



Directional correlation coefficient measures for Pythagorean fuzzy sets: their applications to medical diagnosis and cluster analysis

Mingwei Lin^{1,2} · Chao Huang¹ · Riqing Chen² · Hamido Fujita^{3,4} · Xing Wang¹

Received: 4 July 2020 / Accepted: 16 December 2020 / Published online: 20 January 2021
© The Author(s) 2021

Abstract

Compared to the intuitionistic fuzzy sets, the Pythagorean fuzzy sets (PFSs) can provide the decision makers with more freedom to express their evaluation information. There exist some research results on the correlation coefficient between PFSs, but sometimes they fail to deal with the problems of disease diagnosis and cluster analysis. To tackle the drawbacks of the existing correlation coefficients between PFSs, some novel directional correlation coefficients are put forward to compute the relationship between two PFSs by taking four parameters of the PFSs into consideration, which are the membership degree, non-membership degree, strength of commitment, and direction of commitment. Afterwards, two practical examples are given to show the application of the proposed directional correlation coefficient in the disease diagnosis, and the application of the proposed weighted directional correlation coefficient in the cluster analysis. Finally, they are compared with the previous correlation coefficients that have been developed for PFSs.

Keywords Pythagorean fuzzy set · Correlation coefficient · Medical diagnosis · Cluster analysis · Information energy

Introduction

The decision making refers to the human activities that rank the alternatives and select the optimal ones [26, 38, 39, 55]. It is extremely common in our daily life and national

industrial development [10, 13, 23]. For example, there exist some startup companies that need to look for the venture capital. The investor evaluates the startup companies and then selects the optimal alternative for investing. In the early stage, the decision-making problems are very simple, so human beings can use crisp numbers to express their opinions [5, 6, 30, 31]. As the human activities develop quickly, it is difficult for the human beings to describe the vague and imprecise information. The knowledge system based on the fuzzy set theory was put forward by Zadeh [62] to model the fuzzy information. However, the fuzzy set is not composed of the non-membership degree. To extend the modeling capability of the fuzzy set, the concept of intuitionistic fuzzy sets (IFSs) was proposed in [4]. The IFSs utilize a pair of values from the unit interval to describe the membership degree (MD) and non-membership degree (NMD) simultaneously. The sum of the MD and NMD in each IFS is less than or equal to 1.

To provide the decision makers (DMs) with more freedom to express the MDs and NMDs [22], Yager [61] gave an extension of IFSs, which is called the Pythagorean fuzzy sets (PFSs) and is also characterized by the MD and NMD. However, the square sum of the MD and NMD in each PFS is less than or equal to 1. As shown in Fig. 1, the reachable space of each Pythagorean fuzzy number (PFN) is larger

✉ Riqing Chen
riqing.chen@fafu.edu.cn

✉ Xing Wang
wangxing@fjnu.edu.cn

Mingwei Lin
linmwcs@163.com

Chao Huang
heminmax@163.com

Hamido Fujita
hfujita-799@acm.org

- 1 College of Mathematics and Informatics, Fujian Normal University, Fuzhou 350117, Fujian, China
- 2 Digital Fujian Institute of Big Data for Agriculture and Forestry, Fujian Agriculture and Forestry University, Fuzhou 350002, China
- 3 Faculty of Information Technology, Ho Chi Minh City University of Technology, Ho Chi Minh City, Vietnam
- 4 Faculty of Software and Information Science, Iwate Prefectural University, Iwate 020-0693, Japan

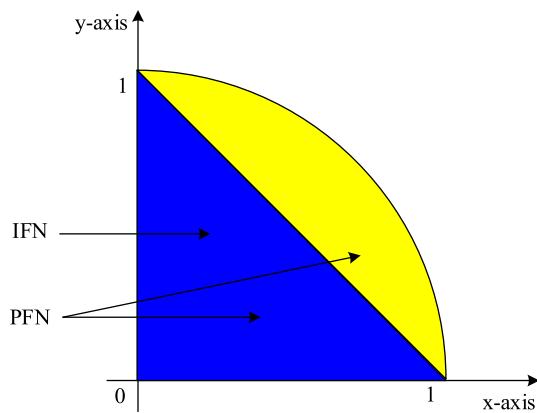


Fig. 1 The reachable spaces of IFNs and PFNs

than that of each intuitionistic fuzzy number (IFN). It indicates that the PFSs provide the DMs with more freedom to express their information than the IFNs from the geometric thinking. Hence, each IFN can also be an PFN, while the reverse is not true. For example, a pair of values, denoted by $(0.6, 0.8)$, is not an IFN since $0.6 + 0.8 > 1$. However, it is an PFN since $0.6^2 + 0.8^2 = 1$. Because of its stronger modeling capability, the PFSs have been paid much attention [1, 12, 44, 52–54, 59, 64]. For example, Garg [18] combined the PFSs with the linguistic term sets [32] to put forward the concept of linguistic Pythagorean fuzzy sets. The novel distance measures were devised for PFSs in [41]. Akram et al. [2] utilized a revised closeness index and the TOPSIS method to handle the group decision-making problems with Pythagorean fuzzy sets. Yager [60] presented a generalized form of IFNs and PFNs, called the q -rung orthopair fuzzy sets. Xiao and Ding [56] put forward the concept of divergence measure for PFSs and applied it into medical diagnosis. Some similarity measures have been put forward for PFSs and used to process multicriteria decision-making problems [51] [54]. Lots of aggregation operators have been developed for fusing PFSs [14, 16, 19, 20].

The correlation coefficient, which is an important concept from the statistics, is utilized to measure the relationship between two random variables and is widely applied in the statistical analysis [3] [46]. It has been extended to compute the relationship between fuzzy information [35] and has wide applications in the medical diagnosis [43], pattern recognition [42], cluster analysis [47], and decision-making [33]. Chiang et al. [9] introduced a method from the mathematical statistics to compute the correlation coefficient between fuzzy data. In [40], a correlation coefficient between two IFNs was put forward. In [28], a novel correlation coefficient measure was developed for hesitant fuzzy sets, whose value is in the interval $[-1, 1]$. In [11], the correlation and correlation coefficient between q -rung orthopair fuzzy sets were developed by Du and their properties were

discussed. Although there exist many research results reporting the correlation coefficient for different kinds of fuzzy information, they cannot be utilized to calculate the relationship between PFSs since the PFSs show a unique information representation form. To fill in this gap for PFSs, Garg [21] firstly developed a correlation coefficient and its weighted version to compute the relationship between two PFSs considering their MD, NMD, and hesitation degree (HD). Chen [7] devised some Pearson-like correlation coefficients for PFSs and used them to develop the Pythagorean fuzzy compromise approach. The existing studies have the drawback that the correlation coefficient value between two unequal PFSs may equal to 1. To address this issue, Singh et al. [45] put forward some correlation coefficients for PFSs. To consider the inverse correlation relation between PFSs, Thao [49] proposed a novel correlation coefficient for PFSs, the value of which is in the interval $[-1, 1]$. However, all of these existing correlation coefficients for PFSs still have some drawbacks as follows:

1. For the existing studies in [7, 49], there exists the case that the denominator equals to zero during the computing process of correlation coefficient value between two PFSs. It is meaningless in mathematics. Hence, they may be invalid for some cases.
2. As mentioned by Yager [61], the strength of commitment (SoC) and direction of commitment (DoC) are a pair of important parameters that can be used to determine a unique PFN. Moreover, Li et al. [27] and Zeng et al. [63] reported that SoC and DoC are important for developing the distance measure between two PFSs. However, all the existing studies only consider the MDs, NMDs, and HDs of PFSs when calculating the correlation coefficient between PFSs, but ignore the impact of SoC and DoC.

To solve the above drawbacks, we put forward some directional correlation coefficients to measure the relationship between PFSs by considering the MDs, NMDs, SoCs, and DoCs of PFNs. Our contributions can be summarized as follows:

1. We analyze the drawbacks of the existing correlation coefficients using some counterexamples. To overcome these drawbacks, we propose a novel directional correlation coefficient and also its weighted version for PFSs. The properties of the novel directional correlation coefficient and its weighted version are also discussed.
2. The proposed directional correlation coefficients are applied to deal with the problems of medical diagnosis and cluster analysis via two case studies and also a series of comparative analyses are performed to verify

the superiority of our proposed directional correlation coefficients.

The remainder of this paper is organized as follows. Section “Preliminaries” provides some basic knowledge of IFSs and PFSs. The drawbacks of the existing correlation coefficients of PFSs are analyzed and then some novel directional correlation coefficients are studied in Sect. “Directional correlation coefficients for PFSs”. Section “Case study and comparative analysis” provides two case studies to show the applications of the proposed directional correlation coefficients in the medical diagnosis and cluster analysis. Some valuable conclusions are drawn in Sect. “Conclusions”.

Preliminaries

In this section, the mathematical forms of the intuitionistic fuzzy sets and Pythagorean fuzzy sets are presented.

Intuitionistic fuzzy sets

The concept of intuitionistic fuzzy sets (IFSs) was proposed by Atanassov [4], which is an extension of fuzzy sets. The IFSs consist of the membership function and non-membership function of each element in the universe of discourse. The membership function and non-membership function return the membership degree (MD) and non-membership degree (NMD) of each element belonging to an IFS. The values of MD and NMD in the IFSs are in the unit interval [0, 1] and their sum is also in the unit interval [0, 1]. Hence, the mathematical form of an IFS can be expressed as.

Definition 1 [4]. Let $X = \{x_1, x_2, \dots, x_n\}$ denote a universe of discourse, then an IFS I on X can be defined as.

$$I = \left\{ (x, \mu_I(x), \nu_I(x)) \mid x \in X \right\} \tag{1}$$

where $\mu_I(x)$ and $\nu_I(x)$ are the membership function and non-membership function of each element x , respectively. The return values from the membership function and non-membership function form a pair of MD and NMD, which are denoted as (μ_I, ν_I) . Each pair (μ_I, ν_I) is called an IFN [25] and it satisfies that $0 \leq \mu_I \leq 1$, $0 \leq \nu_I \leq 1$, and $0 \leq \mu_I + \nu_I \leq 1$. The hesitation degree of an IFN, denoted as π_I , is computed as $1 - \mu_I - \nu_I$. Hence, the reachable space of each IFN is a triangle, which is enclosed by x axis, y axis, and the line $\mu_I + \nu_I = 1$ as shown in Fig. 1.

Pythagorean fuzzy sets

In some situations, the DMs may provide pairs of MD and NMD, which do not satisfy the restrictive condition of IFSs. For example, a pair of MD and NMD, (0.5, 0.6), is given by the DMs. The sum of MD and NMD is larger than 1, but their square sum, $0.5^2 + 0.6^2$, is smaller than 1. If the DMs are required to change their information to satisfy the restrictive condition of IFSs, then the information will be not original and deviate from the DM’s real opinion. This case will result in unreasonable decision results. To overcome this drawback, Yager [61] put forward an extension of IFSs, called Pythagorean fuzzy sets (PFSs), to model the information, where the square sum of MD and NMD is less than or equal to 1. The mathematical form of PFSs is expressed as follows.

Definition 2 [61]. Let $X = \{x_1, x_2, \dots, x_n\}$ denote a universe of discourse, then a PFS P on X can be defined as

$$P = \left\{ (x, \mu_P(x), \nu_P(x)) \mid x \in X \right\} \tag{2}$$

where $\mu_P(x)$ and $\nu_P(x)$ are the membership function and non-membership function of each element x , respectively. The return values from the membership function and non-membership function form a pair of MD and NMD, which are denoted as (μ_P, ν_P) . Each pair (μ_P, ν_P) is called a PFN [17] and it satisfies that $0 \leq \mu_P \leq 1$, $0 \leq \nu_P \leq 1$, and $0 \leq \mu_P^2 + \nu_P^2 \leq 1$. The hesitation degree of an PFN, which is denoted as π_P , is computed as $\sqrt{1 - \mu_P^2 - \nu_P^2}$. Thus, the reachable space of each PFN is a quarter circle, which is enclosed by x axis, y axis, and the curve $\mu_P^2 + \nu_P^2 = 1$. As shown in Fig. 1, the PFSs provide the DMs with more freedom to express their information than the IFSs.

Yager [61] put forward another way of expressing an PFN using a pair of parameters (r_P, d_P) . The parameter $r_P \in [0, 1]$ is named the strength of commitment (SoC) of an PFN and the parameter $d_P \in [0, 1]$ is named the direction of commitment (DoC). These two parameters are associated with the MD and NMD of an PFN. The relationships among the SoC, DoC, MD, and NMD are expressed as

$$\mu_P = r_P \cos \theta \tag{3}$$

$$\nu_P = r_P \sin \theta, \tag{4}$$

where $\theta = \left(1 - d_P\right) \frac{\pi}{2}$. The term θ denotes the intersection angle and $\theta \in \left[0, \frac{\pi}{2}\right]$.

The relationships among the SoC, DoC, MD, and NMD can be interpreted geometrically as shown in Fig. 2. When the value of θ is fixed, the values of MD and NMD increase

as the value of SoC increases. When the value of r_p is fixed, as the value of θ increases, the value of MD decreases, while the value of NMD increases. Hence, the parameters SoC and DoC are very important to determine an PFN. From Fig. 2, it can be known that the value of θ determines the direction of r_p . Moreover, $d_p = 1 - \frac{2\theta}{\pi}$. Hence d_p is called the direction of r_p .

Directional correlation coefficients for PFSSs

In this section, the drawbacks of the existing correlation coefficients of PFSSs are analyzed using the counterexamples and then some novel directional correlation coefficients are devised for PFSSs. Finally, their properties are discussed.

drawbacks of these four research studies are analyzed as follows.

1. Drawback of the correlation coefficient proposed by Garg [21].

Motivated by the correlation coefficient equations between IFSs given in [58], Garg [21] proposed the correlation coefficient between PFSSs as follows.

Definition 3 [21]. Let $X = \{x_1, x_2, \dots, x_n\}$ be an universe of discourse, $P_1 = \left\{ (x, \mu_{P_1}(x), \nu_{P_1}(x)) \mid x \in X \right\}$ and $P_2 = \left\{ (x, \mu_{P_2}(x), \nu_{P_2}(x)) \mid x \in X \right\}$ be two PFSSs on X , then the correlation coefficient between these two PFSSs is computed as.

$$K_1(P_1, P_2) = \frac{\sum_{i=1}^n \left(\mu_{P_1}^2(x_i) \mu_{P_2}^2(x_i) + \nu_{P_1}^2(x_i) \nu_{P_2}^2(x_i) + \pi_{P_1}^2(x_i) \pi_{P_2}^2(x_i) \right)}{\sqrt{\sum_{i=1}^n \left(\mu_{P_1}^4(x_i) + \nu_{P_1}^4(x_i) + \pi_{P_1}^4(x_i) \right)} \sqrt{\sum_{i=1}^n \left(\mu_{P_2}^4(x_i) + \nu_{P_2}^4(x_i) + \pi_{P_2}^4(x_i) \right)}} \quad (5)$$

Drawbacks of the existing correlation coefficients for PFSSs

To the best of our knowledge, there are only four research studies reporting the correlation coefficients of PFSSs, which are presented in the literatures [7, 21, 45, 49]. Here, the

where the terms $\pi_{P_1}(x_i)$ and $\pi_{P_2}(x_i)$ denote the HDs of the element x_i belonging to the PFSSs P_1 and P_2 , respectively.

From Definition 3, it can be noted that the correlation coefficient presented in [21] only considers the MD, NMD, and HD of each PFSS, but ignores the SoC and DoC, which are also important parameters when determining the values of PFSSs. Hence, this correlation coefficient given in [21] may result in unreasonable results in some situations. In the following part, a counterexample is provided to show its drawback.

Example 1 Let X be an universe of discourse, $P_1 = \left(\frac{2}{3}, \frac{2}{3} \right)$, $P_2 = \left(\frac{1}{3}, \frac{2}{3} \right)$, and $P = \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}} \right)$ be three PFSSs on X . If Eq. (5) is used to compute the correlation coefficient between two PFSSs, then we can get

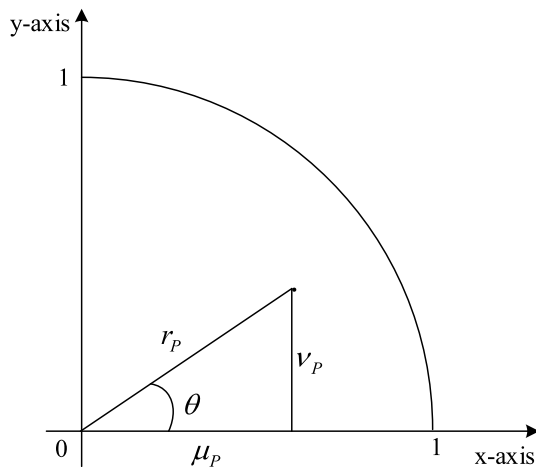


Fig. 2 The geometric interpretation of the relationships among the SoC, DoC, MD, and NMD

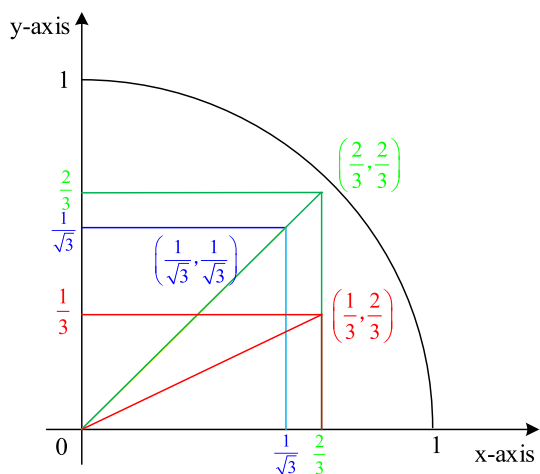


Fig. 3 The geometric interpretation of the relationships among PFSs

$P_2 = \left\{ (x, \mu_{P_2}(x), \nu_{P_2}(x)) \mid x \in X \right\}$ be two PFSs on X , then the correlation coefficient between these two PFSs is computed as

$$K_2(P_1, P_2) = \frac{1}{3} (k_\mu(P_1, P_2) + k_\nu(P_1, P_2) + k_\pi(P_1, P_2)) \tag{6}$$

with

$$k_\mu(P_1, P_2) = \frac{\sum_{i=1}^n ((\mu_{P_1}(x_i))^2 - \bar{\mu}_{P_1}^2)((\mu_{P_2}(x_i))^2 - \bar{\mu}_{P_2}^2)}{\sqrt{\sum_{i=1}^n ((\mu_{P_1}(x_i))^2 - \bar{\mu}_{P_1}^2)^2} \cdot \sqrt{\sum_{i=1}^n ((\mu_{P_2}(x_i))^2 - \bar{\mu}_{P_2}^2)^2}} \tag{7}$$

$$k_\nu(P_1, P_2) = \frac{\sum_{i=1}^n ((\nu_{P_1}(x_i))^2 - \bar{\nu}_{P_1}^2)((\nu_{P_2}(x_i))^2 - \bar{\nu}_{P_2}^2)}{\sqrt{\sum_{i=1}^n ((\nu_{P_1}(x_i))^2 - \bar{\nu}_{P_1}^2)^2} \cdot \sqrt{\sum_{i=1}^n ((\nu_{P_2}(x_i))^2 - \bar{\nu}_{P_2}^2)^2}} \tag{8}$$

$$K_1(P_1, P) = \frac{\left(\frac{2}{3}\right)^2 \left(\frac{1}{\sqrt{3}}\right)^2 + \left(\frac{2}{3}\right)^2 \left(\frac{1}{\sqrt{3}}\right)^2 + \left(\frac{1}{3}\right)^2 \left(\frac{1}{\sqrt{3}}\right)^2}{\sqrt{\left(\left(\frac{2}{3}\right)^4 + \left(\frac{2}{3}\right)^4 + \left(\frac{1}{3}\right)^4\right)} \sqrt{\left(\left(\frac{1}{\sqrt{3}}\right)^4 + \left(\frac{1}{\sqrt{3}}\right)^4 + \left(\frac{1}{\sqrt{3}}\right)^4\right)}} = 0.9045$$

$$K_1(P_2, P) = \frac{\left(\frac{1}{3}\right)^2 \left(\frac{1}{\sqrt{3}}\right)^2 + \left(\frac{2}{3}\right)^2 \left(\frac{1}{\sqrt{3}}\right)^2 + \left(\frac{2}{3}\right)^2 \left(\frac{1}{\sqrt{3}}\right)^2}{\sqrt{\left(\left(\frac{1}{3}\right)^4 + \left(\frac{2}{3}\right)^4 + \left(\frac{2}{3}\right)^4\right)} \sqrt{\left(\left(\frac{1}{\sqrt{3}}\right)^4 + \left(\frac{1}{\sqrt{3}}\right)^4 + \left(\frac{1}{\sqrt{3}}\right)^4\right)}} = 0.9045$$

From the above computing result, it can be seen that $K_1(P_1, P) = K_1(P_2, P)$. However, the actual situation is that $K_1(P_1, P) \neq K_1(P_2, P)$ as shown in Fig. 3. Hence, Eq. (5) cannot handle this example well.

2. Drawback of the correlation coefficient proposed by Chen [7]

Inspired by the correlation coefficient between IFSs given in [48], Chen [7] proposed the correlation coefficient between IFSs as follows.

Definition 4 [7]. Let $X = \{x_1, x_2, \dots, x_n\}$ be an universe of discourse, $P_1 = \left\{ (x, \mu_{P_1}(x), \nu_{P_1}(x)) \mid x \in X \right\}$ and

$$k_\pi(P_1, P_2) = \frac{\sum_{i=1}^n ((\pi_{P_1}(x_i))^2 - \bar{\pi}_{P_1}^2)((\pi_{P_2}(x_i))^2 - \bar{\pi}_{P_2}^2)}{\sqrt{\sum_{i=1}^n ((\pi_{P_1}(x_i))^2 - \bar{\pi}_{P_1}^2)^2} \cdot \sqrt{\sum_{i=1}^n ((\pi_{P_2}(x_i))^2 - \bar{\pi}_{P_2}^2)^2}} \tag{9}$$

where $\bar{\mu}_{P_1} = \frac{1}{n} \sum_{i=1}^n \mu_{P_1}(x_i)$, $\bar{\mu}_{P_2} = \frac{1}{n} \sum_{i=1}^n \mu_{P_2}(x_i)$, $\bar{\nu}_{P_1} = \frac{1}{n} \sum_{i=1}^n \nu_{P_1}(x_i)$, $\bar{\nu}_{P_2} = \frac{1}{n} \sum_{i=1}^n \nu_{P_2}(x_i)$, $\bar{\pi}_{P_1} = \frac{1}{n} \sum_{i=1}^n \pi_{P_1}(x_i)$ and $\bar{\pi}_{P_2} = \frac{1}{n} \sum_{i=1}^n \pi_{P_2}(x_i)$. $\pi_{P_1}(x_i)$ and $\pi_{P_2}(x_i)$ are the HDs of the element x_i belonging to the PFSs P_1 and P_2 .

Example 2 Let X be an universe of discourse, $P_1 = \left\{ \left(\frac{2}{3}, \frac{2}{3}\right), \left(\frac{1}{3}, \frac{2}{3}\right) \right\}$ and $P_2 = \left\{ \left(\frac{2}{3}, \frac{1}{3}\right), \left(\frac{2}{3}, \frac{2}{3}\right) \right\}$ be two PFSs on X , then we have

$$k_\mu(P_1, P_2) = \frac{\left(\left(\frac{2}{3}\right)^2 - \left(\frac{1}{2}\right)^2\right)\left(\left(\frac{2}{3}\right)^2 - \left(\frac{2}{3}\right)^2\right) + \left(\left(\frac{1}{3}\right)^2 - \left(\frac{1}{2}\right)^2\right)\left(\left(\frac{2}{3}\right)^2 - \left(\frac{2}{3}\right)^2\right)}{\sqrt{\left(\left(\frac{2}{3}\right)^2 - \left(\frac{1}{2}\right)^2\right)^2 + \left(\left(\frac{1}{3}\right)^2 - \left(\frac{1}{2}\right)^2\right)^2} \cdot \sqrt{\left(\left(\frac{2}{3}\right)^2 - \left(\frac{2}{3}\right)^2\right)^2 + \left(\left(\frac{2}{3}\right)^2 - \left(\frac{2}{3}\right)^2\right)^2}} = \frac{0}{0}$$

From the result, it can be known that Eq. (6) is invalid for Example 2 since the denominator cannot be zero in mathematics.

3. Drawback of the correlation coefficient proposed by Singh et al. [45].

For most of the existing correlation coefficients, the correlation coefficient value between two unequal PFSs is 1, which is unreasonable and ineffective. To overcome this drawback, Singh et al. [45] put forward a novel correlation coefficient as follows:

Definition 5 [45]. Let $X = \{x_1, x_2, \dots, x_n\}$ be an universe of discourse, $P_1 = \{ (x, \mu_{P_1}(x), \nu_{P_1}(x)) \mid x \in X \}$ and $P_2 = \{ (x, \mu_{P_2}(x), \nu_{P_2}(x)) \mid x \in X \}$ be two PFSs on X , then the correlation coefficient between these two PFSs is computed as.

$$K_3(P_1, P_2) = \frac{1}{2n} \sum_{i=1}^n [\alpha_i(1 - \Delta\mu_i) + \beta_i(1 - \Delta\nu_i)] \quad (10)$$

where

$$\begin{aligned} \alpha_i &= \frac{c - \Delta\mu_i - \Delta\mu_{\max}}{c - \Delta\mu_{\min} - \Delta\mu_{\max}}, \beta_i = \frac{c - \Delta\nu_i - \Delta\nu_{\max}}{c - \Delta\nu_{\min} - \Delta\nu_{\max}}, c > 2, \\ \Delta\mu_i &= \left| (\mu_{P_1}(x_i))^2 - (\mu_{P_2}(x_i))^2 \right|, \\ \Delta\nu_i &= \left| (\nu_{P_1}(x_i))^2 - (\nu_{P_2}(x_i))^2 \right|, \\ \Delta\mu_{\min} &= \min_i \left\{ \left| (\mu_{P_1}(x_i))^2 - (\mu_{P_2}(x_i))^2 \right| \right\}, \\ \Delta\mu_{\max} &= \max_i \left\{ \left| (\mu_{P_1}(x_i))^2 - (\mu_{P_2}(x_i))^2 \right| \right\}, \\ \Delta\nu_{\min} &= \min_i \left\{ \left| (\nu_{P_1}(x_i))^2 - (\nu_{P_2}(x_i))^2 \right| \right\}, \\ \Delta\nu_{\max} &= \max_i \left\{ \left| (\nu_{P_1}(x_i))^2 - (\nu_{P_2}(x_i))^2 \right| \right\}. \end{aligned}$$

Example 3 Let X be an universe of discourse, $P_1 = (u, u)$ and $P = (1, 0)$ be two PFSs on X . If Eq. (10) is used to compute the correlation coefficient between two PFSs, then we can get

$$K_3(P_1, P) = \frac{1}{2} [1 \times (1 - |u^2 - 1^2|) + 1 \times (1 - |u^2 - 0^2|)] = \frac{1}{2}.$$

From the above result, it can be seen that $K_3(P_1, P)$ always equals to $\frac{1}{2}$ when $\mu_{P_1} = \nu_{P_1} = u$. Hence, it is unreasonable and ineffective.

4. Drawback of the correlation coefficient proposed by Thao [49]

Based on the variance and covariance of PFSs, Thao [49] proposed the correlation coefficient between PFSs as follows.

Definition 6 [49]. Let $X = \{x_1, x_2, \dots, x_n\}$ be an universe of discourse, $P_1 = \{ (x, \mu_{P_1}(x), \nu_{P_1}(x)) \mid x \in X \}$ and $P_2 = \{ (x, \mu_{P_2}(x), \nu_{P_2}(x)) \mid x \in X \}$ be two PFSs on X , then the correlation coefficient between these two PFSs is computed as

$$K_4(P_1, P_2) = \frac{COV(P_1, P_2)}{\sqrt{D(P_1)D(P_2)}} \quad (11)$$

where

$$\begin{aligned} COV(P_1, P_2) &= \frac{1}{n-1} \sum_{i=1}^n d_i(P_1)d_i(P_2), D(P_1) \\ &= \frac{1}{n-1} \sum_{i=1}^n d_i^2(P_1), D(P_2) \\ &= \frac{1}{n-1} \sum_{i=1}^n d_i^2(P_2), \end{aligned}$$

$$\begin{aligned} d_i(P_1) &= \left((\mu_{P_1}(x_i))^2 - \left(\frac{1}{n} \sum_{i=1}^n \mu_{P_1}(x_i) \right)^2 \right) \\ &\quad - \left((\nu_{P_1}(x_i))^2 - \left(\frac{1}{n} \sum_{i=1}^n \nu_{P_1}(x_i) \right)^2 \right), \end{aligned}$$

$$\begin{aligned} d_i(P_2) &= \left((\mu_{P_2}(x_i))^2 - \left(\frac{1}{n} \sum_{i=1}^n \mu_{P_2}(x_i) \right)^2 \right) \\ &\quad - \left((\nu_{P_2}(x_i))^2 - \left(\frac{1}{n} \sum_{i=1}^n \nu_{P_2}(x_i) \right)^2 \right). \end{aligned}$$

Example 4 Let X be an universe of discourse, $P_1 = \left(\frac{2}{3}, \frac{2}{3}\right)$ and $P_2 = \left(\frac{2}{5}, \frac{1}{5}\right)$ be two PFSs on X , then we have

$$K_4(P_1, P_2) = \frac{COV(P_1, P_2)}{\sqrt{D(P_1)D(P_2)}} = \frac{\frac{1}{1-1} \times 0 \times 0}{\sqrt{\frac{1}{1-1} \times 0^2 \times 0^2}}$$

From the above result, it is seen that Eq. (11) is invalid for Example 4 since the denominator cannot be zero in mathematics.

Directional correlation coefficients and their properties

To solve the defects of the existing correlation coefficients of PFSs, we propose the information energy, correlation, and novel directional correlation coefficients for PFSs by considering their MDs, NMDs, SoCs, and DoCs in this section.

Let $X = \{x_1, x_2, \dots, x_n\}$ be an universe of discourse, $P = \{ (x, \mu_P(x), \nu_P(x)) \mid x \in X \}$ denote an PFS on X , then the information energy of the PFS P is computed as

$$T(P) = \sum_{i=1}^n \left((\mu_P(x_i))^4 + (v_P(x_i))^4 + (r_P(x_i))^4 + (d_P(x_i))^4 \right) \tag{12}$$

where $r_P(x_i)$ and $d_P(x_i)$ denote the strength of commitment and the direction of commitment that are associated with the membership degree and non-membership degree of the element x_i .

Definition 7 Let $X = \{x_1, x_2, \dots, x_n\}$ denote an universe of discourse, $P_1 = \left\{ (x, \mu_{P_1}(x), \nu_{P_1}(x)) \mid x \in X \right\}$ and

Based on the information energy and correlation, the correlation coefficient between any two PFSs is defined as follows.

Definition 8 Let $X = \{x_1, x_2, \dots, x_n\}$ denote an universe of discourse, $P_1 = \left\{ (x, \mu_{P_1}(x), \nu_{P_1}(x)) \mid x \in X \right\}$ and $P_2 = \left\{ (x, \mu_{P_2}(x), \nu_{P_2}(x)) \mid x \in X \right\}$ be two PFSs on X , then the correlation coefficient between the PFSs P_1 and P_2 is defined as

$$\begin{aligned} \partial_1(P_1, P_2) &= \frac{C(P_1, P_2)}{[T(P_1)T(P_2)]^{1/2}} \\ &= \frac{\sum_{i=1}^n \left((\mu_{P_1}(x_i))^2(\mu_{P_2}(x_i))^2 + (v_{P_1}(x_i))^2(v_{P_2}(x_i))^2 + (r_{P_1}(x_i))^2(r_{P_2}(x_i))^2 + (d_{P_1}(x_i))^2(d_{P_2}(x_i))^2 \right)}{\sqrt{\sum_{i=1}^n \left((\mu_{P_1}(x_i))^4 + (v_{P_1}(x_i))^4 + (r_{P_1}(x_i))^4 + (d_{P_1}(x_i))^4 \right)} \sqrt{\sum_{i=1}^n \left((\mu_{P_2}(x_i))^4 + (v_{P_2}(x_i))^4 + (r_{P_2}(x_i))^4 + (d_{P_2}(x_i))^4 \right)}} \end{aligned} \tag{14}$$

$P_2 = \left\{ (x, \mu_{P_2}(x), \nu_{P_2}(x)) \mid x \in X \right\}$ be two PFSs on X , then the correlation between these two PFSs P_1 and P_2 is defined as

$$\begin{aligned} C(P_1, P_2) &= \sum_{i=1}^n \left((\mu_{P_1}(x_i))^2(\mu_{P_2}(x_i))^2 + (v_{P_1}(x_i))^2(v_{P_2}(x_i))^2 \right. \\ &\quad \left. + (r_{P_1}(x_i))^2(r_{P_2}(x_i))^2 + (d_{P_1}(x_i))^2(d_{P_2}(x_i))^2 \right) \end{aligned} \tag{13}$$

Apparently, the correlation between the PFSs P_1 and P_2 satisfies the following properties:

1. $C(P_1, P_1) = T(P_1)$;
2. $C(P_1, P_2) = C(P_2, P_1)$.

Theorem 1 Let $X = \{x_1, x_2, \dots, x_n\}$ denote an universe of discourse, $P_1 = \left\{ (x, \mu_{P_1}(x), \nu_{P_1}(x)) \mid x \in X \right\}$ and $P_2 = \left\{ (x, \mu_{P_2}(x), \nu_{P_2}(x)) \mid x \in X \right\}$ be two PFSs on X , then the correlation coefficient between the PFSs P_1 and P_2 has the following properties:

1. $\partial_1(P_1, P_2) = \partial_1(P_2, P_1)$;
2. $0 \leq \partial_1(P_1, P_2) \leq 1$;
3. $\partial_1(P_1, P_2) = 1$ if $P_1 = P_2$.

Proof (1) It is straightforward.

(2) From Definition 8, it is straightforward that $0 \leq \partial_1(P_1, P_2)$. We only need to prove that $\partial_1(P_1, P_2) \leq 1$.

$$\begin{aligned} C(P_1, P_2) &= \sum_{i=1}^n \left((\mu_{P_1}(x_i))^2(\mu_{P_2}(x_i))^2 + (v_{P_1}(x_i))^2(v_{P_2}(x_i))^2 + (r_{P_1}(x_i))^2(r_{P_2}(x_i))^2 + (d_{P_1}(x_i))^2(d_{P_2}(x_i))^2 \right) \\ &= \left((\mu_{P_1}(x_1))^2(\mu_{P_2}(x_1))^2 + (v_{P_1}(x_1))^2(v_{P_2}(x_1))^2 + (r_{P_1}(x_1))^2(r_{P_2}(x_1))^2 + (d_{P_1}(x_1))^2(d_{P_2}(x_1))^2 \right) + \\ &\quad \left((\mu_{P_1}(x_2))^2(\mu_{P_2}(x_2))^2 + (v_{P_1}(x_2))^2(v_{P_2}(x_2))^2 + (r_{P_1}(x_2))^2(r_{P_2}(x_2))^2 + (d_{P_1}(x_2))^2(d_{P_2}(x_2))^2 \right) + \dots + \\ &\quad \left((\mu_{P_1}(x_n))^2(\mu_{P_2}(x_n))^2 + (v_{P_1}(x_n))^2(v_{P_2}(x_n))^2 + (r_{P_1}(x_n))^2(r_{P_2}(x_n))^2 + (d_{P_1}(x_n))^2(d_{P_2}(x_n))^2 \right) \end{aligned} \tag{15}$$

According to the Cauchy–Schwarz inequality, we have

$$\begin{aligned}
 (C(P_1, P_2))^2 &\leq \left[(\mu_{P_1}(x_1))^4 + (v_{P_1}(x_1))^4 + (r_{P_1}(x_1))^4 + (d_{P_1}(x_1))^4 + (\mu_{P_1}(x_2))^4 + (v_{P_1}(x_2))^4 + (r_{P_1}(x_2))^4 + (d_{P_1}(x_2))^4 + \dots \right. \\
 &\quad \left. + (\mu_{P_1}(x_n))^4 + (v_{P_1}(x_n))^4 + (r_{P_1}(x_n))^4 + (d_{P_1}(x_n))^4 \right] \times \left[(\mu_{P_2}(x_1))^4 + (v_{P_2}(x_1))^4 + (r_{P_2}(x_1))^4 + (d_{P_2}(x_1))^4 \right. \\
 &\quad \left. + (\mu_{P_2}(x_2))^4 + (v_{P_2}(x_2))^4 + (r_{P_2}(x_2))^4 + (d_{P_2}(x_2))^4 + \dots + (\mu_{P_2}(x_n))^4 + (v_{P_2}(x_n))^4 + (r_{P_2}(x_n))^4 + (d_{P_2}(x_n))^4 \right] \tag{16} \\
 &= \sum_{i=1}^n \left((\mu_{P_1}(x_i))^4 + (v_{P_1}(x_i))^4 + (r_{P_1}(x_i))^4 + (d_{P_1}(x_i))^4 \right) \times \sum_{i=1}^n \left((\mu_{P_2}(x_i))^4 + (v_{P_2}(x_i))^4 + (r_{P_2}(x_i))^4 + (d_{P_2}(x_i))^4 \right) \\
 &= T(P_1) \cdot T(P_2)
 \end{aligned}$$

Hence, $\partial_1(P_1, P_2) \leq 1$, which completes the proof.

(3) For each $x_i \in X$, if $P_1 = P_2$, then $\mu_{P_1}(x_i) = \mu_{P_2}(x_i)$, $v_{P_1}(x_i) = v_{P_2}(x_i)$, $r_{P_1}(x_i) = r_{P_2}(x_i)$, $d_{P_1}(x_i) = d_{P_2}(x_i)$. Hence, $\partial_1(P_1, P_2) = 1$.

Example 5 Recently, the oil leakage event of Benz cars happens in China and has become the hot topic on the Internet. The reporter intends to investigate the correlation coefficient between the automobile after-sale service and the customer loyalty. The automobile after-sale service contains two factors: maintenance cost (x_1) and service attitude (x_2). The customer loyalty involves two aspects: customer satisfaction (y_1) and trust (y_2). The PFSs are utilized to express

Step 2 Use Eq. (12) to compute the information energy of the PFS P_2 as

$$\begin{aligned}
 T(P_2) &= \sum_{i=1}^n \left((\mu_{P_2}(x_i))^4 + (v_{P_2}(x_i))^4 + (r_{P_2}(x_i))^4 + (d_{P_2}(x_i))^4 \right) \\
 &= \left(0.7^4 + 0.6^4 + (\sqrt{0.85})^4 + 0.5489^4 \right) \\
 &\quad + \left(0.6^4 + 0.4^4 + (\sqrt{0.52})^4 + 0.6257^4 \right) \\
 &= 1.1830 + 0.5789 \\
 &= 1.7619
 \end{aligned}$$

Step 3 Use Eq. (13) to compute the correlation between the PFSs P_1 and P_2 as

$$\begin{aligned}
 C(P_1, P_2) &= \sum_{i=1}^n \left((\mu_{P_1}(x_i))^2 (\mu_{P_2}(x_i))^2 + (v_{P_1}(x_i))^2 (v_{P_2}(x_i))^2 + (r_{P_1}(x_i))^2 (r_{P_2}(x_i))^2 + (d_{P_1}(x_i))^2 (d_{P_2}(x_i))^2 \right) \\
 &= \left(0.6^2 * 0.7^2 + 0.5^2 * 0.6^2 + (\sqrt{0.61})^2 * (\sqrt{0.85})^2 + 0.5577^2 * 0.7422^2 \right) \\
 &\quad + \left(0.7^2 * 0.6^2 + 0.3^2 * 0.4^2 + (\sqrt{0.58})^2 * (\sqrt{0.52})^2 + 0.5489^2 * 0.6257^2 \right) \\
 &= 0.9562 + 0.6104 \\
 &= 1.5666
 \end{aligned}$$

the information of the respondents regarding the factors of the automobile after-sale service (P_1) and customer loyalty (P_2). The decision-making information collected from the questionnaire is expressed as

$$\begin{aligned}
 P_1 &= \{ (x_1, 0.6, 0.5), (x_2, 0.7, 0.3) \} \\
 P_2 &= \{ (y_1, 0.7, 0.6), (y_2, 0.6, 0.4) \}.
 \end{aligned}$$

Step 1 Use Eq. (12) to compute the information energy of the PFS P_1 as

$$\begin{aligned}
 T(P_1) &= \sum_{i=1}^n \left((\mu_{P_1}(x_i))^4 + (v_{P_1}(x_i))^4 + (r_{P_1}(x_i))^4 + (d_{P_1}(x_i))^4 \right) \\
 &= \left(0.6^4 + 0.5^4 + (\sqrt{0.61})^4 + 0.5577^4 \right) + \left(0.7^4 + 0.3^4 + (\sqrt{0.58})^4 + 0.7422^4 \right) \\
 &= 0.6609 + 0.8880 \\
 &= 1.5489
 \end{aligned}$$

Step 4 Use Eq. (14) to compute the directional correlation coefficient between the PFSs P_1 and P_2 as

$$\partial_1(P_1, P_2) = \frac{C(P_1, P_2)}{[T(P_1)T(P_2)]^{1/2}} = \frac{1.5666}{\sqrt{1.5489}\sqrt{1.7619}} = 0.9483$$

In the multi-attribute decision-making problems [11], the attributes usually own different weight values. Nevertheless, the proposed directional correlation coefficient formula in

Table 1 The medical reference data set containing the symptom values of the diseases

	Cough	Headache	Temperature	Stomach pain	Chest pain
Viral fever	(0.3, 0.4)	(0.7, 0.1)	(0.6, 0.2)	(0.1, 0.7)	(0.5, 0.2)
Malaria	(0.5, 0.1)	(0.3, 0.2)	(0.7, 0.1)	(0.6, 0.2)	(0.5, 0.1)
Typhoid	(0.6, 0.2)	(0.9, 0.1)	(0.4, 0.3)	(0.5, 0.2)	(0.1, 0.9)
Stomach problem	(0.4, 0.3)	(0.1, 0.7)	(0.2, 0.6)	(0.7, 0.1)	(0.2, 0.5)
Chest problem	(0.2, 0.6)	(0.1, 0.9)	(0.3, 0.4)	(0.2, 0.5)	(0.9, 0.1)

Table 2 Symptom values for four patients

	Cough	Headache	Temperature	Stomach pain	Chest pain
Luna	(0.4, 0.4)	(0.6, 0.6)	(0.5, 0.5)	(0.2, 0.2)	(0.5, 0.5)
Rose	(0.3, 0.3)	(0.6, 0.6)	(0.2, 0.2)	(0.3, 0.3)	(0.4, 0.4)
MingK	(0.4, 0.4)	(0.7, 0.7)	(0.3, 0.3)	(0.2, 0.2)	(0.7, 0.7)
Bob	(0.4, 0.4)	(0.6, 0.6)	(0.5, 0.5)	(0.6, 0.6)	(0.5, 0.5)

the form of (14) does not consider the weight information of the elements in the PFSs. To handle this case, the weighted directional correlation coefficient is devised. Suppose that the weight vector of the elements in the $X = \{x_1, x_2, \dots, x_n\}$ is denoted as $\omega = (\omega_1, \omega_2, \dots, \omega_n)$, which satisfies that $\omega_i \geq 0$ and $\sum_{i=1}^n \omega_i = 1$. Then the weighted information energy of the PFS P is computed as

$$T_\omega(P) = \sum_{i=1}^n \omega_i \left((\mu_P(x_i))^4 + (v_P(x_i))^4 + (r_P(x_i))^4 + (d_P(x_i))^4 \right) \tag{17}$$

and the weighted correlation between two PFSs P_1 and P_2 is calculated as

$$C_\omega(P_1, P_2) = \sum_{i=1}^n \omega_i \left((\mu_{P_1}(x_i))^2 (\mu_{P_2}(x_i))^2 + (v_{P_1}(x_i))^2 (v_{P_2}(x_i))^2 + (r_{P_1}(x_i))^2 (r_{P_2}(x_i))^2 + (d_{P_1}(x_i))^2 (d_{P_2}(x_i))^2 \right) \tag{18}$$

Hence, the weighted directional correlation coefficient between two PFSs P_1 and P_2 is computed as

$$\begin{aligned} \partial_2(P_1, P_2) &= \frac{C_\omega(P_1, P_2)}{[T_\omega(P_1)T_\omega(P_2)]^{1/2}} \\ &= \frac{\sum_{i=1}^n \omega_i \left((\mu_{P_1}(x_i))^2 (\mu_{P_2}(x_i))^2 + (v_{P_1}(x_i))^2 (v_{P_2}(x_i))^2 + (r_{P_1}(x_i))^2 (r_{P_2}(x_i))^2 + (d_{P_1}(x_i))^2 (d_{P_2}(x_i))^2 \right)}{\sqrt{\sum_{i=1}^n \omega_i \left((\mu_{P_1}(x_i))^4 + (v_{P_1}(x_i))^4 + (r_{P_1}(x_i))^4 + (d_{P_1}(x_i))^4 \right)} \sqrt{\sum_{i=1}^n \omega_i \left((\mu_{P_2}(x_i))^4 + (v_{P_2}(x_i))^4 + (r_{P_2}(x_i))^4 + (d_{P_2}(x_i))^4 \right)}} \end{aligned} \tag{19}$$

Table 3 Correlation coefficients between the symptom values of the patients and the symptom values of the diseases

	Viral fever	Malaria	Typhoid	Stomach problem	Chest problem
Luna	0.7353	0.6550	0.8138	0.7219	0.7796
Rose	0.7415	0.6225	0.8262	0.7056	0.7684
MingK	0.6307	0.5018	0.8198	0.6289	0.8127
Bob	0.7928	0.6807	0.7683	0.8027	0.7467

When $\omega = (1/n, 1/n, \dots, 1/n)$, Eq. (19) can reduce to Eq. (14).

Theorem 2 Let $X = \{x_1, x_2, \dots, x_n\}$ denote an universe of discourse and $\omega = (\omega_1, \omega_2, \dots, \omega_n)$ be the weight vector of the elements in X , $P_1 = \left\{ (x, \mu_{P_1}(x), v_{P_1}(x)) \mid x \in X \right\}$ and $P_2 = \left\{ (x, \mu_{P_2}(x), v_{P_2}(x)) \mid x \in X \right\}$ be two PFSs on X , then the weighted directional correlation coefficient between P_1 and P_2 shows the following properties:

1. $\partial_2(P_1, P_2) = \partial_2(P_2, P_1)$;
2. $0 \leq \partial_2(P_1, P_2) \leq 1$;
3. $\partial_2(P_1, P_2) = 1$ if $P_1 = P_2$.

Proof The process of proof is similar to that of Theorem 1, so it is omitted here.

Fig. 4 The visualization of correlation coefficients between the symptom values of the patients and the symptom values of the diseases

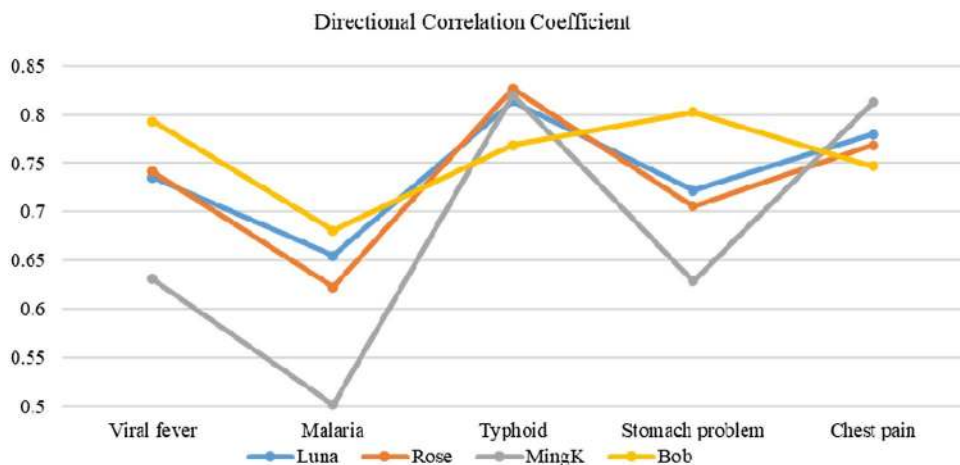


Table 4 Correlation coefficients between the symptom values of the patients and the symptom values of the diseases when Garg’s correlation coefficient [21] is used

	Viral fever	Malaria	Typhoid	Stomach problem	Chest problem
Luna	0.8567	0.8206	0.8124	0.8567	0.8124
Rose	0.8811	0.8183	0.7971	0.8811	0.7971
MingK	0.6988	0.5989	0.8552	0.6988	0.8552
Bob	0.8924	0.8158	0.7643	0.8924	0.7643

Table 5 Correlation coefficients between the symptom values of the patients and the symptom values of the diseases when Chen’s correlation coefficient [7] is used

	Viral fever	Malaria	Typhoid	Stomach problem	Chest problem
Luna	0.0473	− 0.3315	0.4485	0.0473	0.4485
Rose	0.2338	− 0.3737	0.5445	0.2338	0.5445
MingK	− 0.0031	− 0.4591	0.6289	− 0.0031	0.6289
Bob	0.5023	0.1994	0.1500	0.5023	0.1500

Table 6 Correlation coefficients between the symptom values of the patients and the symptom values of the diseases when the correlation coefficient of Singh et al. [45] is used

	Viral fever	Malaria	Typhoid	Stomach problem	Chest problem
Luna	0.8017	0.7612	0.7126	0.8017	0.7126
Rose	0.8013	0.7781	0.7380	0.8013	0.7380
MingK	0.7472	0.6857	0.7329	0.7472	0.7329
Bob	0.7996	0.7816	0.7009	0.7996	0.7009

Table 7 Correlation coefficients between the symptom values of the patients and the symptom values of the diseases when Thao’s correlation coefficient [49] is used

	Viral fever	Malaria	Typhoid	Stomach problem	Chest problem
Luna	NULL	NULL	NULL	NULL	NULL
Rose	NULL	NULL	NULL	NULL	NULL
MingK	NULL	NULL	NULL	NULL	NULL
Bob	NULL	NULL	NULL	NULL	NULL

Case study and comparative analysis

The correlation coefficient is an important indicator, which can be used in many practical applications. In this section, two practical cases are presented to demonstrate the applications of correlation coefficient in the medical diagnosis and cluster analysis.

Case study: medical diagnosis

Example 6 This example originated in [57]. Suppose that a traditional Chinese physician (TCP) performs the medical diagnosis for four patients

$D = \{d_1(Luna), d_2(Rose), d_3(MingK), d_4(Bob)\}$ according to their symptom values $C = \{c_1(Cough), c_2(Headache), c_3(Temperature), c_4(Stomach\ pain), c_5(Chest\ pain)\}$. The diseases $R = \{r_1(Viral\ fever), r_2(Malaria), r_3(Typhoid), r_4(Stomach\ problem), r_5(Chest\ problem)\}$ have different values for symptoms. Since the TCPs mainly use four methods: observation, olfaction, inquiry, and palpation when the diseases of the patients are diagnosed, it is very difficult for TCPs to provide crisp numbers as the values of symptoms. Considering the uncertainty and fuzziness of diagnosis processes, the Pythagorean fuzzy sets are used to represent the values of symptoms. According to the clinical experience accumulated by the veteran TCPs, a medical reference data set composed of the symptom values of the diseases is constructed as shown in Table 1. Through the above four methods, the TCP gives the symptom values of four patients, which are shown in Table 2. To conduct the medical diagnosis for these four patients, we use Eq. (14) to compute the correlation coefficients between their symptom values and the symptom values of the diseases in the medical reference data set, which are shown in Table 3. Then the patient $d_j(j = 1, 2, 3, 4)$ suffers from the disease $r_i(i = 1, 2, \dots, 5)$ that satisfies the condition that $\max(\partial_1(r_i, C_j))$, where C_j denotes the symptom values of the patient d_j .

Figure 4 depicts the visualization of correlation coefficients between the symptom values of the patients and the symptom values of the diseases.

As shown in Table 3 and Fig. 4, it can be seen that Luna, Rose, and MingK suffer from Typhoid, while Bob suffers from Stomach problem.

In order to compare our proposed directional correlation coefficient with Garg’s correlation coefficient [21], Chen’s correlation coefficient [7], the correlation coefficient of Singh et al. [45], and Thao’s correlation coefficient [49], Eqs. (5), (6), (10) and (11) are used to handle Example 6 and then the results are obtained as shown in Tables 4, 5, 6, 7.

As listed in Tables 4 and 6, the correlation coefficient value between the symptom values of Luna and the symptom values of Viral fever equals to the correlation coefficient value between the symptom values of Luna and the symptom values of Stomach problem. Hence, the disease of Luna cannot be diagnosed in such a case. The patients Rose and Bob also have the same situation that their diseases cannot be diagnosed when Garg’s correlation coefficient [21] and the correlation coefficient of Singh et al. [45] are used.

As depicted in Table 5, the correlation coefficient value between the symptom values of Luna and the symptom values of Typhoid equals to the correlation coefficient value between the symptom values of Luna and the symptom values of Chest problem. Therefore, the disease of Luna cannot

be diagnosed when Chen’s correlation coefficient [7] is used. The patients Rose and MingK also have the same situation.

As shown in Table 7, the correlation coefficient value between the symptom values of each patient and the symptom values of each disease is NULL. That is because the denominator is zero during the computing process of correlation coefficient values. It violates the principle of mathematics. Therefore, the diseases of four patients cannot be diagnosed.

According to the above analyses, it can be known that our proposed directional correlation coefficient is able to accurately diagnose the diseases of four patients. Hence, it shows the feasibility and superiority of our proposed directional correlation coefficient.

Case study: cluster analysis

Cluster analysis is a kind of statistical analysis method guiding the process that splits a set of data into some clusters [24, 36, 37, 65], where the data in the same clusters have more similar characteristic than those in the different clusters. It is an unsupervised process, which is very different from the classification analysis [41]. It has been extended by research scholars into the fuzzy decision-making field [34] and some clustering algorithms were developed to cluster intuitionistic fuzzy sets [15, 50], hesitant fuzzy sets [8], and probabilistic linguistic term sets [29]. Here, we do not plan to develop a novel clustering algorithm to handle PFSs, but use the existing clustering algorithm presented in [8] to validate the effectiveness of our proposed weighted directional correlation coefficient. Suppose that there exists a set of PFSs $\{P_1, P_2, \dots, P_n\}$, then the cluster analysis process guided by this clustering algorithm is presented as follows:

Algorithm 1 (Clustering algorithm)

Step 1 Compute the weighted directional correlation coefficient $\partial_2(P_i, P_j)$ between any two PFSs P_i and P_j in $\{P_1, P_2, \dots, P_n\}$ using Eq. (19) and use all the obtained weighted directional correlation coefficients to form a correlation matrix (CM) $\Gamma_w = (\partial_{ij})_{n \times n}$ where $\partial_{ij} = \partial_2(P_i, P_j)$.

Step 2 Check whether the CM can satisfy the condition that $\Gamma_w^2 \subseteq \Gamma_w$, where $\Gamma_w^2 = \Gamma_w \circ \Gamma_w = (\partial'_{ij})_{n \times n}$ and $\partial'_{ij} = \max_l \{\min\{\partial_{il}, \partial_{lj}\}\}$. If it fails to satisfy the condition, then the equivalent CM $\Gamma_w^{2^k}$ is computed as

$$\Gamma_w \rightarrow \Gamma_w^2 \rightarrow \Gamma_w^4 \rightarrow \dots \rightarrow \Gamma_w^{2^k} \rightarrow \dots \tag{20}$$

until $\Gamma_w^{2^k} = \Gamma_w^{2^{(k+1)}}$.

Step 3 Choose a certain number of values for the confidence level ϑ from the unit interval $[0, 1]$ so as to construct ϑ -cutting matrices $\Gamma_\vartheta = (\chi_{ij}^\vartheta)_{n \times n}$, where the element χ_{ij}^ϑ can be determined as

$$\chi_{ij}^{\vartheta} = \begin{cases} 0, & \text{if } \chi_{ij} < \vartheta \\ 1, & \text{if } \chi_{ij} \geq \vartheta \end{cases} \quad i, j = 1, 2, \dots, n \quad (21)$$

Step 4 The set of PFSs $\{P_1, P_2, \dots, P_n\}$ can be clustered using the following rule: If all the elements in the i th row of Γ_{ϑ} are equal to the elements in the same positions of j th row, then the PFSs P_i and P_j can be considered to be with the similar characteristics and they can be placed in the same clusters.

Step 5 End.

In the following part, a case concerning the analysis of informatization level of smart cities is given to show the superiority of the proposed directional correlation coefficient when it is applied to cluster analysis. In this example, the weighted directional correlation coefficient formula given in (19) is used to compute the correlation coefficient between two PFSs since the informatization level involves

considered, which are IT infrastructure (C_1), information technology talent (C_2), information resource and application (C_3), information result transformation efficiency (C_4), economic benefit (C_5). The weight information of these indexes is set to $w = (0.2, 0.15, 0.3, 0.1, 0.25)^T$. The government information steering committee evaluates the informatization level of ten cities according to their indexes using the PFSs and records the information as shown in Table 8.

Firstly, the clustering algorithm with our proposed weighted directional correlation coefficient is used to group these ten cities according the above evaluation information as follows:

Step 1 Compute the weighted directional correlation coefficient $\partial_2(P_i, P_j)$ between any two PFSs P_i and P_j in $\{P_1, P_2, \dots, P_{10}\}$ using Eq. (19) and use them to form a correlation matrix (CM) $\Gamma_w = (\partial_{ij})_{n \times n}$ as

$$\Gamma_w = \begin{bmatrix} 1.0000 & 0.4434 & 0.7868 & 0.5949 & 0.4027 & 0.3941 & 0.7701 & 0.3685 & 0.6618 & 0.7449 \\ 0.4434 & 1.0000 & 0.4357 & 0.7798 & 0.8876 & 0.6094 & 0.5083 & 0.7048 & 0.5896 & 0.6643 \\ 0.7868 & 0.4357 & 1.0000 & 0.7886 & 0.4490 & 0.6108 & 0.6548 & 0.4420 & 0.5848 & 0.6469 \\ 0.5949 & 0.7798 & 0.7886 & 1.0000 & 0.6129 & 0.8619 & 0.6502 & 0.6378 & 0.5768 & 0.6280 \\ 0.4027 & 0.8876 & 0.4490 & 0.6129 & 1.0000 & 0.4240 & 0.4553 & 0.7272 & 0.6495 & 0.7402 \\ 0.3941 & 0.6094 & 0.6108 & 0.8619 & 0.4240 & 1.0000 & 0.5852 & 0.5824 & 0.2564 & 0.4832 \\ 0.7701 & 0.5083 & 0.6548 & 0.6502 & 0.4553 & 0.5852 & 1.0000 & 0.4072 & 0.6667 & 0.7505 \\ 0.3685 & 0.7048 & 0.4420 & 0.6378 & 0.7272 & 0.5824 & 0.4072 & 1.0000 & 0.6736 & 0.7503 \\ 0.6618 & 0.5896 & 0.5848 & 0.5768 & 0.6495 & 0.2564 & 0.6667 & 0.6736 & 1.0000 & 0.8811 \\ 0.7449 & 0.6643 & 0.6469 & 0.6280 & 0.7402 & 0.4832 & 0.7505 & 0.7503 & 0.8811 & 1.0000 \end{bmatrix}$$

some different factors and the factors usually show different weights or importance.

Step 2 The equivalent CM is computed using the following process

$$\Gamma_w^2 = \begin{bmatrix} 1.0000 & 0.6643 & 0.7868 & 0.7868 & 0.7402 & 0.6108 & 0.7701 & 0.7449 & 0.7449 & 0.7505 \\ 0.6643 & 1.0000 & 0.7798 & 0.7798 & 0.8876 & 0.7798 & 0.6643 & 0.7272 & 0.6736 & 0.7402 \\ 0.7868 & 0.7798 & 1.0000 & 0.7886 & 0.6469 & 0.7886 & 0.7701 & 0.6469 & 0.6618 & 0.7449 \\ 0.7868 & 0.7798 & 0.7886 & 1.0000 & 0.7798 & 0.8619 & 0.6548 & 0.7048 & 0.6502 & 0.6643 \\ 0.7402 & 0.8876 & 0.6469 & 0.7798 & 1.0000 & 0.6129 & 0.7402 & 0.7402 & 0.7402 & 0.7402 \\ 0.6108 & 0.7798 & 0.7886 & 0.8619 & 0.6129 & 1.0000 & 0.6502 & 0.6378 & 0.5896 & 0.6280 \\ 0.7701 & 0.6643 & 0.7701 & 0.6548 & 0.7402 & 0.6502 & 1.0000 & 0.7503 & 0.7505 & 0.7505 \\ 0.7449 & 0.7272 & 0.6469 & 0.7048 & 0.7402 & 0.6378 & 0.7503 & 1.0000 & 0.7503 & 0.7503 \\ 0.7449 & 0.6736 & 0.6618 & 0.6502 & 0.7402 & 0.5896 & 0.7505 & 0.7503 & 1.0000 & 0.8811 \\ 0.7505 & 0.7402 & 0.7449 & 0.6643 & 0.7402 & 0.6280 & 0.7505 & 0.7503 & 0.8811 & 1.0000 \end{bmatrix}$$

Example 7 The informatization level of a city has become an important measure for the popularity and economic strength of a city, which has attracted much attention from government officials in recent years. For example, Digital China Summit has been successfully held for two times in Fuzhou, China. To analyze the informatization level of ten cities, denoted as $\{P_1, P_2, \dots, P_{10}\}$, five indexes should be

$$\Gamma_w^4 = \begin{bmatrix} 1.0000 & 0.7798 & 0.7868 & 0.7868 & 0.7798 & 0.7868 & 0.7701 & 0.7503 & 0.7505 & 0.7505 \\ 0.7798 & 1.0000 & 0.7798 & 0.7798 & 0.8876 & 0.7798 & 0.7701 & 0.7402 & 0.7402 & 0.7449 \\ 0.7868 & 0.7798 & 1.0000 & 0.7886 & 0.7798 & 0.7886 & 0.7701 & 0.7503 & 0.7505 & 0.7505 \\ 0.7868 & 0.7798 & 0.7886 & 1.0000 & 0.7798 & 0.8619 & 0.7701 & 0.7449 & 0.7449 & 0.7505 \\ 0.7798 & 0.8876 & 0.7798 & 0.7798 & 1.0000 & 0.7798 & 0.7402 & 0.7402 & 0.7402 & 0.7402 \\ 0.7868 & 0.7798 & 0.7886 & 0.8619 & 0.7798 & 1.0000 & 0.7701 & 0.7272 & 0.6736 & 0.7449 \\ 0.7701 & 0.7701 & 0.7701 & 0.7701 & 0.7402 & 0.7701 & 1.0000 & 0.7503 & 0.7505 & 0.7505 \\ 0.7503 & 0.7402 & 0.7503 & 0.7449 & 0.7402 & 0.7272 & 0.7503 & 1.0000 & 0.7503 & 0.7503 \\ 0.7505 & 0.7402 & 0.7505 & 0.7449 & 0.7402 & 0.6736 & 0.7505 & 0.7503 & 1.0000 & 0.8811 \\ 0.7505 & 0.7449 & 0.7505 & 0.7505 & 0.7402 & 0.7449 & 0.7505 & 0.7503 & 0.8811 & 1.0000 \end{bmatrix}$$

$$\Gamma_w^8 = \begin{bmatrix} 1.0000 & 0.7798 & 0.7868 & 0.7868 & 0.7798 & 0.7868 & 0.7701 & 0.7503 & 0.7505 & 0.7505 \\ 0.7798 & 1.0000 & 0.7798 & 0.7798 & 0.8876 & 0.7798 & 0.7701 & 0.7503 & 0.7505 & 0.7505 \\ 0.7868 & 0.7798 & 1.0000 & 0.7886 & 0.7798 & 0.7886 & 0.7701 & 0.7503 & 0.7505 & 0.7505 \\ 0.7868 & 0.7798 & 0.7886 & 1.0000 & 0.7798 & 0.8619 & 0.7701 & 0.7503 & 0.7505 & 0.7505 \\ 0.7798 & 0.8876 & 0.7798 & 0.7798 & 1.0000 & 0.7798 & 0.7701 & 0.7503 & 0.7505 & 0.7505 \\ 0.7868 & 0.7798 & 0.7886 & 0.8619 & 0.7798 & 1.0000 & 0.7701 & 0.7503 & 0.7505 & 0.7505 \\ 0.7701 & 0.7701 & 0.7701 & 0.7701 & 0.7701 & 0.7701 & 1.0000 & 0.7503 & 0.7505 & 0.7505 \\ 0.7503 & 0.7503 & 0.7503 & 0.7503 & 0.7503 & 0.7503 & 0.7503 & 1.0000 & 0.7503 & 0.7503 \\ 0.7505 & 0.7505 & 0.7505 & 0.7505 & 0.7505 & 0.7505 & 0.7505 & 0.7503 & 1.0000 & 0.8811 \\ 0.7505 & 0.7505 & 0.7505 & 0.7505 & 0.7505 & 0.7505 & 0.7505 & 0.7503 & 0.8811 & 1.0000 \end{bmatrix}$$

$$\Gamma_w^{16} = \begin{bmatrix} 1.0000 & 0.7798 & 0.7868 & 0.7868 & 0.7798 & 0.7868 & 0.7701 & 0.7503 & 0.7505 & 0.7505 \\ 0.7798 & 1.0000 & 0.7798 & 0.7798 & 0.8876 & 0.7798 & 0.7701 & 0.7503 & 0.7505 & 0.7505 \\ 0.7868 & 0.7798 & 1.0000 & 0.7886 & 0.7798 & 0.7886 & 0.7701 & 0.7503 & 0.7505 & 0.7505 \\ 0.7868 & 0.7798 & 0.7886 & 1.0000 & 0.7798 & 0.8619 & 0.7701 & 0.7503 & 0.7505 & 0.7505 \\ 0.7798 & 0.8876 & 0.7798 & 0.7798 & 1.0000 & 0.7798 & 0.7701 & 0.7503 & 0.7505 & 0.7505 \\ 0.7868 & 0.7798 & 0.7886 & 0.8619 & 0.7798 & 1.0000 & 0.7701 & 0.7503 & 0.7505 & 0.7505 \\ 0.7701 & 0.7701 & 0.7701 & 0.7701 & 0.7701 & 0.7701 & 1.0000 & 0.7503 & 0.7505 & 0.7505 \\ 0.7503 & 0.7503 & 0.7503 & 0.7503 & 0.7503 & 0.7503 & 0.7503 & 1.0000 & 0.7503 & 0.7503 \\ 0.7505 & 0.7505 & 0.7505 & 0.7505 & 0.7505 & 0.7505 & 0.7505 & 0.7503 & 1.0000 & 0.8811 \\ 0.7505 & 0.7505 & 0.7505 & 0.7505 & 0.7505 & 0.7505 & 0.7505 & 0.7503 & 0.8811 & 1.0000 \end{bmatrix}$$

It can be noted that $\Gamma_w^{16} = \Gamma_w^8$. Hence, Γ_w^8 is an equivalent CM.

Step 3 Ten values of the confidence level of the equivalent CM Γ_w^8 can be obtained as 1.0000, 0.8876, 0.8811, 0.8619, 0.7886, 0.7868, 0.7798, 0.7701, 0.7505, 0.7503. According to these values of the confidence level, ten ϑ -cutting matrices can be obtained.

For example, when $\vartheta = 0.8876$, then

$$\Gamma_{w,\vartheta=0.8876} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Step 4 Based on the obtained ϑ -cutting matrices, the informatization level of ten cities can be clustered as shown in Table 9.

Step 5 End.

To compare the weighted directional correlation coefficient with the existing correlation coefficients [7], 21, 45, these existing correlation coefficients are used to handle Example 7.

First, the correlation coefficient proposed by Garg [21] is used to compute the CM and equivalent CM as

$$\Gamma_w^v = \begin{bmatrix} 1.0000 & 0.8097 & 0.9203 & 0.7773 & 0.8328 & 0.5636 & 0.8191 & 0.6812 & 0.7842 & 0.8885 \\ 0.8097 & 1.0000 & 0.8049 & 0.9022 & 0.9203 & 0.7150 & 0.6977 & 0.7660 & 0.7095 & 0.7991 \\ 0.9203 & 0.8049 & 1.0000 & 0.9022 & 0.8097 & 0.7150 & 0.7660 & 0.6977 & 0.7095 & 0.7991 \\ 0.7773 & 0.9022 & 0.9022 & 1.0000 & 0.7773 & 0.8664 & 0.7292 & 0.7292 & 0.6310 & 0.6999 \\ 0.8328 & 0.9203 & 0.8097 & 0.7773 & 1.0000 & 0.5636 & 0.6812 & 0.8191 & 0.7842 & 0.8885 \\ 0.5636 & 0.7150 & 0.7150 & 0.8664 & 0.5636 & 1.0000 & 0.6524 & 0.6524 & 0.3082 & 0.5359 \\ 0.8191 & 0.6977 & 0.7660 & 0.7292 & 0.6812 & 0.6524 & 1.0000 & 0.6652 & 0.7565 & 0.8367 \\ 0.6812 & 0.7660 & 0.6977 & 0.7292 & 0.8191 & 0.6524 & 0.6652 & 1.0000 & 0.7565 & 0.8367 \\ 0.7842 & 0.7095 & 0.7095 & 0.6310 & 0.7842 & 0.3082 & 0.7565 & 0.7565 & 1.0000 & 0.8922 \\ 0.8885 & 0.7991 & 0.7991 & 0.6999 & 0.8885 & 0.5359 & 0.8367 & 0.8367 & 0.8922 & 1.0000 \end{bmatrix}$$

$$\Gamma_w^{v16} = \begin{bmatrix} 1.0000 & 0.9022 & 0.9203 & 0.9022 & 0.9022 & 0.8664 & 0.8367 & 0.8367 & 0.8885 & 0.8885 \\ 0.9022 & 1.0000 & 0.9022 & 0.9022 & 0.9203 & 0.8664 & 0.8367 & 0.8367 & 0.8885 & 0.8885 \\ 0.9203 & 0.9022 & 1.0000 & 0.9022 & 0.9022 & 0.8664 & 0.8367 & 0.8367 & 0.8885 & 0.8885 \\ 0.9022 & 0.9022 & 0.9022 & 1.0000 & 0.9022 & 0.8664 & 0.8367 & 0.8367 & 0.8885 & 0.8885 \\ 0.9022 & 0.9203 & 0.9022 & 0.9022 & 1.0000 & 0.8664 & 0.8367 & 0.8367 & 0.8885 & 0.8885 \\ 0.8664 & 0.8664 & 0.8664 & 0.8664 & 0.8664 & 1.0000 & 0.8367 & 0.8367 & 0.8664 & 0.8664 \\ 0.8367 & 0.8367 & 0.8367 & 0.8367 & 0.8367 & 0.8367 & 1.0000 & 0.8367 & 0.8367 & 0.8367 \\ 0.8367 & 0.8367 & 0.8367 & 0.8367 & 0.8367 & 0.8367 & 0.8367 & 1.0000 & 0.8367 & 0.8367 \\ 0.8885 & 0.8885 & 0.8885 & 0.8885 & 0.8885 & 0.8664 & 0.8367 & 0.8367 & 1.0000 & 0.8922 \\ 0.8885 & 0.8885 & 0.8885 & 0.8885 & 0.8885 & 0.8664 & 0.8367 & 0.8367 & 0.8922 & 1.0000 \end{bmatrix}$$

When Garg's correlation coefficient [21] is used, the obtained clustering result is listed in Table 10.

Similarly, the correlation coefficient that was proposed by Chen [7] is also used to compute the CM and equivalent CM as

$$\Gamma_w^v = \begin{bmatrix} 1.0000 & -0.1297 & 0.4634 & -0.2997 & 0.0069 & -0.5037 & 0.2763 & -0.4232 & 0.3488 & 0.3832 \\ -0.1297 & 1.0000 & -0.0997 & 0.3859 & 0.4634 & 0.0440 & -0.5237 & -0.0586 & -0.0210 & -0.2645 \\ 0.4634 & -0.0997 & 1.0000 & 0.3859 & -0.1297 & 0.0440 & -0.0586 & -0.5237 & -0.0210 & -0.2645 \\ -0.2997 & 0.3866 & 0.3859 & 1.0000 & -0.2997 & 0.5521 & -0.2649 & -0.2649 & 0.3332 & -0.8589 \\ 0.0069 & 0.4864 & -0.1297 & -0.2997 & 1.0000 & -0.5037 & -0.4232 & 0.2763 & 0.3488 & 0.3832 \\ -0.5037 & 0.0442 & 0.0440 & 0.5521 & -0.5037 & 1.0000 & 0.0084 & 0.0084 & -0.7170 & -0.6490 \\ 0.2763 & -0.5237 & -0.0586 & -0.2649 & -0.4232 & 0.0084 & 1.0000 & 0.0013 & 0.2104 & 0.2409 \\ 0.4232 & -0.0586 & -0.5237 & -0.2649 & 0.2763 & 0.0084 & 0.0013 & 1.0000 & 0.2104 & 0.2409 \\ 0.3488 & 0.0210 & -0.0210 & 0.3332 & 0.3488 & -0.7170 & 0.2104 & 0.2104 & 1.0000 & 0.6917 \\ 0.3832 & -0.2645 & -0.2645 & -0.8589 & 0.3832 & -0.6490 & 0.2409 & 0.2409 & 0.6917 & 1.0000 \end{bmatrix}$$

$$\Gamma_w^{//8} = \begin{bmatrix} 1.0000 & 0.3859 & 0.4634 & 0.3859 & 0.3859 & 0.3859 & 0.2763 & 0.2763 & 0.3832 & 0.3832 \\ 0.3859 & 1.0000 & 0.3859 & 0.3859 & 0.4634 & 0.3859 & 0.2763 & 0.2763 & 0.3832 & 0.3832 \\ 0.4634 & 0.3859 & 1.0000 & 0.3859 & 0.3859 & 0.3859 & 0.2763 & 0.2763 & 0.3832 & 0.3832 \\ 0.3859 & 0.3859 & 0.3859 & 1.0000 & 0.3859 & 0.5521 & 0.2763 & 0.2763 & 0.3832 & 0.3832 \\ 0.3859 & 0.4634 & 0.3859 & 0.3859 & 1.0000 & 0.3859 & 0.2763 & 0.2763 & 0.3832 & 0.3832 \\ 0.3859 & 0.3859 & 0.3859 & 0.5521 & 0.3859 & 1.0000 & 0.2763 & 0.2763 & 0.3832 & 0.3832 \\ 0.2763 & 0.2763 & 0.2763 & 0.2763 & 0.2763 & 0.2763 & 1.0000 & 0.2763 & 0.2763 & 0.2763 \\ 0.2763 & 0.2763 & 0.2763 & 0.2763 & 0.2763 & 0.2763 & 0.2763 & 1.0000 & 0.2763 & 0.2763 \\ 0.3832 & 0.3832 & 0.3832 & 0.3832 & 0.3832 & 0.3832 & 0.2763 & 0.2763 & 1.0000 & 0.6919 \\ 0.3832 & 0.3832 & 0.3832 & 0.3832 & 0.3832 & 0.3832 & 0.2763 & 0.2763 & 0.6919 & 1.0000 \end{bmatrix}$$

When Chen’s correlation coefficient [7] is used, the obtained clustering result is shown in Table 11.

Finally, the correlation coefficient of Singh et al. [45] is used to compute the CM and equivalent CM as

1. When the correlation coefficients proposed by Garg [21], Chen [7], and Singh et al. [45] are used, there may exist the case that two PFSs P_i and P_j are different, but the correlation coefficient between the

$$\Gamma_w''' = \begin{bmatrix} 1.0000 & 0.6906 & 0.8531 & 0.7621 & 0.6942 & 0.6728 & 0.7923 & 0.6341 & 0.7728 & 0.8148 \\ 0.6906 & 1.0000 & 0.6755 & 0.8351 & 0.8351 & 0.7217 & 0.6763 & 0.7323 & 0.7276 & 0.7639 \\ 0.8531 & 0.6755 & 1.0000 & 0.8351 & 0.6906 & 0.7217 & 0.7323 & 0.6763 & 0.7276 & 0.7639 \\ 0.7621 & 0.8351 & 0.8351 & 1.0000 & 0.7621 & 0.8322 & 0.7322 & 0.7322 & 0.7027 & 0.7771 \\ 0.6942 & 0.8351 & 0.6906 & 0.7621 & 1.0000 & 0.6728 & 0.6341 & 0.7923 & 0.7728 & 0.8148 \\ 0.6728 & 0.7217 & 0.7217 & 0.8322 & 0.6728 & 1.0000 & 0.7204 & 0.7204 & 0.5867 & 0.6830 \\ 0.7923 & 0.6763 & 0.7323 & 0.7322 & 0.6341 & 0.7204 & 1.0000 & 0.5577 & 0.7204 & 0.7364 \\ 0.6341 & 0.7323 & 0.6763 & 0.7322 & 0.7923 & 0.7204 & 0.5577 & 1.0000 & 0.7204 & 0.7364 \\ 0.7728 & 0.7276 & 0.7276 & 0.7027 & 0.7728 & 0.5867 & 0.7204 & 0.7204 & 1.0000 & 0.8429 \\ 0.8148 & 0.7639 & 0.7639 & 0.7771 & 0.8148 & 0.6830 & 0.7364 & 0.7364 & 0.8429 & 1.0000 \end{bmatrix}$$

$$\Gamma_w'''^{16} = \begin{bmatrix} 1.0000 & 0.8351 & 0.8531 & 0.8351 & 0.8351 & 0.8322 & 0.7923 & 0.7923 & 0.8148 & 0.8148 \\ 0.8351 & 1.0000 & 0.8351 & 0.8351 & 0.8351 & 0.8322 & 0.7923 & 0.7923 & 0.8148 & 0.8148 \\ 0.8531 & 0.8351 & 1.0000 & 0.8351 & 0.8351 & 0.8322 & 0.7923 & 0.7923 & 0.8148 & 0.8148 \\ 0.8351 & 0.8351 & 0.8351 & 1.0000 & 0.8351 & 0.8322 & 0.7923 & 0.7923 & 0.8148 & 0.8148 \\ 0.8351 & 0.8351 & 0.8351 & 0.8351 & 1.0000 & 0.8322 & 0.7923 & 0.7923 & 0.8148 & 0.8148 \\ 0.8322 & 0.8322 & 0.8322 & 0.8322 & 0.8322 & 1.0000 & 0.7923 & 0.7923 & 0.8148 & 0.8148 \\ 0.7923 & 0.7923 & 0.7923 & 0.7923 & 0.7923 & 0.7923 & 1.0000 & 0.7923 & 0.7923 & 0.7923 \\ 0.7923 & 0.7923 & 0.7923 & 0.7923 & 0.7923 & 0.7923 & 0.7923 & 1.0000 & 0.7923 & 0.7923 \\ 0.8148 & 0.8148 & 0.8148 & 0.8148 & 0.8148 & 0.8148 & 0.7923 & 0.7923 & 1.0000 & 0.8429 \\ 0.8148 & 0.8148 & 0.8148 & 0.8148 & 0.8148 & 0.8148 & 0.7923 & 0.7923 & 0.8429 & 1.0000 \end{bmatrix}$$

When the correlation coefficient of Singh et al. [45] is used, the obtained clustering result is listed in Table 12.

It can be observed that the clustering result obtained from the clustering algorithm with the proposed weighted directional correlation coefficient is different from the clustering results that are obtained from the clustering algorithms with the correlation coefficients proposed by Garg [21], Chen [7] and Singh et al. [45]. The reasons can be analyzed as follows:

PFSs P_i and P_k equals to the correlation coefficient between the PFSs P_j and P_k . For example, in the CM Γ_w' , $\partial_2(P_2, P_4) = \partial_2(P_3, P_4) = 0.9022$. In the CM Γ_w'' , $\partial_2(P_7, P_4) = \partial_2(P_8, P_4) = -0.2649$. In the CM Γ_w''' , $\partial_2(P_2, P_6) = \partial_2(P_3, P_6) = 0.7217$. In the CM Γ_w , the drawback is overcome and it does not incur the above case since the proposed weighted directional correlation coefficient considers not only the MDs and NMDs, but also the SoCs and DoCs. Thus, the clustering result obtained from the clustering algorithm with the pro-

Table 8 The evaluation information for the informatization level of ten cities

	C_1	C_2	C_3	C_4	C_5
P_1	(0.2, 0.5)	(0.1, 0.8)	(0.6, 0.2)	(0.3, 0.4)	(0.1, 0.3)
P_2	(0.6, 0.1)	(0.7, 0.2)	(0.3, 0.4)	(0.1, 0.8)	(0.5, 0.2)
P_3	(0.1, 0.6)	(0.2, 0.7)	(0.4, 0.3)	(0.8, 0.1)	(0.2, 0.5)
P_4	(0.5, 0.5)	(0.4, 0.4)	(0.3, 0.3)	(0.7, 0.7)	(0.5, 0.5)
P_5	(0.5, 0.2)	(0.8, 0.1)	(0.2, 0.6)	(0.4, 0.3)	(0.3, 0.1)
P_6	(0.6, 0.6)	(0.2, 0.2)	(0.1, 0.1)	(0.5, 0.5)	(0.7, 0.7)
P_7	(0.2, 0.7)	(0.1, 0.3)	(0.8, 0.1)	(0.4, 0.3)	(0.6, 0.2)
P_8	(0.7, 0.2)	(0.3, 0.1)	(0.1, 0.8)	(0.3, 0.4)	(0.2, 0.6)
P_9	(0.2, 0.2)	(0.4, 0.4)	(0.7, 0.7)	(0.6, 0.6)	(0.1, 0.1)
P_{10}	(0.3, 0.3)	(0.5, 0.5)	(0.6, 0.6)	(0.2, 0.2)	(0.4, 0.4)

posed weighted directional correlation coefficient is more reasonable since the correlation coefficient is an indicator used to perform cluster analysis and it has a great influence on the clustering results.

- From Tables 9, 10, 11 and 12, it can be seen that the confidence level value belongs to the interval [0.7503, 0.8876] when our proposed weighted directional correlation coefficient is utilized. The confidence level value belongs to the interval [0.8367, 0.9203] when the correlation coefficient proposed by Garg [21]

is used. The confidence level value is in the interval [0.2763, 0.6919] when the correlation coefficient that was proposed by Chen [7] is used. The confidence level value is in the interval [0.7923, 0.8531] when the correlation coefficient proposed by Singh et al. [45] is used. Therefore, the confidence level obtained from Chen [7] is much smaller than that obtained from our proposed weighted directional correlation coefficient, Garg’s correlation coefficient [21], and the correlation coefficient of Singh et al. [45]. Moreover, some of the confidence level values are smaller than 0.5. It does not satisfy the principle in statistics. The interval of the confidence level obtained from our proposed weighted directional correlation coefficient is bigger than that obtained from Garg’s correlation coefficient [21] and correlation coefficient of Singh et al. [45]. It implies that the differences of confidence level values obtained from our proposed weighted directional correlation coefficient are bigger. Therefore, it can better reflect the differences among the different clusters.

According to the above analysis, it can be observed that our proposed weighted directional correlation coefficient can achieve more reasonable result than the correlation coefficients proposed by Garg [21], Chen [7], and Singh et al. [45] when the correlation coefficient is applied to cluster analysis.

Table 9 The clustering result of the informatization level of ten cities when the weighted directional correlation coefficient is used

Class	Confidence level	Clustering result
10	$0.8876 < \lambda \leq 1$	$\{P_1\}, \{P_2\}, \{P_3\}, \{P_4\}, \{P_5\}, \{P_6\}, \{P_7\}, \{P_8\}, \{P_9\}, \{P_{10}\}$
9	$0.8811 < \lambda \leq 0.8876$	$\{P_1\}, \{P_2, P_5\}, \{P_3\}, \{P_4\}, \{P_6\}, \{P_7\}, \{P_8\}, \{P_9\}, \{P_{10}\}$
8	$0.8619 < \lambda \leq 0.8811$	$\{P_1\}, \{P_2, P_5\}, \{P_3\}, \{P_4\}, \{P_6\}, \{P_7\}, \{P_8\}, \{P_9, P_{10}\}$
7	$0.7886 < \lambda \leq 0.8619$	$\{P_1\}, \{P_2, P_5\}, \{P_3\}, \{P_4, P_6\}, \{P_7\}, \{P_8\}, \{P_9, P_{10}\}$
6	$0.7868 < \lambda \leq 0.7886$	$\{P_1\}, \{P_2, P_5\}, \{P_3, P_4, P_6\}, \{P_7\}, \{P_8\}, \{P_9, P_{10}\}$
5	$0.7798 < \lambda \leq 0.7868$	$\{P_1, P_3, P_4, P_6\}, \{P_2, P_5\}, \{P_7\}, \{P_8\}, \{P_9, P_{10}\}$
4	$0.7701 < \lambda \leq 0.7798$	$\{P_1, P_2, P_3, P_4, P_5, P_6\}, \{P_7\}, \{P_8\}, \{P_9, P_{10}\}$
3	$0.7505 < \lambda \leq 0.7701$	$\{P_1, P_2, P_3, P_4, P_5, P_6, P_7\}, \{P_8\}, \{P_9, P_{10}\}$
2	$0.7503 < \lambda \leq 0.7505$	$\{P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_9, P_{10}\}, \{P_8\}$
1	$0 < \lambda \leq 0.7503$	$\{P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9, P_{10}\}$

Table 10 The clustering result of the informatization level of ten cities when Garg’s correlation coefficient [21] is used

Class	Confidence level	Clustering result
10	$0.9203 < \lambda \leq 1$	$\{P_1\}, \{P_2\}, \{P_3\}, \{P_4\}, \{P_5\}, \{P_6\}, \{P_7\}, \{P_8\}, \{P_9\}, \{P_{10}\}$
8	$0.9022 < \lambda \leq 0.9203$	$\{P_1, P_3\}, \{P_2, P_5\}, \{P_4\}, \{P_6\}, \{P_7\}, \{P_8\}, \{P_9\}, \{P_{10}\}$
6	$0.8922 < \lambda \leq 0.9022$	$\{P_1, P_2, P_3, P_4, P_5\}, \{P_6\}, \{P_7\}, \{P_8\}, \{P_9\}, \{P_{10}\}$
5	$0.8885 < \lambda \leq 0.8922$	$\{P_1, P_2, P_3, P_4, P_5\}, \{P_6\}, \{P_7\}, \{P_8\}, \{P_9, P_{10}\}$
4	$0.8664 < \lambda \leq 0.8885$	$\{P_1, P_2, P_3, P_4, P_5, P_9, P_{10}\}, \{P_6\}, \{P_7\}, \{P_8\}$
3	$0.8367 < \lambda \leq 0.8664$	$\{P_1, P_2, P_3, P_4, P_5, P_6, P_9, P_{10}\}, \{P_7\}, \{P_8\}$
1	$0 < \lambda \leq 0.8367$	$\{P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9, P_{10}\}$

Table 11 The clustering result of the informatization level of ten cities when Chen’s correlation coefficient [7] is used

Class	Confidence level	Clustering result
10	$0.6919 < \lambda \leq 1$	$\{P_1\}, \{P_2\}, \{P_3\}, \{P_4\}, \{P_5\}, \{P_6\}, \{P_7\}, \{P_8\}, \{P_9\}, \{P_{10}\}$
9	$0.5521 < \lambda \leq 0.6919$	$\{P_1\}, \{P_2\}, \{P_3\}, \{P_4\}, \{P_5\}, \{P_6\}, \{P_7\}, \{P_8\}, \{P_9, P_{10}\}$
8	$0.4634 < \lambda \leq 0.5521$	$\{P_1\}, \{P_2\}, \{P_3\}, \{P_4, P_6\}, \{P_5\}, \{P_7\}, \{P_8\}, \{P_9, P_{10}\}$
6	$0.3859 < \lambda \leq 0.4634$	$\{P_1, P_3\}, \{P_2, P_5\}, \{P_4, P_6\}, \{P_7\}, \{P_8\}, \{P_9, P_{10}\}$
4	$0.3832 < \lambda \leq 0.3859$	$\{P_1, P_2, P_3, P_4, P_5, P_6\}, \{P_7\}, \{P_8\}, \{P_9, P_{10}\}$
3	$0.2763 < \lambda \leq 0.3832$	$\{P_1, P_2, P_3, P_4, P_5, P_6, P_9, P_{10}\}, \{P_7\}, \{P_8\}$
1	$0 < \lambda \leq 0.2763$	$\{P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9, P_{10}\}$

Table 12 The clustering result of the informatization level of ten cities when the correlation coefficient of Singh et al. [45] is used

Class	Confidence level	Clustering result
10	$0.8531 < \lambda \leq 1$	$\{P_1\}, \{P_2\}, \{P_3\}, \{P_4\}, \{P_5\}, \{P_6\}, \{P_7\}, \{P_8\}, \{P_9\}, \{P_{10}\}$
8	$0.8429 < \lambda \leq 0.8531$	$\{P_1, P_3\}, \{P_2, P_5\}, \{P_4\}, \{P_6\}, \{P_7\}, \{P_8\}, \{P_9\}, \{P_{10}\}$
7	$0.8351 < \lambda \leq 0.8429$	$\{P_1, P_3\}, \{P_2, P_5\}, \{P_4\}, \{P_6\}, \{P_7\}, \{P_8\}, \{P_9, P_{10}\}$
5	$0.8322 < \lambda \leq 0.8351$	$\{P_1, P_2, P_3, P_4, P_5\}, \{P_6\}, \{P_7\}, \{P_8\}, \{P_9, P_{10}\}$
4	$0.8148 < \lambda \leq 0.8322$	$\{P_1, P_2, P_3, P_4, P_5, P_6\}, \{P_7\}, \{P_8\}, \{P_9, P_{10}\}$
3	$0.7923 < \lambda \leq 0.8148$	$\{P_1, P_2, P_3, P_4, P_5, P_6, P_9, P_{10}\}, \{P_7\}, \{P_8\}$
1	$0 < \lambda \leq 0.7923$	$\{P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9, P_{10}\}$

Conclusions

In this paper, we focus on developing some novel directional correlation coefficient measures for PFSs by considering their MD, NMD, SoC, and DoC. The drawbacks of the existing correlation coefficients are analyzed. Afterwards, we propose the information energy, correlation, and two novel directional correlation coefficients for PFSs by considering their MDs, NMDs, SoCs, and DoCs. Their properties are investigated. The proposed directional correlation coefficients are used in two important applications, which are medical diagnosis and cluster analysis. An extended example concerning the Chinese medical diagnosis is presented to show the application of the proposed directional correlation coefficient and the superiority in the medical diagnosis. Another case concerning the analysis of informatization level of smart cities is given to show the application of the proposed weighted directional correlation coefficient and also verify the superiority in the cluster analysis. Comparative analyses show that our proposed directional correlation coefficients are better than the existing correlation coefficients when they are used to solve the problems of the medical diagnosis and cluster analysis.

Compared with the previous correlation coefficients, the proposed directional correlation coefficients show the following advantages:

1. The proposed directional correlation coefficients include the important SoC and DoC parameters to better measure the relationship between two PFSs, which have already been verified to be effective in the distance measures.
2. For the proposed directional correlation coefficients of PFSs, there does not exist the case that the denominator equals to zero. They can overcome the drawbacks of the existing correlation coefficients.

In the future research, we intend to use the proposed directional correlation coefficients for improving the TOPSIS method and apply it to solve multicriteria decision-making problems. Moreover, we will focus on the applications of PFSs in IoT and cloud computing.

Acknowledgements This work was supported in part by the National Natural Science Foundation of China under Grant nos. 61872086 and 61972093, Digital Fujian Institute of Big Data for Agriculture and Forestry under Grant no. KJG18019A, and the startup project of doctoral research of Fujian Normal University.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long

as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adabitar Firozja M, Agheli B, Baloui Jamkhaneh E (2020) A new similarity measure for Pythagorean fuzzy sets. *Compl Intell Syst* 6:67–74
- Akram M, Dudek WA, Ilyas F (2019) Group decision-making based on Pythagorean fuzzy TOPSIS method. *Int J Intell Syst* 34(7):1455–1475
- Arora R, Garg H (2018) A robust correlation coefficient measure of dual hesitant fuzzy soft sets and their application in decision making. *Eng Appl Artif Intell* 72:80–92
- Atanassov K (1986) Intuitionistic fuzzy sets. *Fuzzy Sets Syst* 20(1):87–96
- Cao Q, Liu XD, Wang ZW, Zhang ST, Wu J (2020) Recommendation decision-making algorithm for sharing accommodation using probabilistic hesitant fuzzy sets and bipartite network projection. *Compl Intell Syst* 6:431–445
- Can GF, Demirok S (2019) Universal usability evaluation by using an integrated fuzzy multi criteria decision making approach. *Int J Intell Comput Cybernet* 12(2):194–223
- Chen TY (2019) Multiple criteria decision analysis under complex uncertainty: a pearson-like correlation-based Pythagorean fuzzy compromise approach. *Int J Intell Syst* 34(1):114–151
- Chen N, Xu ZS, Xia MM (2013) Correlation coefficients of hesitant fuzzy sets and their applications to cluster analysis. *Appl Math Model* 37(4):2197–2211
- Chiang DA, Lin NP (1999) Correlation of fuzzy sets. *Fuzzy Sets Syst* 102(2):221–226
- DavoudabadiMohagheghi MMSV (2020) A new last aggregation method of multi-attributes group decision making based on concepts of TODIM, WASPAS and TOPSIS under interval-valued intuitionistic fuzzy uncertainty. *Knowl Inf Syst* 62:1371–1391
- Du WS (2019) Correlation and correlation coefficient of generalized orthopair fuzzy sets. *Int J Intell Syst* 34(4):564–583
- Fatih AKM, Gul M (2019) AHP-TOPSIS integration extended with Pythagorean fuzzy sets for information security risk analysis. *Compl Intell Syst* 5:113–126
- Feng X, Shang XP, Xu Y, Wang J (2020) A method to multi-attribute decision-making based on interval-valued q-rung dual hesitant linguistic Maclaurin symmetric mean operators. *Compl Intell Syst* 6:447–468
- Garg H (2020) Neutrality operations-based Pythagorean fuzzy aggregation operators and its applications to multiple attribute group decision-making process. *J Ambient Intell Human Comput* 11:3021–3041
- Garg H, Rani D (2019) A robust correlation coefficient measure of complex intuitionistic fuzzy sets and their applications in decision-making. *Appl Intell* 49(2):496–512
- Garg H (2019) Novel neutrality operation-based Pythagorean fuzzy geometric aggregation operators for multiple attribute group decision analysis. *Int J Intell Syst* 34(10):2459–2489
- Garg H (2019) New logarithmic operational laws and their aggregation operators for Pythagorean fuzzy set and their applications. *Int J Intell Syst* 34(1):82–106
- Garg H (2018) Linguistic Pythagorean fuzzy sets and its applications in multiattribute decision-making process. *Int J Intell Syst* 33(6):1234–1263
- Garg H (2018) Generalised Pythagorean fuzzy geometric interactive aggregation operators using Einstein operations and their application to decision making. *J Exp Theor Artif Intell* 30(6):763–794
- Garg H (2018) Some methods for strategic decision-making problems with immediate probabilities in Pythagorean fuzzy environment. *Int J Intell Syst* 33(4):687–712
- Garg H (2016) A novel correlation coefficients between Pythagorean fuzzy sets and its applications to decision-making processes. *Int J Intell Syst* 31(12):1234–1252
- Huang C, Lin MW, Xu ZS (2020) Pythagorean fuzzy MULTIMOORA method based on distance measure and score function: its application in multicriteria decision making process. *Knowl Inf Syst* 62:4373–4406
- Jan N, Zedam L, Mahmood T, Rak E, Ali Z (2020) Generalized dice similarity measures for q-rung orthopair fuzzy sets with applications. *Compl Intell Syst* 6:545–558
- Kuo RJ, Lin TC, Zulvia FE, Tsai CY (2018) A hybrid metaheuristic and kernel intuitionistic fuzzy c-means algorithm for cluster analysis. *Appl Soft Comput J* 67:299–308
- Lei Q, Xu ZS (2015) Fundamental properties of intuitionistic fuzzy calculus. *Knowl-Based Syst* 76:1–16
- Li HM, Lv LL, Li F, Wang LY, Xia Q (2020) A novel approach to emergency risk assessment using FMEA with extended MULTIMOORA method under interval-valued Pythagorean fuzzy environment. *Int J Intell Comput Cybernet* 13(1):41–65
- Li DQ, Zeng WY (2018) Distance measure of Pythagorean fuzzy sets. *Int J Intell Syst* 33(2):348–361
- Liao HC, Xu ZS, Zeng XJ (2015) Novel correlation coefficients between hesitant fuzzy sets and their application in decision making. *Knowl-Based Syst* 82:115–127
- Lin MW, Huang C, Xu ZS, Chen RQ (2020) Evaluating IoT platforms using integrated probabilistic linguistic MCDM method. *IEEE Internet Things J*. <https://doi.org/10.1109/JIOT.2020.2997133>
- Lin MW, Xu WS, Lin ZP, Chen RQ (2020) Determine OWA operator weights using kernel density estimation. *Econ Res Ekonomika Istraživanja* 33(1):1441–1464
- Lin MW, Huang C, Xu ZS (2020) MULTIMOORA based MCDM model for site selection of car sharing station under picture fuzzy environment. *Sustain Cities Soc* 53:101873
- Lin MW, Li XM, Chen LF (2020) Linguistic q-rung orthopair fuzzy sets and their interactional partitioned Heronian mean aggregation operators. *Int J Intell Syst* 35(2):217–249
- Lin MW, Wang HB, Xu ZS (2020) TODIM-based multi-criteria decision-making method with hesitant fuzzy linguistic term sets. *Artif Intell Rev* 53:3647–3671
- Lin MW, Huang C, Xu ZS (2019) TOPSIS method based on correlation coefficient and entropy measure for linguistic Pythagorean fuzzy sets and its application to multiple attribute decision making. *Complex* 2019:6967390. <https://doi.org/10.1155/2019/6967390>
- Lin MW, Chen ZY, Liao HC, Xu ZS (2019) ELECTRE II method to deal with probabilistic linguistic term sets and its application to edge computing. *Nonlinear Dyn* 96:2125–2143
- Ma ZZ, Zhu JJ, Ponnambalam K, Zhang ST (2019) A clustering method for large-scale group decision-making with multi-stage hesitant fuzzy linguistic terms. *Inf Fusion* 50:231–250
- Mehta V, Bawa S, Singh J (2020) Analytical review of clustering techniques and proximity measures. *Artif Intell Rev* 53:5995–6023

38. Mishra S, Sahoo MN, Bakshi S, Rodrigues JJPC (2020) Dynamic resource allocation in fog-cloud hybrid systems using multicriteria AHP techniques. *IEEE Internet Things J* 7(9):8993–9000
39. Mohamed A, Najafabadi MK, Wah YB, Kamaru Zaman EA, Maskat R (2020) The state of the art and taxonomy of big data analytics: view from new big data framework. *Artif Intell Rev* 53:989–1037
40. Mitchell HB (2004) A correlation coefficient for intuitionistic fuzzy sets. *Int J Intell Syst* 19(5):483–490
41. Nguyen XT, Nguyen VD, Nguyen VH, Garg H (2019) Exponential similarity measures for Pythagorean fuzzy sets and their applications to pattern recognition and decision-making process. *Compl Intell Syst* 5:217–228
42. Nikfalazar S, Yeh CH, Bedingfield S, Khorshidi HA (2020) Missing data imputation using decision trees and fuzzy clustering with iterative learning. *Knowl Inf Syst* 62:2419–2437
43. Peng XD, Dai JG (2020) A bibliometric analysis of neutrosophic set: two decades review from 1998 to 2017. *Artif Intell Rev* 53:199–255
44. Peng XD (2019) New similarity measure and distance measure for Pythagorean fuzzy set. *Compl Intell Syst* 5:101–111
45. Singh S, Ganie AH (2020) On some correlation coefficients in Pythagorean fuzzy environment with applications. *Int J Intell Syst* 35:682–717
46. Sun J, Wang J, Chen J, Ding G, Lin F (2020) Clustering analysis for internet of spectrum devices: real-world data analytics and applications. *IEEE Internet Things J* 7(5):4485–4496
47. Song CY, Xu ZS, Zhao H (2019) New correlation coefficients between probabilistic hesitant fuzzy sets and their applications in cluster analysis. *Int J Fuzzy Syst* 21(2):355–368
48. Szmids E, Kacprzyk J (2010) Correlation of intuitionistic fuzzy sets. *Lecture Notes Comput Sci* 6178:169–177
49. Thao NX (2020) A new correlation coefficient of the Pythagorean fuzzy sets and its applications. *Soft Comput* 24:9467–9478
50. Thao NX, Ali M, Smarandache F (2019) An intuitionistic fuzzy clustering algorithm based on a new correlation coefficient with application in medical diagnosis. *J Intell Fuzzy Syst* 36(1):189–198
51. Verma R, Merigo JM (2019) On generalized similarity measures for Pythagorean fuzzy sets and their applications to multiple attribute decision-making. *Int J Intell Syst* 34(10):2556–2583
52. Wan SP, Jin Z, Dong JY (2020) A new order relation for Pythagorean fuzzy numbers and application to multi-attribute group decision making. *Knowl Inf Syst* 62:751–785
53. Wan SP, Jin Z, Dong JY (2018) Pythagorean fuzzy mathematical programming method for multi-attribute group decision making with Pythagorean fuzzy truth degrees. *Knowl Inf Syst* 55:437–466
54. Wang J, Gao H, Wei GW (2019) The generalized Dice similarity measures for Pythagorean fuzzy multiple attribute group decision making. *Int J Intell Syst* 34(6):1158–1183
55. Wang LY, Xia Q, Li HM, Cao YC (2019) Multi-criteria decision making method based on improved cosine similarity measure with interval neutrosophic sets. *Int J Intell Comput Cybern* 12(3):414–423
56. Xiao FY, Ding WP (2019) Divergence measure of Pythagorean fuzzy sets and its application in medical diagnosis. *Appl Soft Comput* 79:254–267
57. Xu ZS, Xia MM (2011) On distance and correlation measures of hesitant fuzzy information. *Int J Intell Syst* 26:410–425
58. Xu ZS, Chen J, Wu JJ (2008) Clustering algorithm for intuitionistic fuzzy sets. *Inf Sci* 178(19):3775–3790
59. Yager RR (2019) Extending set measures to Pythagorean fuzzy sets. *Int J Fuzzy Syst* 21(2):343–354
60. Yager RR (2017) Generalized orthopair fuzzy sets. *IEEE Trans Fuzzy Syst* 25(5):1222–1230
61. Yager RR (2014) Pythagorean membership grades in multicriteria decision making. *IEEE Trans Fuzzy Syst* 22(4):958–965
62. Zadeh LA (1965) Fuzzy sets. *Inf Control* 8(3):338–353
63. Zeng WY, Li DQ, Yin Q (2018) Distance and similarity measures of Pythagorean fuzzy sets and their applications to multiple criteria group decision making. *Int J Intell Syst* 33(11):2236–2254
64. Zhang XL, Xu ZS (2014) Extension of TOPSIS to multiple criteria decision making with Pythagorean fuzzy sets. *Int J Intell Syst* 29(12):1061–1078
65. Zhang XL (2018) Pythagorean fuzzy clustering analysis: A hierarchical clustering algorithm with the ratio index-based ranking methods. *Int J Intell Syst* 33(9):1798–1822

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.