Knowledge-Based Systems 82 (2015) 115–127

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Novel correlation coefficients between hesitant fuzzy sets and their application in decision making

Huchang Liao [a], Zeshui Xu [a,*], Xiao-Jun Zeng [b]

[a] Business School, Sichuan University, Chengdu, Sichuan 610065, China
[b] School of Computer Science, University of Manchester, Manchester M13 9PL, United Kingdom

ARTICLE INFO

Article history:
Received 18 October 2014
Received in revised form 3 February 2015
Accepted 23 February 2015
Available online 3 March 2015

Keywords:
Hesitant fuzzy set
Correlation coefficient
Decision making
Medical diagnosis
Cluster analysis

ABSTRACT

Hesitant fuzzy set (HFS) is now attracting more and more scholars' attention due to its efficiency in representing comprehensively uncertain and vague information. Considering that correlation coefficient is one of the most widely used indices in data analysis, in this paper, after pointing out the weakness of the existing correlation coefficients between HFSs, we propose a novel correlation coefficient formulation to measure the relationship between two HFSs. As a departure, some new concepts, such as the mean of a hesitant fuzzy element (HFE), the hesitant degree of a HFE, the mean of a HFS, the variance of a HFS and the correlation between two HFSs are defined. Based on these concepts, a novel correlation coefficient formulation between two HFSs is developed. Afterwards, the upper and lower bounds of the correlation coefficient are defined. A theorem is given to determine these two bounds. It is stated that the correlation coefficient between two HFSs should also be hesitant and thus the upper and lower bounds can further help to identify the correlation coefficient between HFSs. The significant characteristic of the introduced correlation coefficient is that it lies in the interval $[-1,1]$, which is in accordance with the classical correlation coefficient in statistics, whereas all the old correlation coefficients between HFSs in the literature are within unit interval $[0,1]$. The weighted correlation coefficient is also proposed to make it more applicable. In order to show the efficiency of the proposed correlation coefficients, they are implemented in medical diagnosis and cluster analysis. Some numerical examples are given to support the findings and also illustrate the applicability and efficiency of the proposed correlation coefficient between HFSs.

1. Introduction

As uncertainty takes place almost everywhere in our daily life, many different tools have been developed to recognize, represent, manipulate, and tackle such uncertainty. Among the most popular theories to handle uncertainty include the probability theory and fuzzy set theory, which are proposed to interpret statistical uncertainty and fuzzy uncertainty, respectively. These two types of models possess philosophically different kinds of information: the probability theory conveys information about relative frequencies, while the fuzzy set theory represents similarities of objects to imprecisely defined properties [1]. After introduced by Zadeh [2], fuzzy set has attracted the attention of many people from academia, industry and government. The fuzzy set theory is now more and more popular in aiding decision making process [3,4]. A fuzzy set A on a universe of objects, X, is characterized by a membership function μ_A which takes the values in the interval $[0,1]$, i.e., $\mu_A : X \to [0,1]$. The value of μ_A at x, $\mu_A(x)$, represents the grade of membership of x in A and is a point in $[0,1]$ [5]. After the pioneering work of Zadeh, the fuzzy set theory has been extended in a number of directions, the most impressive one of which relates to the representation of the membership grades of the underlying fuzzy set [6]. As in Zadeh's fuzzy set, the membership grades in a fuzzy set are expressed as precise values drawn from the unit interval $[0,1]$, the fuzzy set cannot capture the human ability in expressing imprecise and vague membership grades of the fuzzy set. To circumvent this flaw, some notable extensions have been introduced to allow for imprecision and uncertainty in the membership grades of the fuzzy set, such as the type 2 fuzzy set [7], the interval type-2 fuzzy set [8], the intuitionistic fuzzy set [9–11], the interval-valued intuitionistic fuzzy set [12], the hesitant fuzzy set [13], and the hesitant fuzzy linguistic term set [14–16]. In this paper, we focus on the hesitant fuzzy set.

The motivation for introducing hesitant fuzzy set (HFS) is that when defining the membership grade of an element x to the set

* Corresponding author. Tel.: +86 28 85414255.
E-mail addresses: liaohuchang@163.com (H. Liao), xuzeshui@263.net (Z. Xu).

http://dx.doi.org/10.1016/j.knosys.2015.02.020
0950-7051/

A, the difficulty in establishing such a grade is not because we have some possibility distribution (as in type 2 fuzzy set), or a margin of error (as in interval fuzzy set and intuitionistic fuzzy set), but because we have a set of possible values [13]. HFS, permitting the membership grade of an element to a given set represented by several possible values between 0 and 1, is capable of determining the membership grades especially when experts have several different values on it. HFS shows many advantages over traditional fuzzy set and its other extensions, especially in group decision making with anonymity [17]. Thus, it has attracted many scholars. After introducing HFS, Torra [13] gave some basic operations such as complement, union and intersection on HFSs. Xia and Xu [17] named the elements of a HFS as hesitant fuzzy elements (HFEs) and then defined the addition and multiplication operations over HFEs. Afterwards, Liao and Xu [18] established the subtraction and division operation for HFSs. Torra and Narukawa [19] presented an extension principle permitting to generalize the existing operations on fuzzy sets to HFSs, based on which, different aggregation operators, such as the HFWA, HFWG, HFOWA, HFOWG, HFHA, HFHG operators [17], were proposed to fuse the hesitant fuzzy information in multiple criteria decision making (MCDM). Later, Liao and Xu [20,21] further proposed some hesitant fuzzy hybrid weighted aggregation operators, which have many advantages over the aforementioned operators. Liao et al. [22] also proposed some dynamic aggregation operators to aggregate multiple stage hesitant fuzzy information. In order to better aid the decision making process, Liao et al. [23] introduced the hesitant fuzzy preference relation (HFPR) and investigated its multiplicative consistency. Based on the HFPR, Zhu and Xu [24] proposed a goal programming method to derive a ranking from a HFPR and applied it to group decision making. Some decision making methods have been developed for decision making with hesitant fuzzy formation, such as the hesitant fuzzy VIKOR method [25], the hesitant fuzzy TOPSIS method [26], the hesitant fuzzy TODIM method [27] and the satisfaction degree based interactive decision making method [28]. All these achievements contribute to the theory of HFS and show that HFS is a powerful tool to handle uncertainty and vagueness in the process of decision making.

Correlation is one of the most widely used indices in data analysis, pattern recognition, machine learning, decision making, etc. It measures how well two variables move together in a linear fashion. The correlation coefficient, which firstly appeared in Karl Pearson's proposal relating to statistics, has been extended into different fuzzy circumstances. Different forms of fuzzy correlation and correlation coefficients have been proposed, such as the fuzzy correlation and correlation coefficients [29–34], the intuitionistic fuzzy correlation and correlation coefficients [35–41], and the hesitant fuzzy correlation and correlation coefficients [42,43] (for details of the review over these different forms of correlation measures, please refer to Section 2). As presented above, HFS is efficient in expressing vagueness and uncertainty. In this paper, we focus on the correlation measures between HFSs.

To measure the correlation between two HFSs, Xu and Xia [42] proposed several correlation coefficient formulas for HFEs and discussed their properties. Later, Chen et al. [43] defined some correlation coefficient formulas for HFSs and applied them to cluster analysis. However, the correlation coefficients they introduced have several flaws, which are clarified as follows:

(1) Both of Xu and Xia [42]'s and Chen et al. [43]'s correlation coefficient formulas are based on the assumption that each pair of HFEs has equal length. If the lengths of different HFEs are different, the pessimistic and optimistic principles should be implemented to fill some artificial values into the short HFEs. It should be stated that filling some artificial values into a HFE would change the information of the original HFE. Both of the mean and variance of the original HFE are replaced. Thus, such an approach is less well justified theoretically and less reliable practically.
(2) In traditional statistics, the correlation coefficient lies in the interval $[-1,1]$. However, the correlation coefficients proposed by Chen et al. [43] are always positive but ignore the negative correlation situation. This also reduces the theoretical support and applicability of their proposed hesitant fuzzy correlation coefficients.
(3) As HFEs or HFSs considered are hesitant, the correlation coefficient should have certain degree of hesitant rather than just a fixed value. However, such a hesitant feature has not been considered or modeled in Xu and Xia [42]'s or Chen et al. [43]'s correlation coefficient formulas.

To circumvent these drawbacks, in this paper, we propose some different correlation measures for HFSs. The novelties of this paper are summarized as follows:

(1) We introduce the mean and hesitant degree for a HFE.
(2) We propose a novel formula to calculate the correlation coefficient between two HFSs and compare it with other definitions reported in the literature. The new correlation coefficient proposed in this paper lies in the interval $[-1,1]$.
(3) We define the hesitant degree of a correlation coefficient between two HFSs.
(4) We introduce the weighted correlation coefficient between HFSs to broaden the applicability in practical decision making.
(5) We apply the proposed correlation coefficients between HFSs to practical decision making processes such as medical diagnosis and cluster analysis.

The rest of this paper is organized as follows: Section 2 reviews the concepts of HFS as well as the state of the art in correlation coefficients related to fuzzy sets and their extensions. The weakness of the existing correlation coefficients between HFSs is also clarified in this section. In Section 3, based on some predefined concepts, a novel correlation coefficient definition is proposed. A theorem is given to determine the lower and upper bounds of the correlation coefficient. The weighted correlation coefficients are also established. The correlation coefficient between HFSs is applied into medical diagnosis and cluster analysis in Section 4. Some numerical examples are given to illustrate the correlation coefficient in this section as well. The paper ends with some concluding remarks in Section 5.

2. Preliminaries

To facilitate the presentation, let us first review some basic knowledge.

2.1. Hesitant fuzzy set

As the membership grades in classical fuzzy set are expressed as precise values draw from [0,1], different extensions of fuzzy set which allow for imprecision and uncertainty in the membership grades have been proposed. The most notable one is the hesitant fuzzy set. When an expert considers several values to determine the membership degree of an element to a set, the concept of HFS was introduced:

Definition 1 [13]. Let X be a reference set and $M=\{\mu_1, \mu_2,\ldots,\mu_N\}$ be a set of N membership functions. Then, a hesitant fuzzy set (HFS) associated with M, that is H_M, is defined as follows: