $\mathscr{O}$ that belong to $A \in \mathscr{A}$ with degree a; that is, $\mathrm{count}_A(a) = |\{o \in \mathscr{O} \mid A(o) = a\}|$.

Assuming $x \in \{X(o) \mid o \in \mathscr{O}\}$ and $y \in \{Y(o) \mid o \in \mathscr{O}\}$, we obtain, from (15), (17), and (18),

$$\begin{aligned} E[\sigma(x,y)] &\approx \iint_0^1 \frac{(x \otimes y)\cdot \mathrm{count}_X(x) \cdot \mathrm{count}_Y(y)}{n^3}\,\mathrm{d}x\,\mathrm{d}y \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{X(o_i)\otimes Y(o_j)}{n^3}. \end{aligned} \quad (19)$$

Since X and Y are independent, $E[\mathrm{fsupp}(X \rightharpoonup Y)] = n \cdot E[\sigma(x,y)]$ and hence expected value of $\mathrm{fsupp}(X \rightharpoonup Y)$ is

$$E[\mathrm{fsupp}(X \rightharpoonup Y)] \approx \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{X(o_i)\otimes Y(o_j)}{n^2}. \quad (20)$$

Now, we are ready to define notions of *expected support* and *expected confidence*.

Definition 2. Let $\otimes$ be a t-norm, and let X, Y be sets of fuzzy attributes such that $\mathrm{fsupp}(X) > 0$ and $n > 0$. Then, the *expected fuzzy support* $\widehat{\mathrm{fsupp}}(X \rightharpoonup Y)$ and the *expected fuzzy confidence* $\widehat{\mathrm{fconf}}(X \rightharpoonup Y)$ are defined as follows:

$$\begin{aligned} \widehat{\mathrm{fsupp}}(X \rightharpoonup Y) &:= \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{X(o_i)\otimes Y(o_j)}{n^2}, \\ \widehat{\mathrm{fconf}}(X \rightharpoonup Y) &:= \frac{\widehat{\mathrm{fsupp}}(X \rightharpoonup Y)}{\mathrm{fsupp}(X)}. \end{aligned} \quad (21)$$

Proposition 3. *Let X, Y be sets of fuzzy attributes. Then,*

(1) *if $fsupp(X \rightharpoonup Y) > 0$, then $\widehat{fsupp}(X \rightharpoonup Y) > 0$,*

(2) *$\widehat{fsupp}(X \rightharpoonup Y) \le \min(fsupp(X), fsupp(Y))$.*

Proof. (1) If $\mathrm{fsupp}(X \rightharpoonup Y) > 0$, then

$$\sum_{i=1}^{n} X(o_i) \otimes Y(o_i) > 0 \quad (22)$$

and hence

$$\sum_{i=1}^{n}\sum_{j=1}^{n} X(o_i) \otimes Y(o_j) > 0. \quad (23)$$

Therefore also $\widehat{\mathrm{fsupp}}(X \rightharpoonup Y) > 0$.

(2) For any t-norm $\otimes$ holds that $X(o_i) \otimes Y(o_j) \le Y(o_j)$. Therefore,

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{X(o_i)\otimes Y(o_j)}{n^2} \le \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{Y(o_j)}{n^2} = \mathrm{fsupp}(Y) \quad (24)$$

and similarly for $\mathrm{fsupp}(X)$. Hence, $\widehat{\mathrm{fsupp}}(X \rightharpoonup Y) \le \min(\mathrm{fsupp}(X), \mathrm{fsupp}(Y))$. □

Remark 4. Note that the reverse direction of the first implication of Proposition 3 is not generally true. For example, for $X := (1,0)$ and $Y := (0,1)$, we have $\widehat{\mathrm{fsupp}}(X \rightharpoonup Y) = 1/4$ but $\mathrm{fsupp}(X \rightharpoonup Y) = 0$.

Proposition 5. *Let X, Y be sets of fuzzy attributes and let $\otimes := \otimes_{prod}$; that is, the product t-norm is being used as a conjunction. Then,*

$$\widehat{fsupp}(X \rightharpoonup Y) = fsupp(X) \cdot fsupp(Y). \quad (25)$$

Proof. If $\otimes := \otimes_{\mathrm{prod}}$, $\widehat{\mathrm{fsupp}}(X \rightharpoonup Y)$ equals

$$\begin{aligned} &\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{X(o_i)\cdot Y(o_j)}{n^2} \\ &\quad = \frac{1}{n^2}\sum_{i=1}^{n} X(o_i) \sum_{i=1}^{n} Y(o_i) = \mathrm{fsupp}(X)\cdot\mathrm{fsupp}(Y). \end{aligned} \quad (26)$$

□

Proposition 6. *Let X, Y be sets of fuzzy attributes and let $\otimes := \otimes_{\min}$; that is, the minimum t-norm is being used as a conjunction. Then,*

$$\begin{aligned} fsupp(X)\cdot fsupp(X) &\le \widehat{fsupp}(X \rightharpoonup Y) \\ &\le fsupp(X) \otimes_{\min} fsupp(Y). \end{aligned} \quad (27)$$

Proof. Let $\otimes := \otimes_{\min}$. The first inequality follows from the fact that $\otimes_{\min}(x,y) \ge x \cdot y$ as

$$\begin{aligned} \widehat{\mathrm{fsupp}}(X \rightharpoonup Y) &= \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{X(o_i)\otimes_{\min} Y(o_j)}{n^2} \\ &\ge \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{X(o_i)\cdot Y(o_j)}{n^2} \\ &= \mathrm{fsupp}(X)\cdot\mathrm{fsupp}(Y). \end{aligned} \quad (28)$$

The second inequality follows directly from Proposition 3. □

Proposition 7. *Let X, Y be sets of fuzzy attributes and let $\otimes := \otimes_{Łuk}$; that is, the Łukasiewicz t-norm is being used as a conjunction. Then,*

$$\begin{aligned} fsupp(X) \otimes_{Łuk} fsupp(Y) &\le \widehat{fsupp}(X \rightharpoonup Y) \\ &\le fsupp(X)\cdot fsupp(Y). \end{aligned} \quad (29)$$

Proof. Let $\otimes := \otimes_{Łuk}$, $s_X := \mathrm{fsupp}(X)$, and $s_Y := \mathrm{fsupp}(Y)$. To prove the first inequality, it suffices to prove

$$\widehat{\mathrm{fsupp}}(X \rightharpoonup Y) \ge \max(0, s_X + s_Y - 1) \quad (30)$$

which is obvious for $s_X + s_Y \le 1$. Let us therefore assume $s_X + s_Y > 1$; then, (30) can be rewritten as

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{X(o_i)\otimes_{Łuk} Y(o_j)}{n^2} \ge s_X + s_Y - 1. \quad (31)$$

Then,

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{n} X(o_i) \otimes_{\text{Łuk}} Y(o_j) \\ = \sum_{i=1}^{n}\sum_{j=1}^{n} \left(X(o_i) + Y(o_j)\right) - \sum_{\forall k}(t_k) - |r|,\end{aligned} \tag{32}$$

where t is a sequence of numbers $(X(o_i) + Y(o_j) \mid X(o_i) + Y(o_j) < 1)$ and $r = \{(i, j) \mid X(o_i) + Y(o_j) \geq 1\}$, for $i, j \in \{1, 2, \ldots, n\}$. Since $|t| + |r| = n^2$ and each $t_k < 1$, we can immediately see that

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{n} X(o_i) \otimes_{\text{Łuk}} Y(o_j) \\ \geq \sum_{i=1}^{n}\sum_{j=1}^{n} \left(X(o_i) + Y(o_j)\right) - n^2 \\ = n^2 s_X + n^2 s_Y - n^2 = n^2 (s_X + s_Y - 1);\end{aligned} \tag{33}$$

hence, (30) holds.

The second inequality follows from $\otimes_{\text{Łuk}}(x, y) \leq x \cdot y$ because then we have

$$\begin{aligned}\widehat{\text{fsupp}}(X \rightharpoonup Y) = \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{X(o_i) \otimes_{\text{Łuk}} Y(o_j)}{n^2} \\ \leq \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{X(o_i) \cdot Y(o_j)}{n^2} = s_X \cdot s_Y.\end{aligned} \tag{34}$$

□

4. Correct Definition of Lift, Leverage, and Conviction for Fuzzy Association Rules

Now, we are ready to provide a correct definition of lift, leverage, and conviction in the framework of fuzzy association rules.

Definition 8. Let X, Y be sets of fuzzy attributes with $n := |\mathcal{O}| > 0$. Then, lift, leverage, and conviction of a fuzzy association rule $X \rightharpoonup Y$ are defined as follows:

$$\begin{aligned}\text{flift}(X \rightharpoonup Y) &:= \frac{\text{fsupp}(X \rightharpoonup Y)}{\widehat{\text{fsupp}}(X \rightharpoonup Y)}; \\ \text{flever}(X \rightharpoonup Y) &:= \text{fsupp}(X \rightharpoonup Y) - \widehat{\text{fsupp}}(X \rightharpoonup Y); \\ \text{fconv}(X \rightharpoonup Y) &:= \frac{\widehat{\text{fsupp}}(X \rightharpoonup \neg Y)}{\text{fsupp}(X \rightharpoonup \neg Y)}.\end{aligned} \tag{35}$$

Let us now study some interesting properties of the newly defined notions.

Proposition 9. *Let X, Y be sets of fuzzy attributes. Then,*

(1) *$flift(X \rightharpoonup Y) = flift(Y \rightharpoonup X)$;*

(2) *$flift(X \rightharpoonup Y) = fconf(X \rightharpoonup Y)/\widehat{fconf}(X \rightharpoonup Y)$;*

(3) *$0 \leq flift(X \rightharpoonup Y) \leq n$;*

(4) *if $fsupp(X \rightharpoonup Y) > 0$, then $flift(X \rightharpoonup Y) > 0$.*

Proof. (1) and (2) directly follow from the definitions and from the fact that t-norms are commutative.

(3) Since the membership degrees are defined on interval $[0, 1]$, their sums cannot be negative either. Hence, $\text{flift}(X \rightharpoonup Y) \geq 0$. Next, assume to the contrary that $\text{flift}(X \rightharpoonup Y) > n$ as follows:

$$\text{flift}(X \rightharpoonup Y) = n \frac{\sum_{i=1}^{n} X(o_i) \otimes Y(o_i)}{\sum_{i=1}^{n}\sum_{j=1}^{n} X(o_i) \otimes Y(o_j)} > n; \tag{36}$$

therefore,

$$\sum_{i=1}^{n} X(o_i) \otimes Y(o_i) > \sum_{i=1}^{n}\sum_{j=1}^{n} X(o_i) \otimes Y(o_j), \tag{37}$$

which is a contradiction.

(4) If $\text{fsupp}(X \rightharpoonup Y) > 0$, then, from Proposition 3, we know that also $\widehat{\text{fsupp}}(X \rightharpoonup Y) > 0$. Therefore, $\text{flift}(X \rightharpoonup Y)$ exists and is greater than 0. □

Proposition 10. *Let X, Y be sets of fuzzy attributes. Then,*

(1) *$flever(X \rightharpoonup Y) = flever(Y \rightharpoonup X)$,*

(2) *$flever(X \rightharpoonup Y) = fsupp(X)(fconf(X \rightharpoonup Y) - \widehat{fconf}(X \rightharpoonup Y))$,*

(3) *$1/n - 1 \leq flever(X \rightharpoonup Y) \leq fsupp(X \rightharpoonup Y)(1 - 1/n)$.*

Proof. (1) and (2) directly follow from the definitions and from the fact that t-norms are commutative.

(3) Let $u := \sum_{\forall i} X(o_i) \otimes Y(o_i)$ and $r := \sum_{\forall i}\sum_{\forall j \neq i} X(o_i) \otimes Y(o_j)$. Evidently, $\text{fsupp}(X \rightharpoonup Y) = u/n$ and $\widehat{\text{fsupp}}(X \rightharpoonup Y) = (u + r)/n^2$. Then, $\text{flever}(X \rightharpoonup Y) = u/n - (u + r)/n^2$. Obviously, $0 \leq u \leq n$ and $0 \leq r \leq n^2 - n$. Therefore,

$$\begin{aligned}-\frac{n^2 - n}{n^2} = \frac{1}{n} - 1 \\ \leq \text{flever}(X \rightharpoonup Y) \leq \frac{u}{n} - \frac{u}{n^2} \\ = \text{fsupp}(X \rightharpoonup Y)\left(1 - \frac{1}{n}\right).\end{aligned} \tag{38}$$

□

Proposition 11. *Let X, Y be sets of fuzzy attributes. Then,*

(1) *$fconv(X \rightharpoonup Y) = 1/flift(X \rightharpoonup \neg Y)$;*

(2) *$fconv(X \rightharpoonup Y) = fconv(\neg Y \rightharpoonup \neg X)$;*

(3) *$fconv(X \rightharpoonup Y) = \widehat{fconf}(X \rightharpoonup \neg Y)/fconf(X \rightharpoonup \neg Y)$;*

(4) *$1/n \leq fconv(X \rightharpoonup Y)$.*

Proof. Everything directly follows from Definition 8 and Proposition 9. □

Corollary 12. *Let X, Y be sets of fuzzy attributes and let $\otimes := \otimes_{prod}$; that is, the product t-norm is being used as a conjunction. Then,*

$$\begin{aligned} flift(X \rightharpoonup Y) &= \frac{fsupp\,(X \rightharpoonup Y)}{fsupp\,(X)\cdot fsupp\,(Y)} = \frac{fconf\,(X \rightharpoonup Y)}{fsupp\,(Y)}; \\ flever\,(X \rightharpoonup Y) &= fsupp\,(X \rightharpoonup Y) - fsupp\,(X)\cdot fsupp\,(Y); \\ fconv\,(X \rightharpoonup Y) &= \frac{fsupp\,(X)\cdot fsupp\,(\neg Y)}{fsupp\,(X \rightharpoonup \neg Y)} \\ &= \frac{1 - fsupp\,(Y)}{1 - fconf\,(X \rightharpoonup Y)}. \end{aligned} \tag{39}$$

Proof. Everything directly follows from Definitions 2 and 8, from Proposition 5, and from the fact that $\mathrm{fsupp}(\neg Y) = 1 - \mathrm{fsupp}(Y)$. □

Corollary 12 copies properties that are well known for crisp variants of lift, leverage, and conviction. In practice, the use of these equations is much more convenient than the original ones from Definition 8. However, note that Corollary 12 holds only if product t-norm $\otimes_{\mathrm{prod}}$ is used. For other t-norms such as minimum $\otimes_{\min}$ or Łukasiewicz $\otimes_{\text{Łuk}}$, Definition 8 must not be oversimplified that way. See the subsequent corollaries for more details.

Corollary 13. *Let X, Y be sets of fuzzy attributes and let $\otimes := \otimes_{\min}$; that is, the minimum t-norm is being used as a conjunction. Then,*

$$\begin{gathered} \frac{s_{XY}}{s_X \otimes_{\min} s_Y} \le flift\,(X \rightharpoonup Y) \le \frac{s_{XY}}{s_X \cdot s_Y}, \\ s_{XY} - s_X \otimes_{\min} s_Y \le flever\,(X \rightharpoonup Y) \le s_{XY} - s_X \cdot s_Y, \\ \frac{s_X \cdot s_{\overline{Y}}}{s_{X\overline{Y}}} \le fconv\,(X \rightharpoonup Y) \le \frac{s_X \otimes_{\min} s_{\overline{Y}}}{s_{X\overline{Y}}}, \end{gathered} \tag{40}$$

where $s_X := fsupp(X)$, $s_Y := fsupp(Y)$, $s_{\overline{Y}} := fsupp(\neg Y)$, $s_{XY} := fsupp(X \rightharpoonup Y)$, and $s_{X\overline{Y}} := fsupp(X \neg Y)$.

Proof. Everything directly follows from Definitions 2 and 8 and from Propositions 6 and 11. □

Corollary 14. *Let X, Y be sets of fuzzy attributes and let $\otimes := \otimes_{\text{Łuk}}$; that is, the Łukasiewicz t-norm is being used as a conjunction. Then,*

$$\begin{gathered} \frac{s_{XY}}{s_X \cdot s_Y} \le flift\,(X \rightharpoonup Y) \le \frac{s_{XY}}{s_X \otimes_{\text{Łuk}} s_Y}; \\ s_{XY} - s_X \cdot s_Y \le flever\,(X \rightharpoonup Y) \le s_{XY} - s_X \otimes_{\text{Łuk}} s_Y; \\ \frac{s_X \otimes_{\text{Łuk}} s_{\overline{Y}}}{s_{X\overline{Y}}} \le fconv\,(X \rightharpoonup Y) \le \frac{s_X \cdot s_{\overline{Y}}}{s_{X\overline{Y}}}, \end{gathered} \tag{41}$$

where $s_X := fsupp(X)$, $s_Y := fsupp(Y)$, $s_{\overline{Y}} := fsupp(\neg Y)$, $s_{XY} := fsupp(X \rightharpoonup Y)$, and $s_{X\overline{Y}} := fsupp(X \rightharpoonup \neg Y)$.

Table 2: Comparison of lift, leverage, and conviction computed with different t-norms on stochastically independent data generated randomly from uniform distribution.

	Naive lift	Naive leverage	Naive conviction
Łukasiewicz t-norm	1.022	0.004	1.021
Product t-norm	1.012	0.003	1.012
Minimum t-norm	1.010	0.003	1.011

Proof. Everything directly follows from Definitions 2 and 8 and from Propositions 7 and 11. □

Experiment 2. The same data as in Experiment 1 were processed accordingly to correct definitions of lift, leverage, and conviction (see Definition 8). The results can be found in Table 2. Since the data are randomly generated from uniform distribution, they are stochastically independent; hence, lift and conviction should be close to 1 and lift should be close to 0 regardless of t-norm being used. As one can see, the results in Table 2 are perfectly in compliance with our expectations.

5. Conclusion

Lift is a ratio of observed support (resp., confidence) to the support (resp., confidence) that is expected under the assumption of independence. Leverage is similar to lift, since it is a difference between observed and expected support. Conviction is often treated as an alternative to confidence; nevertheless, it is defined on the basis of observed and expected support too.

In this paper, a correct definition of lift, leverage, and conviction for fuzzy data was provided. It should be stressed here that there already exist some research papers that use incorrect (a.k.a. "naive") definition of fuzzy lift (e.g., [18]) that is also discussed here.

The naive definition does not preserve interpretation of positive/negative relationship. For crisp lift and conviction (resp., leverage), the stochastically independent attributes X and Y yield $\mathrm{lift}(X \rightharpoonup Y) \approx 1$, $\mathrm{conv}(X \rightharpoonup Y) \approx 1$, and $\mathrm{lever}(X \rightharpoonup Y) \approx 0$. As shown in Experiment 1, this is no more the case for naive definitions of these measures. On the other hand, Experiment 2 shows that correct definitions of lift, leverage, and conviction again recover that feature for fuzzy data.

All the lift, leverage, and conviction definitions are similar to their "crisp" alternatives (i.e., definitions for binary data) if the t-norm being used is the product $\otimes_{\mathrm{prod}}$. For Łukasiewicz $\otimes_{\text{Łuk}}$ and minimum $\otimes_{\min}$ t-norms, a more complicated computation takes place.

In [20], an algorithm was developed in for fast evaluation of fuzzy lift. A future research will therefore address improvements of association rules search algorithms by introducing fast computations of other measures, adding pruning heuristics based on boundary conditions provided by Corollaries 13 and 14. Also, other interest measures may be studied and their applicability to fuzzy rules may be considered.

Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This work was supported by the European Regional Development Fund in the Project of IT4Innovations Centre of Excellence (CZ.1.05/1.1.00/02.0070, VP6).

References

[1] M. Hahsler and K. Hornik, "New probabilistic interest measures for association rules," *Intelligent Data Analysis*, vol. 11, no. 5, pp. 437–455, 2007.

[2] R. Agrawal, T. Imielinski, and A. Swami, "Mining association rules between sets of items in large databases," in *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pp. 207–216, Washington, DC, USA, May 1993.

[3] A. Berrado and G. C. Runger, "Using metarules to organize and group discovered association rules," *Data Mining and Knowledge Discovery*, vol. 14, no. 3, pp. 409–431, 2007.

[4] G. I. Webb, "Discovering significant patterns," *Machine Learning*, vol. 68, no. 1, pp. 1–33, 2007.

[5] R. Srikant and R. Agrawal, "Mining quantitative association rules in large relational tables," *SIGMOD Record*, vol. 25, no. 2, pp. 1–12, 1996.

[6] V. Novák, I. Perfilieva, and J. Močkoř, *Mathematical Principles of Fuzzy Logic*, Kluwer, Boston, Mass, USA, 1999.

[7] H. Kalia, S. Dehuri, and A. Ghosh, "A survey on fuzzy association rule mining," *International Journal of Data Warehousing and Mining*, vol. 9, no. 1, pp. 1–27, 2013.

[8] K. C. C. Chen and W.-H. Au, *Mining Fuzzy Association Rules*, ACM, New York, NY, USA, 1997.

[9] V. Novák, I. Perfilieva, A. Dvořák, G. Chen, Q. Wei, and P. Yan, "Mining pure linguistic associations from numerical data," *International Journal of Approximate Reasoning*, vol. 48, no. 1, pp. 4–22, 2008.

[10] M. Burda, "Fast evaluation of t-norms for fuzzy association rules mining," in *Proceedings of the 14th IEEE International Symposium on Computational Intelligence and Informatics (CINTI '13)*, pp. 465–470, Budapest, Hungary, November 2013.

[11] M. Burda, V. Pavliska, and R. Valášek, "Parallel mining of fuzzy association rules on dense data sets," in *Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE '14)*, pp. 2156–2162, IEEE, Beijing, China, 2014.

[12] S. Brin, R. Motwani, J. D. Ullman, and S. Tsur, "Dynamic itemset counting and implication rules for market basket data," in *Proceedings ACM SIGMOD International Conference on Management of Data (SIGMOD '97)*, pp. 255–264, Tucson, Ariz, USA, 1997.

[13] P. D. McNicholas, T. B. Murphy, and M. O'Regan, "Standardising the lift of an association rule," *Computational Statistics & Data Analysis*, vol. 52, no. 10, pp. 4712–4721, 2008.

[14] P. Hájek, I. Havel, and M. Chytil, "The GUHA method of automatic hypotheses determination," *Computing*, vol. 1, no. 4, pp. 293–308, 1966.

[15] G. Piatetsky-Shapiro, "Discovery, analysis, and presentation of strong rules," in *Knowledge Discovery in Databases*, pp. 229–248, AAAI/MIT Press, 1991.

[16] S. Lallich, O. Teytaud, and E. Prudhomme, "Association rule interestingness: measure and statistical validation," in *Quality Measures in Data Mining*, vol. 43 of *Studies in Computational Intelligence*, pp. 251–275, Springer, Berlin, Germany, 2007.

[17] A. L. Buczak and C. M. Gifford, "Fuzzy association rule mining for community crime pattern discovery," in *Proceedings of the ACM SIGKDD Workshop on Intelligence and Security Informatics (ISI-KDD '10)*, vol. 2, pp. 1–10, ACM, New York, NY, USA, 2010.

[18] D. P. Pancho, J. M. Alonso, J. Alcalá-Fdez, and L. Magdalena, "Interpretability analysis of fuzzy association rules supported by fingrams," in *Proceedings of the 8th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT '13)*, pp. 469–474, September 2013.

[19] R. Agrawal, "Fast discovery of association rules," in *Advances in Knowledge Discovery and Data Mining*, pp. 307–328, AAAI Press/MIT Press, Cambridge, Mass, USA, 1996.

[20] M. Burda, "Lift measure for fuzzy association rules," in *Proceedings of the 7th International Conference on Soft Methods in Probability and Statistics (SMPS '2014)*, Springer, Warsaw, Poland, September 2014.

Hindawi Publishing Corporation
Advances in Fuzzy Systems
Volume 2014, Article ID 864074, 5 pages
http://dx.doi.org/10.1155/2014/864074

Research Article

Revised Max-Min Average Composition Method for Decision Making Using Intuitionistic Fuzzy Soft Matrix Theory

P. Shanmugasundaram,[1] C. V. Seshaiah,[2] and K. Rathi[1]

[1] *Department of Mathematics, Velalar College of Engineering & Technology, Tamil Nadu 638012, India*
[2] *Department of Mathematics, Sri Ramakrishna Engineering College, Tamil Nadu 641022, India*

Correspondence should be addressed to P. Shanmugasundaram; psserode@gmail.com

Received 15 July 2014; Revised 29 July 2014; Accepted 2 August 2014; Published 20 August 2014

Academic Editor: Salvatore Sessa

In this paper a revised Intuitionistic Fuzzy Max-Min Average Composition Method is proposed to construct the decision method for the selection of the professional students based on their skills by the recruiters using the operations of Intuitionistic Fuzzy Soft Matrices. In Shanmugasundaram et al. (2014), Intuitionistic Fuzzy Max-Min Average Composition Method was introduced and applied in Medical diagnosis problem. Sanchez's approach (Sanchez (1979)) for decision making is studied and the concept is modified for the application of Intuitionistic fuzzy soft set theory. Through a survey, the opportunities and selection of the students with the help of Intuitionistic fuzzy soft matrix operations along with Intuitionistic fuzzy max-min average composition method is discussed.

1. Introduction

Soft set theory was initiated by the Russian researcher Molodtsov [1]; he proposed soft set as a completely generic mathematical tool for modeling uncertainties. Maji et al. [2, 3] applied this theory to several directions for dealing with the problems in uncertainty and imprecision. Pei and Miao [4] and Chen et al. [5] improved the work of Maji et al. [3]. Yang and Ji [6] initiated a matrix representation of a fuzzy soft set and applied it in decision making problems. Borah et al. [7] and Neog and Sut [8] extended fuzzy soft matrix theory and its application. Chetia and Das [9–11] proposed intuitionistic fuzzy soft matrix theory. Rajarajeswari and Dhanalakshmi [12–14] proposed new definitions for intuitionistic fuzzy soft matrices.

In real life most of the existing mathematical tools for formal modeling, reasoning, and computing are crisp, deterministic, and precise in nature. The classical crisp mathematical tools are not capable of dealing with the problems involving uncertainty and imprecision. There are many mathematical tools available for modeling complex systems such as probability theory, fuzzy set theory, and interval mathematics. Probability theory is applicable only for a stochastically stable system. Interval mathematics is not sufficiently adaptable for problems withdifferent uncertainties. Setting the membership function value is always a problem in fuzzy set theory. Intuitionistic fuzzy soft set theory (IFSS) may be more applicable to deal with uncertainty and imprecision; the parameterization tool using fuzzy soft set theory enhances the flexibility of its applications. Initially, intuitionistic fuzzy max-min average composition method is used.

In this paper, a new approach is proposed to construct the decision method for student selection using intuitionistic fuzzy soft matrices. In order to make this, addition, subtraction, and the complement of an intuitionistic fuzzy soft matrix (IFSM) are applied. The result is obtained based on the maximum value in the score matrix. An attempt has been made to provide a formal model of the process to study the selection of professional students by using IFSM theory and implement it in the form of field recommendation system. This is the system by which the recruiters use their knowledge to infer the selection from the skills displayed by the test results of the students and based on their selection criteria. A new technique called intuitionistic fuzzy revised max-min average composition method is proposed. Through a case study, it is seen that the proposed method produces better results than the existing max-min composition method [15, 16].

2. Preliminaries

The basic definitions of intuitionistic fuzzy soft set theory that are useful for subsequent discussions are given.

Definition 1 (see [1]). Suppose that U is the universe of discourse and E is a set of parameters; let $P(U)$ denote the power set of U. A pair (F, E) is called a soft set over U, where F is a mapping given by $F: E \rightarrow P(U)$. Clearly, a soft set is a mapping from parameters to $P(U)$, and it is not a set, but a parameterized family of subsets of the universe of discourse.

Definition 2 (see [1, 2]). Let U be the universe of discourse and let E be the set of parameters. Let $A \subseteq E$. A pair (F, A) is called fuzzy soft set over U, where F is a mapping given by F: $A \rightarrow I^U$ and I^U denotes the collection of all fuzzy subsets of U.

Definition 3 (see [1, 2]). Let U be a universal set, E a set of parameters, and $A \subset E$. Let $F(U)$ denote the set of all fuzzy subsets of U. A fuzzy soft set (f_A, E) on the universe U is defined as the set of ordered pairs $(f_A, E) = \{(e, (f_A(e)) : e \in E, f_A(e) \in F(U)\}$, where $f_A : E \rightarrow F(U)$. Here, f_A is called an approximate function of the fuzzy soft set (f_A, E). The set $f_A(e)$ is called e approximate value set or e-approximate set which consists of related objects of the parameter $e \in E$.

Definition 4 (see [3]). Let U be an initial universe set and let E be the set of parameters. Let IF^U denote the collection of all intuitionistic fuzzy subsets of U. Let $A \subset E$. A pair $(F; A)$ is called an intuitionistic fuzzy soft set over U, where F is a mapping given by $F : A \rightarrow \mathrm{IF}^U$.

Definition 5 (see [12]). Let $U = \{c_1, c_2, c_3, \ldots, c_m\}$ be the universal set and let $E = \{e_1, e_2, e_3, \ldots, e_n\}$ be the set of parameters. Let $A \subseteq E$ and let (F, A) be a fuzzy soft set in the fuzzy soft class (U, E). Then fuzzy soft set (F, A) is a matrix form as $A_{m\times n} = [a_{ij}]_{m\times n}$ or $A = [a_{ij}]$, $i = 1, 2, \ldots, m$ and $j = 1, 2, 3, \ldots, n$, where

$$a_{ij} = \begin{cases} (\mu_j(c_i), \nu_j(c_i)) & \text{if } e_j \in A \\ (0,1) & \text{if } e_j \notin A. \end{cases} \tag{1}$$

$\mu_j(c_i)$ represents the membership of c_i in the intuitionistic fuzzy set $F(e_j)$.

$\nu_j(c_i)$ represents the nonmembership of c_i in the intuitionistic fuzzy set $F(e_j)$.

Example 6 (see [14]). Suppose that $U = \{P_1, P_2, P_3, P_4\}$ is a set of students and $E = \{S_1, S_2, S_3, S_4\}$ is a set of parameters, which stand for results, technical skills, analytical skills, presentation skills, and communication skills, respectively. Consider the mapping from parameters set to the set of all intuitionistic fuzzy subsets of power set U. Then intuitionistic fuzzy soft set (F, A) describes the skills of the students with respect to the given parameters, for finding the best student of an academic year. Consider $A = \{S_1, S_2, S_3, S_4\}$; then intuitionistic fuzzy soft set is

$$\begin{aligned} (F, A) = \{F(S_1) &= \{(P_1, 0.8, 0.1), (P_2, 0.3, 0.6), \\ &\quad (P_3, 0.4, 0.5), (P_4, 0.6, 0.1)\} \\ F(S_2) &= \{(P_1, 0.6, 0.1), (P_2, 0.4, 0.4), \\ &\quad (P_3, 0.5, 0.4), (P_4, 0.8, 0.1)\} \\ F(S_3) &= \{(P_1, 0.4, 0.6), (P_2, 0.4, 0.5), \\ &\quad (P_3, 0.8, 0.2), (P_4, 0.3, 0.4)\} \\ F(S_4) &= \{(P_1, 0.6, 0.1), (P_2, 0.7, 0.2), \\ &\quad (P_3, 0.6, 0.3), (P_4, 0.7, 0.2)\}\}. \end{aligned} \tag{2}$$

This intuitionistic fuzzy soft set is represented by the following intuitionistic fuzzy soft matrix:

$$A = \begin{array}{c} \\ P_1 \\ P_2 \\ P_3 \\ P_4 \end{array} \begin{array}{c} \begin{array}{cccc} S_1 & S_2 & S_3 & S_4 \end{array} \\ \begin{bmatrix} (0.8, 0.1) & (0.6, 0.1) & (0.4, 0.6) & (0.6, 0.1) \\ (0.3, 0.6) & (0.4, 0.4) & (0.4, 0.5) & (0.7, 0.2) \\ (0.4, 0.5) & (0.5, 0.4) & (0.8, 0.2) & (0.6, 0.3) \\ (0.6, 0.1) & (0.8, 0.1) & (0.3, 0.4) & (0.7, 0.2) \end{bmatrix} \end{array}. \tag{3}$$

Definition 7 (see [12]). If $A = [a_{ij}] \in \mathrm{IFSM}_{m\times n}$ and $B = [b_{ij}] \in \mathrm{IFSM}_{m\times n}$, then we define the addition and subtraction of intuitionistic fuzzy soft matrices of A and B as

$$\begin{aligned} A + B &= \{\max[\mu_A(a_{ij}), \mu_B(b_{ij})], \\ &\quad \min[\nu_A(a_{ij}), \nu_B(b_{ij})]\}, \quad \forall i, j \\ A - B &= \{\min[\mu_A(a_{ij}), \mu_B(b_{ij})], \\ &\quad \max[\nu_A(a_{ij}), \nu_B(b_{ij})]\}, \quad \forall i, j. \end{aligned} \tag{4}$$

Definition 8 (see [12]). Let $A = [a_{ij}] \in \mathrm{IFSM}_{m\times n}$, where $a_{ij} = (\mu_j(c_i), \nu_j(c_i))$ for all i, j. Then A^C is called an intuitionistic fuzzy soft complement matrix if $A^C = [d_{ij}]_{m\times n}$, where $d_{ij} = (\nu_j(c_i), \mu_j(c_i))$ for all i, j.

Definition 9 (see [13]). If $A = [a_{ij}] \in \mathrm{IFSM}_{m\times n}$, $B = [b_{jk}] \in \mathrm{IFSM}_{n\times p}$, then max min composition fuzzy soft matrix relation of A and B is defined as

$$A * B = [c_{ik}]_{m\times p}, \tag{5}$$

where

$$\begin{aligned} c_{ik} = \Big\{&\mathrm{Max}\Big\{\mathop{\mathrm{Min}}_j[\mu_A(a_{ij}), \mu_B(b_{jk})]\Big\}, \\ &\mathrm{Min}\Big\{\mathop{\mathrm{Max}}_j[\nu_A(a_{ij}), \nu_B(b_{jk})]\Big\}\Big\}. \end{aligned} \tag{6}$$

Definition 10. If $A = [a_{ij}] \in \text{IFSM}_{m\times n}$ and $B = [b_{jk}] \in \text{IFSM}_{n\times p}$, then a new operation called revised intuitionistic fuzzy max-min average composition for fuzzy soft matrix relation is defined as

$$A\Phi B = \left\{\text{Max}\left\{\frac{\mu_A(a_{ij}) + \mu_B(b_{jk})}{2}\right\}, \text{Min}\left\{\frac{\nu_A(a_{ij}) + \nu_B(b_{jk})}{2}\right\}\right\}, \quad \forall i, j. \tag{7}$$

Example 11. Consider that $A = \begin{pmatrix} (0.8,0.1) & (0.4,0.5) \\ (0.7,0.3) & (0.4,0.6) \end{pmatrix}$ and $B = \begin{pmatrix} (0.6,0.3) & (0.8,0.2) \\ (0.7,0.3) & (0.5,0.5) \end{pmatrix}$ are the two intuitionistic fuzzy soft matrices; then the addition, subtraction, complement, max-min composition, and max-min average composition of fuzzy soft matrix relations are

$$\begin{gathered} A + B = \begin{pmatrix} (0.8, 0.1) & (0.8, 0.2) \\ (0.7, 0.3) & (0.5, 0.5) \end{pmatrix} \\ A - B = \begin{pmatrix} (0.6, 0.3) & (0.4, 0.5) \\ (0.7, 0.3) & (0.4, 0.6) \end{pmatrix} \\ A^C = \begin{pmatrix} (0.1, 0.8) & (0.5, 0.4) \\ (0.3, 0.7) & (0.6, 0.4) \end{pmatrix} \\ A * B = \begin{pmatrix} (0.6, 0.3) & (0.8, 0.2) \\ (0.6, 0.3) & (0.7, 0.3) \end{pmatrix} \\ A\Phi B = \begin{pmatrix} (0.70, 0.20) & (0.80, 0.15) \\ (0.65, 0.3) & (0.75, 0.25) \end{pmatrix}. \end{gathered} \tag{8}$$

Definition 12. If $A = [a_{ij}] \in \text{IFSM}_{m\times n}$, $B = [b_{ij}] \in \text{IFSM}_{m\times n}$, and A^C, B^C are the complement of A and B then the score matrix of A and B is defined as $S(A, B) = (1/2)[V + W]$, where V is the matrix defined as $V = [(\mu_{A\Phi B} - \nu_{A^C\phi B^C})]$ and W is the matrix defined as $W = [(\mu_{A^C\Phi B^C} - \nu_{A\phi B})]$.

3. Intuitionistic Fuzzy Max-Min Average Composition Method for Decision Making

In this section an application of intuitionistic fuzzy soft set theory using max-min average composition method for decision making is presented.

In a given set of systems, let $P = \{P_1, P_2, \ldots, P_m\}$ be the set of m students and let $S = \{S_1, S_2, \ldots, S_n\}$ be the set of n skills and let $D = \{D_1, D_2, \ldots, D_k\}$ be the set of k opportunities.

Now construct an intuitionistic fuzzy soft set (F, S) over P, where F is a mapping $F : S \to \text{IF}^P$ and IF^P is the collection of all intuitionistic fuzzy subsets of P. This intuitionistic fuzzy soft set gives a matrix A called student-skills matrix (as in Example 6).

Then construct another intuitionistic fuzzy soft set (G, D) over S, where G is a mapping $G : D \to \text{IF}^S$ and IF^S is the collection of all intuitionistic fuzzy subsets of S. This intuitionistic fuzzy soft set gives a matrix B called skills-opportunities matrix, where each element denotes the weight of the skills for certain opportunities.

Using Definition 8, obtain the intuitionistic fuzzy soft complement matrices A^c, B^c.

Using Definitions 10 and 12, compute $A\Phi B$, $A^c\Phi B^c$, and the score matrix $S(A, B)$.

Finally find the maximum value for each student P_i in the score matrix and then conclude that the student P_i is suitable for the opportunity D_j.

3.1. Algorithm

Step 1. Input the intuitionistic fuzzy soft sets (F, S), (G, D) and obtain the intuitionistic fuzzy soft matrices A, B corresponding to (F, S) and (G, D), respectively.

Step 2. Using Definition 8, obtain the intuitionistic fuzzy soft complement matrices A^c, B^c.

Step 3. Using Definition 10, compute the intuitionistic fuzzy max-min average compositions $A\Phi B$ and $A^c\Phi B^c$.

Step 4. Compute the matrices V, W and obtain the score matrix $S(A, B)$ using Definition 12.

Step 5. Identify the maximum score in $S(A, B)$ for each student P_i to select the suitable opportunity.

4. Case Study

Consider $P = \{P_1, P_2, P_3, P_4\}$ as the universal set, where P_1, P_2, P_3, and P_4 represent the set of students Arun, John, Peter, and Ram. Again consider $S = \{S_1, S_2, S_3, S_4\}$ as the set of skills, where S_1, S_2, S_3, S_4 represent technical skills, analytical skills, presentation skills, and communication skills, respectively, for the case study. Let the possible opportunities relating to the above skills be $D = \{D_1, D_2, D_3\}$, where D_1, D_2, D_3 represent hardware, software, and others.

Suppose that IFSS (F, S) over P, where F is a mapping $F : S \to \text{IF}^P$, gives the description of student-skills relation based on the test result by using intuitionistic fuzzy matrix relation in the form of IFS:

$$\begin{aligned} (F, S) = \{F(S_1) &= \{(P_1, 0.8, 0.1), (P_2, 0.3, 0.6), (P_3, 0.4, 0.5), (P_4, 0.6, 0.1)\} \\ F(S_2) &= \{(P_1, 0.6, 0.1), (P_2, 0.4, 0.4), (P_3, 0.5, 0.4), (P_4, 0.8, 0.1)\} \\ F(S_3) &= \{(P_1, 0.4, 0.6), (P_2, 0.4, 0.5), (P_3, 0.8, 0.2), (P_4, 0.3, 0.4)\} \\ F(S_4) &= \{(P_1, 0.6, 0.1), (P_2, 0.7, 0.2), (P_3, 0.6, 0.3), (P_4, 0.7, 0.2)\}\}. \end{aligned} \tag{9}$$

This intuitionistic fuzzy soft set is represented by the following intuitionistic fuzzy soft matrix:

$$A = \begin{array}{c} \\ P_1 \\ P_2 \\ P_3 \\ P_4 \end{array} \begin{array}{c} \begin{array}{cccc} S_1 & S_2 & S_3 & S_4 \end{array} \\ \begin{bmatrix} (0.8, 0.1) & (0.6, 0.1) & (0.4, 0.6) & (0.6, 0.1) \\ (0.3, 0.6) & (0.4, 0.4) & (0.4, 0.5) & (0.7, 0.2) \\ (0.4, 0.5) & (0.5, 0.4) & (0.8, 0.2) & (0.6, 0.3) \\ (0.6, 0.1) & (0.8, 0.1) & (0.3, 0.4) & (0.7, 0.2) \end{bmatrix} \end{array}. \quad (10)$$

Suppose that IFSS (G, D) over S, where G is a mapping $G : D \rightarrow \mathrm{IF}^S$, gives the weight of the skills for certain opportunities by field expert using intuitionistic fuzzy matrix relation in the form of IFS:

$$\begin{aligned} \text{Let } (G, D) = \{G(D_1) &= \{(S_1, 0.7, 0.2), (S_2, 0.8, 0.1), \\ &(S_3, 0.4, 0.3), (S_4, 0.5, 0.3)\} \\ G(D_2) &= \{(S_1, 0.5, 0.4), (S_2, 0.4, 0.5), \\ &(S_3, 0.8, 0.1), (S_4, 0.6, 0.2)\} \\ G(D_3) &= \{(S_1, 0.4, 0.5), (S_2, 0.6, 0.3), \\ &(S_3, 0.4, 0.4), (S_4, 0.7, 0.1)\}\}. \end{aligned} \quad (11)$$

This intuitionistic fuzzy soft set is represented by the following intuitionistic fuzzy soft matrix:

$$B = \begin{array}{c} \\ S_1 \\ S_2 \\ S_3 \\ S_4 \end{array} \begin{array}{c} \begin{array}{ccc} D_1 & D_2 & D_3 \end{array} \\ \begin{bmatrix} (0.7, 0.2) & (0.5, 0.4) & (0.4, 0.5) \\ (0.8, 0.1) & (0.4, 0.5) & (0.6, 0.3) \\ (0.4, 0.3) & (0.8, 0.1) & (0.4, 0.4) \\ (0.5, 0.3) & (0.6, 0.2) & (0.7, 0.1) \end{bmatrix} \end{array}. \quad (12)$$

Then calculate the intuitionistic fuzzy soft complement matrices using Definition 8 as follows:

$$A^C = \begin{array}{c} \\ P_1 \\ P_2 \\ P_3 \\ P_4 \end{array} \begin{array}{c} \begin{array}{cccc} S_1 & S_2 & S_3 & S_4 \end{array} \\ \begin{bmatrix} (0.1, 0.8) & (0.1, 0.6) & (0.6, 0.4) & (0.1, 0.6) \\ (0.6, 0.3) & (0.4, 0.4) & (0.5, 0.4) & (0.2, 0.7) \\ (0.5, 0.4) & (0.4, 0.5) & (0.2, 0.8) & (0.3, 0.6) \\ (0.1, 0.6) & (0.1, 0.8) & (0.4, 0.3) & (0.2, 0.7) \end{bmatrix} \end{array}.$$

$$B^C = \begin{array}{c} \\ S_1 \\ S_2 \\ S_3 \\ S_4 \end{array} \begin{array}{c} \begin{array}{ccc} D_1 & D_2 & D_3 \end{array} \\ \begin{bmatrix} (0.2, 0.7) & (0.4, 0.5) & (0.5, 0.4) \\ (0.1, 0.8) & (0.5, 0.4) & (0.3, 0.6) \\ (0.3, 0.4) & (0.1, 0.8) & (0.4, 0.4) \\ (0.3, 0.5) & (0.2, 0.6) & (0.1, 0.7) \end{bmatrix} \end{array} \quad (13)$$

Then the intuitionistic fuzzy max-min average composition relation matrices are obtained using Definition 10 :

$$A\Phi B = \begin{array}{c} \\ p_1 \\ p_2 \\ p_3 \\ p_4 \end{array} \begin{array}{c} \begin{array}{ccc} D_1 & D_2 & D_3 \end{array} \\ \begin{bmatrix} (0.75, 0.10) & (0.65, 0.15) & (0.60, 0.10) \\ (0.60, 0.25) & (0.65, 0.20) & (0.70, 0.15) \\ (0.60, 0.25) & (0.80, 0.20) & (0.65, 0.20) \\ (0.80, 0.10) & (0.65, 0.20) & (0.70, 0.15) \end{bmatrix} \end{array},$$

$$A^c\Phi B^c = \begin{array}{c} \\ p_1 \\ p_2 \\ p_3 \\ p_4 \end{array} \begin{array}{c} \begin{array}{ccc} D_1 & D_2 & D_3 \end{array} \\ \begin{bmatrix} (0.45, 0.40) & (0.40, 0.50) & (0.50, 0.40) \\ (0.40, 0.40) & (0.50, 0.40) & (0.50, 0.35) \\ (0.35, 0.55) & (0.45, 0.45) & (0.50, 0.40) \\ (0.35, 0.35) & (0.30, 0.55) & (0.40, 0.35) \end{bmatrix} \end{array}. \quad (14)$$

Find the matrices V and W using Definition 12 as follows:

$$V = \left[\left(\mu_{A\Phi B} - \nu_{A^C\phi B^C}\right)\right] = \begin{array}{c} \\ P_1 \\ P_2 \\ P_3 \\ P_4 \end{array} \begin{array}{c} \begin{array}{ccc} D_1 & D_2 & D_3 \end{array} \\ \begin{bmatrix} 0.35 & 0.15 & 0.20 \\ 0.20 & 0.25 & 0.35 \\ 0.05 & 0.35 & 0.25 \\ 0.45 & 0.10 & 0.35 \end{bmatrix} \end{array},$$

$$W = \left[\left(\mu_{A^C\Phi B^C} - \nu_{A\phi B}\right)\right] = \begin{array}{c} \\ P_1 \\ P_2 \\ P_3 \\ P_4 \end{array} \begin{array}{c} \begin{array}{ccc} D_1 & D_2 & D_3 \end{array} \\ \begin{bmatrix} 0.35 & 0.25 & 0.40 \\ 0.15 & 0.30 & 0.35 \\ 0.10 & 0.25 & 0.30 \\ 0.25 & 0.10 & 0.25 \end{bmatrix} \end{array}. \quad (15)$$

Using Definition 12, the score matrix for intuitionistic fuzzy max-min average composition method is

$$S(A, B) = \frac{1}{2}[V + W] = \begin{array}{c} \\ P_1 \\ P_2 \\ P_3 \\ P_4 \end{array} \begin{array}{c} \begin{array}{ccc} D_1 & D_2 & D_3 \end{array} \\ \begin{bmatrix} \mathbf{0.35} & 0.20 & 0.30 \\ 0.18 & 0.28 & \mathbf{0.35} \\ 0.08 & \mathbf{0.30} & 0.28 \\ \mathbf{0.35} & 0.10 & 0.30 \end{bmatrix} \end{array}. \quad (16)$$

It is clear from the above result that the interviewer agrees that the max values of Arun (P_1) and Ram (P_4) are 0.35; they are suitable for hardware profession whereas the max value of Peter (P_3) is 0.30 who is suitable for software profession. John (P_2) is suitable for the profession other than hardware and software. This is represented by the chart diagram as shown in Figure 1.

Using Definition 9 and the method discussed in [13], the same problem is solved and the respective score matrix denoted by $S_{\text{max-min}}$ is obtained as follows:

$$S_{\text{max-min}} = \begin{array}{c} \\ P_1 \\ P_2 \\ P_3 \\ P_4 \end{array} \begin{array}{c} \begin{array}{ccc} D_1 & D_2 & D_3 \end{array} \\ \begin{bmatrix} \mathbf{0.60} & 0.40 & 0.50 \\ 0.30 & \mathbf{0.40} & \mathbf{0.40} \\ 0.30 & \mathbf{0.60} & 0.30 \\ \mathbf{0.70} & 0.40 & 0.50 \end{bmatrix} \end{array}. \quad (17)$$

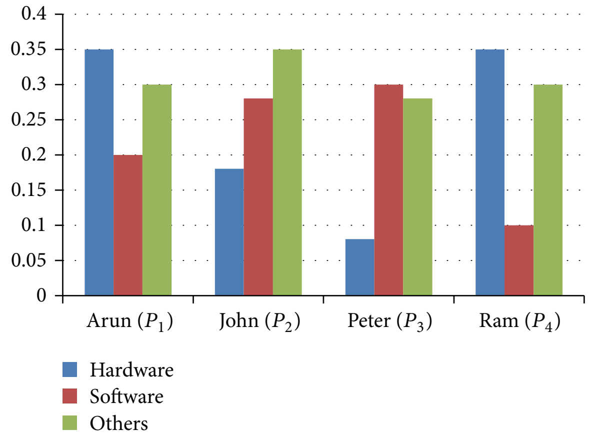

Figure 1

5. Conclusions

Compared with conventional techniques, the proposed approach in decision making effectively reduces the conflict in making decisions. For example, the maximum score in the second row of the score matrix $S_{\text{max-min}}$ is 0.40 which corresponds to more than one opportunity and there arises a conflict in making decision whereas it is not the case in the proposed method. Also, our approach makes it possible to introduce weights for all skills and reduces the confusion about the possibility of two opportunities in a student and also the proposed method is an effective tool for the decision making problems. In the future many decision making problems involving three or more sets of relation can be effectively modeled using our approach.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] D. Molodtsov, "Soft set theory-first results," *Computers and mathematics with Applications*, vol. 37, no. 4-5, pp. 19–31, 1999.

[2] P. K. Maji, R. Biswas, and A. R. Roy, "Fuzzy soft sets," *Journal of Fuzzy Mathematics*, vol. 9, no. 3, pp. 589–602, 2001.

[3] P. K. Maji, A. R. Roy, and R. Biswas, "On intuitionistic fuzzy soft sets," *Journal of Fuzzy Mathematics*, vol. 12, no. 3, pp. 669–683, 2004.

[4] D. Pei and D. Miao, "From soft sets to information systems," in *Proceedings of the IEEE International Conference on Granular Computing*, vol. 2, pp. 617–621, July 2005.

[5] D. Chen, E. C. C. Tsang, D. S. Yeung, and X. Wang, "The parameterization reduction of soft sets and its applications," *Computers and Mathematics with Applications*, vol. 49, no. 5-6, pp. 757–763, 2005.

[6] Y. Yang and C. Ji, "Fuzzy soft matrices and their applications," in *Artificial Intelligence and Computational Intelligence*, vol. 7002 of *Lecture Notes in Computer Science*, pp. 618–627, 2011.

[7] M. J. Borah, T. J. Neog, and D. K. Sut, "Fuzzy soft matrix theory and its decision making," *International Journal of Modern Engineering Research*, vol. 2, pp. 121–127, 2012.

[8] T. J. Neog and D. K. Sut, "An application of fuzzy soft sets in decision making problems using fuzzy soft matrices," *International Journal of Mathematical Archive*, vol. 2, pp. 2258–2263, 2012.

[9] B. Chetia and P. K. Das, "On Fuzzy Soft Matrix Theory," *Journal of the Assam Academy of Mathematics*, vol. 2, pp. 71–83, 2010.

[10] B. Chetia and P. K. Das, "On Fuzzy Soft Sets," *International Journal of Mathematical Archive*, vol. 2, pp. 1–6, 2011.

[11] B. Chetia and P. K. Das, "Some results of intuitionistic fuzzy soft matrix theory," *Pelagia Research Library, Advances in Applied Science Research*, vol. 3, no. 1, pp. 412–423, 2012.

[12] P. Rajarajeswari and P. Dhanalakshmi, "Intuitionistic fuzzy soft matrix theory and its application in decision making," *International Journal of Engineering Research and Technology*, vol. 2, no. 4, pp. 1100–1111, 2013.

[13] P. Rajarajeswari and P. Dhanalakshmi, "Intuitionistic fuzzy soft matrix theory and its application in medical diagnosis," *Annals of Fuzzy Mathematics and Informatics*, vol. 2, pp. 1–11, 2013.

[14] P. Rajarajeswari and P. Dhanalakshmi, "Similarity measures of intuitionistic fuzzy soft sets and their application in medical diagnosis," *International Journal of Mathematical Archive*, vol. 5, no. 5, pp. 143–149, 2014.

[15] P. Shanmugasundaram and C. V. Seshaiah, "An application of intuitionistic fuzzy technique in medical diagnosis," *Australian Journal of Basic and Applied Sciences*, vol. 8, no. 9, pp. 392–395, 2014.

[16] P. Shanmugasundaram, C. V. Seshaiah, and K. Rathi, "Intuitionistic fuzzy soft matrix theory in medical diagnosis using max-min average composition method," Accepted for publication in Journal of Theoretical and Applied Information Technology, 2014.