Journal of KONBiN 2022
Volume 52, Issue 4
DOI 10.2478/jok-2022-0041

Hoang NGUYEN
Gdynia Maritime University (Uniwersytet Morski w Gdyni)

AN APPLICATION OF THE PYTHAGOREAN FUZZY SETS IN THE FAULT DIAGNOSIS
Zastosowanie zbiorów rozmytych Pitagorasa w technicznej diagnostyce

Abstract: *In this paper, a comprehensive review and critical analyses of methods based on the ordinary fuzzy set, Atanassov's intuitionistic fuzzy set, and its extensions have been conducted to show their limitations and defects. Then, a novel similarity measure based on the generalized score function has been introduced that incorporates the significance (importance) of information, making it more intuitive to compare them. The proposed method is employed for the fault diagnosis of steam turbine generator unit under Pythagorean fuzzy environment. Ten fault types of rotating machines are established as failure patterns in nine different vibration frequency ranges, expressed in terms of Pythagorean fuzzy numbers. The superiority of the proposed method in dealing with uncertain and vague information is shown by comparing it with some existing measures in numerical examples.*

Keywords: Pythagorean fuzzy set, score function, similarity measure, fault diagnosis, steam turbine generator

Streszczenie: *W artykule dokonano kompleksowego przeglądu i analiz krytycznych metod opartych na klasycznym zbiorze rozmytym lub intuicjonistycznym zbiorze rozmytym Atanassova i ich rozszerzeniach w celu wykazania ich ograniczeń i wad. Następnie wprowadzono na podstawie miary wiedzy, nową miarę podobieństwa, która uwzględnia znaczenie (ważność) informacji, czyniąc je bardziej intuicyjnymi przy ich porównywaniu. Zaproponowaną metodę weryfikuje się w przypadku diagnozowania uszkodzeń zespołu turbogeneratora w rozmytym środowisku. Dziesięć typów uszkodzeń turbogeneratora jest określanych jako wzorce uszkodzeń wyrażonych za pomocą liczb rozmytych Pitagorasa opisujących ich symptomy w dziewięciu różnych zakresach częstotliwości drgań. Poprzez porównanie z niektórymi istniejącymi miarami w kilku przykładach liczbowych pokazano przewagę proponowanej metody w opisaniu niedokładnych i niepewnych informacji.*

Słowa kluczowe: zbiory rozmyte Pitagorasa, miara wiedzy, miara podobieństwa, diagnostyka techniczna, turbo-generator

1. Introduction

The fault diagnosis of a technical system is critical for its status monitoring in operation and maintenance. Signal-based fault diagnosis is the most popular approach in machine health monitoring. The typical signals used for diagnosing can be vibration signal [4], acoustic emission signal [13] or current signal [3]. Among them, vibration signal analysis is the most widely used method to identify some particular faults because it is easy to measure and can provide highly accurate information about the machine's health condition [15]. Nowadays, most of the fault diagnosis methods are usually based on the vibration signals or the signals transformed from the vibration signals based upon several popular signal processing techniques, including fast Fourier transform (FFT), wavelet transform (WT), and short-time Fourier transform (STFT). The fault diagnosis performance of the signal-based approach highly depends on the procedure of feature extraction in which characteristic features are extracted from vibration signals. After extracting the fault features from the fault signals, an intelligent decision-maker based on machine learning algorithms is used to determine the type of fault occurring. Traditionally, feature extraction exploits signal processing techniques to extract information from the fault signal in the time domain, frequency domain, and time-frequency domain [11]. However, the diagnosing accuracy of the traditional approaches depends on the signal processing technique and requires expert knowledge [6]. Due to the expert knowledge requirement, it is difficult to suggest a generalized framework for feature extraction. Although classical machine learning models, such as support vector machine (SVM) and k-nearest neighbour, have achieved remarkable progress over the past years, some drawbacks still exist when facing higher industrial requirements [5].

The first concept of intuitionistic fuzzy set (IFS) was introduced by Atanassov as a generalization of fuzzy set (FS) theory. The IFS can be viewed as an alternative approach to an ordinary fuzzy set to deal with imperfect information. The uncertainty of information is expressed by a membership function $\mu(x)$ in FSs. But in IFSs, it is represented by membership $\mu(x)$, and non-membership $\mu(x)$ functions, as shown in Fig. 1.

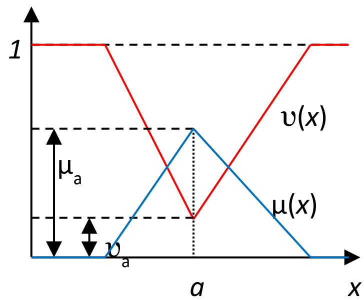

Fig. 1. Example triangular membership and non-membership functions of the IFS

As an extension of the IFS, the Pythagorean fuzzy set (PFS) originally proposed by Yager [14] was regarded to be more flexible in expressing vague and uncertain information due to its bigger domain, where the sum of the squared membership and non-membership degrees must be less than or equal to 1 (see Fig. 2).

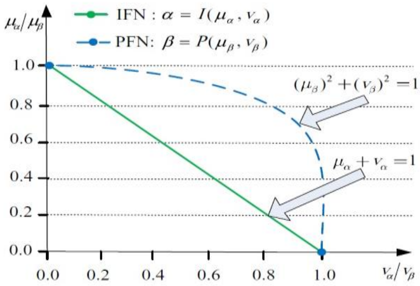

Fig. 2. The domain comparison of the IFS and PFS methods

Since then, more and more methods for PFSs have been developed and widely used in many areas such as image clustering and partitioning, pattern recognition, medical diagnosis and multi-criteria decision-making (MCDM). For instance, Yager [14] developed a useful decision method based on Pythagorean fuzzy aggregation operators to handle Pythagorean fuzzy MCDM problems. Zhang and Xu [18] provided the detailed mathematical expression for PFS and introduced the concept of Pythagorean fuzzy number (PFN). Meanwhile, they also developed a Pythagorean fuzzy TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) for handling the MCDM problem within PFNs. Peng and Yang [10] proposed the division and subtraction operations for PFNs. Also, they developed a Pythagorean fuzzy superiority and inferiority ranking method to solve multicriteria group decision-making problem with PFNs.

Several definitions of similarity measures between IFSs have been proposed recently. However, they have been proved to be unreasonable in some cases. For example, Li et al. [7] examined a comparative analysis of several existing similarity measures and pointed out their unreasonable cases in pattern recognition. Also, Papakostas et al. [9] studied the main properties of the existing distance and similarity measures for IFSs. Based on the transformation into the symmetric triangular fuzzy numbers, Zhang and Yu [16] presented a new similarity measure between IFSs. Similarly, Chen and Chang [2] presented a transformation-based similarity measure and adapted it to the pattern recognition problems. However, it has been shown to be also counterintuitive [8].

Inspired by these limitations, in this paper, we present a similarity measure for PFSs. The proposed measure is defined based on the amount and significance of information depicted in the PFSs. The rest of the paper is organized as follows. In Section II, basic concepts of PFSs are reviewed and some critical analyses of the existing methods for PFSs are conducted. In Section III, generalized knowledge measure, knowledge-based score function and similarity measure for PFSs with their axiomatic properties are introduced.

Section IV illustrates the advantages of the proposed generalized score function in differentiating and ranking the PFSs by some comparative numerical examples. Section V shows the effectiveness of the proposed similarity measure in applications for pattern recognition and medical diagnosis problems. Finally, Section VI provides concluding remarks.

2. Preliminaries

A. Pythagorean Fuzzy Sets

In 2013, R.R. Yager [14] provided two basic representations for PFS by polar coordinates and membership degrees, respectively. In the following, we recall the general definition of PFS, which is similar to the definition of IFS.

Definition 1. [1] A PFS $\boldsymbol{P}$ in a finite universe of discourse $\boldsymbol{X} = \{\boldsymbol{x_1}, \boldsymbol{x_2}, \dots, \boldsymbol{x_n}\}$ is an object with the following form:

$$\boldsymbol{P} = \{\langle \boldsymbol{x_i}, \boldsymbol{\mu_P}(\boldsymbol{x_i}), \boldsymbol{\nu_P}(\boldsymbol{x_i})\rangle | \boldsymbol{x_i} \in \boldsymbol{X}\}, \tag{1}$$

where $\boldsymbol{\mu_P}(\boldsymbol{x_i}), \boldsymbol{\nu_P}(\boldsymbol{x_i}) \in [\boldsymbol{0}, \boldsymbol{1}]$ denote the membership degree and the non-membership degree of an element $\boldsymbol{x_i}$ to PFS $\boldsymbol{P}$, respectively, such that $\boldsymbol{x_i} \in \boldsymbol{X}$,

$$\left(\boldsymbol{\mu_P}(\boldsymbol{x_i})\right)^2 + \left(\boldsymbol{\nu_P}(\boldsymbol{x_i})\right)^2 \leq \boldsymbol{1}\ . \tag{2}$$

The hesitancy degree of $\boldsymbol{x_i}$ to PFS $\boldsymbol{P}$ is given by:

$$\boldsymbol{\pi_P}(\boldsymbol{x_i}) = \sqrt{\boldsymbol{1} - \left(\boldsymbol{\mu_P}(\boldsymbol{x_i})\right)^2 - \left(\boldsymbol{\nu_P}(\boldsymbol{x_i})\right)^2}. \tag{3}$$

For convenience, we denote a single-valued PFS as Pythagorean fuzzy number (PFN) by $\boldsymbol{P} = \langle \boldsymbol{\mu_P}, \boldsymbol{\nu_P}\rangle$.

Afterwards, on the basis of relationship between IFS and PFS, Zhang and Xu [18] defined some operation for PFS as follows:

Definition 2. Let $\boldsymbol{P_1} = \langle \boldsymbol{\mu_{P_1}}, \boldsymbol{\nu_{P_1}}\rangle$ and $\boldsymbol{P_2} = \langle \boldsymbol{\mu_{P_2}}, \boldsymbol{\nu_{P_2}}\rangle$ be two PFNs and $\boldsymbol{\alpha} > \boldsymbol{0}$, then the operations on PFNs are defined in [18] as:

$$\boldsymbol{P_1} \oplus \boldsymbol{P_2} = \left\langle \sqrt{\left(\boldsymbol{\mu_{P_1}}\right)^2 + \left(\boldsymbol{\mu_{P_2}}\right)^2 - \left(\boldsymbol{\mu_{P_1}}\right)^2\left(\boldsymbol{\mu_{P_2}}\right)^2}, \boldsymbol{\nu_{P_1}}\boldsymbol{\nu_{P_1}} \right\rangle, \tag{4}$$

$$P_1 \otimes P_2 = \left\langle \mu_{P_1}\mu_{P_1}, \sqrt{\left(\nu_{P_1}\right)^2 + \left(\nu_{P_2}\right)^2 - \left(\nu_{P_1}\right)^2\left(\nu_{P_2}\right)^2} \right\rangle, \tag{5}$$

$$\alpha P_1 = \left\langle \sqrt{1 - \left(1 - \left(\mu_{P_1}\right)^2\right)^{\alpha}}, \left(\nu_{P_1}\right)^{\alpha} \right\rangle, \tag{6}$$

$$(P_1)^{\alpha} = \left\langle \left(\mu_{P_1}\right)^{\alpha}, \sqrt{1 - \left(1 - \left(\nu_{P_1}\right)^2\right)^{\alpha}} \right\rangle, \tag{7}$$

The operating results were proved to be PFNs as well.
They also utilized a relation between two PFNs defined in [14] as:

$$P_1 \geq P_2 \text{ iff } \mu_{P_1} \geq \mu_{P_2} \text{ and } \nu_{P_1} \leq \nu_{P_2}. \tag{8}$$

In order to compare the magnitude of PFNs, Zhang and Xu [18] defined the score function for PFNs applied as a score-based ranking method as follows:
Definition 3. [18] Let $P_1 = \langle \mu_{P_1}, \nu_{P_1} \rangle$ and $P_2 = \langle \mu_{P_2}, \nu_{P_2} \rangle$ be two PFNs, the score function of PFN is defined as follows:

$$s(P_1) = \left(\mu_{P_1}\right)^2 - \left(\nu_{P_1}\right)^2. \tag{9}$$

Then, a comparison law for PFNs is introduced as:

If $s(P_1) < s(P_2)$ then $P_1 \prec P_2$,
if $s(P_1) = s(P_2)$ then $P_1 \sim P_2$.

B. Critical Analyses of the Existing Measures for PFSs

Remark 1. There exits indistinguishable pairs of PFNs when using the score-based ranking method (9) for PFNs. In other words, in some situations, the score-based ranking method fails to compare the magnitude of PFNs.
Example 1. Let $P_1 = \langle \frac{\sqrt{5}}{3}, 2/3 \rangle$ and $P_2 = \langle \frac{2}{3}, \sqrt{3}/3 \rangle$ be two PFNs, according to Definition 3, we have $s(P_1) = s(P_2)$=1/9, then $P_1 \sim P_2$, meaning that the relation (9) cannot distinguish these two.
Remark 2. The score-based ranking method (9) for PFNs is not held under multiplication by a scalar, i.e. $P_1 \geq P_2$ does not necessarily imply $\alpha P_1 \geq \alpha P_2, \alpha > 0$. In other words, the operation (6) for PFNs is not monotone with respect to the ordering (9).
Example 2. Let $P_1 = \langle 0.5, 0.4 \rangle$, $P_2 = \langle 0.3, 0.2 \rangle$ be two PFNs and α=0.5, then using (6) we have $\alpha P_1 = \langle \sqrt{1 - (1 - 0.5^2)^{0.5}}, (0.4)^{0.5} \rangle = \langle 0.366, 0.632 \rangle$ and $\alpha P_2 = \langle \sqrt{1 - (1 - 0.3^2)^{0.5}}, (0.2)^{0.5} \rangle = \langle 0.214, 0.447 \rangle$ respectively. According to Definition 3, we have $s(P_1) = 0.09$, $s(P_2) = 0.05$, then $P_1 \succ P_2$, whereas $s(\alpha P_1) = -0.266$,

$s(\alpha P_2) = -0.153$, then $\alpha P_1 \prec \alpha P_2$. Thus, the operation (6) for PFNs is not monotone with respect to the ordering (9).

Remark 3. Similarly, the power operation (7) for PFNs is not monotone with respect to the ordering (9).

Example 3. Let $P_1 = \langle 0.4, 0.5\rangle$, $P_2 = \langle 0.2, 0.3\rangle$ be two PFNs and α=0.5, then using (7) we have $P_1{}^{\alpha} = \langle (0.4)^{0.5}, \sqrt{1-(1-0.5^2)^{0.5}}\rangle = \langle 0.632, 0.366\rangle$ and $P_2{}^{\alpha} = \langle (0.2)^{0.5}, \sqrt{1-(1-0.3^2)^{0.5}}\rangle = \langle 0.447, 0.214\rangle$ respectively. According to Definition 3, we have $s(P_1) = -0.09$, $s(P_2) = -0.05$, then $P_1 \prec P_2$, while $s(P_1{}^{\alpha}) = 0.266$, $s(P_2{}^{\alpha}) = 0.153$, then $P_1{}^{\alpha} \succ P_2{}^{\alpha}$. Thus, the operation (7) for PFNs is not monotone with respect to the ordering (9).

Recently, Bakioglu and Atahan [1] proposed a hybrid multi-criteria decision making approach under PFN environment using Pythagorean fuzzy weighted averaging (PFWA), which is defined as follows:

Definition 4. [1] Let $P_i = \langle \mu_{P_i}, \nu_{P_i}\rangle$ be a collection of PFNs and $w = (w_1, \dots, w_n)$ be its weight vector, i=(1, 2,…, n) with $w_i \in [0,1]$ and $\sum_{i=1}^{n} w_i = 1$, the PFWA is defined as:

$$PFWA(P_1, \dots, P_n) = \langle \left(1 - \prod_{i=1}^{n}\left(1-\mu_{P_i}{}^2\right)^{w_i}\right)^{1/2},\ \prod_{i=1}^{n}\left(\nu_{P_i}\right)^{w_i}\rangle. \quad (10)$$

The desirable, monotonicity property of the aggregation operator is defined as follows:

For any three PFNs A, B and C, if $A \succ B$ then $PFWA(A, C) \succ PFWA(B, C)$.

Remark 4. There are some unreasonable results like non-monotonicity provided by the operator (10) as shown below.

Example 4. Let $A = \langle 0.15, 0.3\rangle$, $B = \langle 0.4, 0.5\rangle$ and $C = \langle 0.15, 0.1\rangle$ be three PFNs. Since the score function $s(A) = -0.067$ and $s(B) = -0.09$, so we get $B \prec A$. With the weights $w_1 = w_2 = 0.5$, we aggregate B with C and A with C using (10). Respectively, we get $\mathbf{PFWA}(B, C) = \langle 0.306, 0.223\rangle$, $\mathbf{PFWA}(A, C) = \langle 0.15, 0.173\rangle$. Calculating their score function, we get $s(\mathbf{PFWA}(B, C)) = 0.044$, $s(\mathbf{PFWA}(A, C)) = -0.007$, then we have $\mathbf{PFWA}(B, C) \succ \mathbf{PFWA}(A, C)$ showing that the operator PFWA (10) is non-monotonic on the PFNs.

The another utilization of PFS to the decision making approach is generalized Pythagorean fuzzy weighted Bonferroni mean (GPFWBM) proposed by Zhang et al. [17]

Definition 5. Let s, t, r >0 and $P_i = \langle \mu_i, \nu_i\rangle$ be a collection of PFNs with their weights $w = (w_1, \dots, w_n)$, i=(1, 2,…, n) satisfying $w_i \in [0,1]$ and $\sum_{i=1}^{n} w_i = 1$. Then the GPFWBM is defined in [17] as:

$$GPFWBM^{s,t,r}(P_1, \dots, P_n) = \left\langle \begin{array}{c} \left(\sqrt{1 - \prod_{i,j,k=1}^{n}\left(1-\mu_i^{2s}\mu_j^{2t}\mu_k^{2r}\right)^{w_i w_j w_k}}\right)^{\frac{1}{(s+t+r)}}, \\ \sqrt{\left(1-\left(1-\prod_{i,j,k=1}^{n}\left(1-(1-\nu_i^2)^s\left(1-\nu_j^2\right)^t(1-\nu_k^2)^r\right)^{w_i w_j w_k}\right)\right)^{\frac{1}{(s+t+r)}}} \end{array} \right\rangle \quad (11)$$

Remark 5. It is easily seen that for any three arguments $P_i = \langle 1, 0\rangle$ of the operator (11), the aggregation result will be the same equalling $\langle 1, 0\rangle$ regardless of other arguments and their weights.

3. The similarity measure for PFSs

Apart from the common approach of similarity measures based on the frequently used distance measure, we propose a new similarity measure on the basis of the knowledge measure. The knowledge measure developed by Nguyen [8] depicts knowledge amount of information conveyed by an IFS (sum of the membership and non-membership degrees) and by its inherent fuzziness, which appears as variation between the membership and non-membership functions. Therefore, we propose a similarity measure based on the simpler knowledge measure of the PFSs as follows.

Definition 6. Let $PFS(X)$ denotes the family of all the PFSs over the universe of discourse $X = \{x_1, x_2, \dots, x_n\}$ and an PFS given by $A = \langle \mu_A(x_i), \nu_A(x_i)\rangle$. A mapping $K_N: PFS(X) \longrightarrow [0,1]$ is called a knowledge measure of A, if it satisfies the following properties: $\forall x_i \in X$,

(A1.1) $K_N(A) = 0$ iff $\mu_A(x_i) = \nu_A(x_i) = 0$;

(A1.2) $K_N(A) = 1$ iff A is a crisp set;

(A1.3) $K_N(A^c) = K_N(A)$, where A^c is a complement of A;

(A1.4) $K_N(A) \geq K_N(B)$ iff $\mu_A(x_i) \geq \mu_B(x_i)$ and $\nu_A(x_i) \geq \nu_B(x_i)$;

Theorem 1. For an PFS $A \in PFS(X)$, the function K_N defined by

$$K_N(A) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{2^{1/2}}\left\{\left[\left(\mu_A(x_i)\right)^2 + \left(\nu_A(x_i)\right)^2\right] + \left(\mu_A(x_i) - \nu_A(x_i)\right)^2\right\}^{1/2}, \qquad (12)$$

is a knowledge measure of A.

Proof:

The proof of (A1.1) to (A1.2) is straightforward from Definition 6.

(A1.3), $A^c = \langle \nu_A(x_i), \mu_A(x_i)\rangle$ and

$$K_N(A^c) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{2^{1/2}}\left\{\left[\left(\mu_A(x_i)\right)^2 + \left(\nu_A(x_i)\right)^2\right] + \left(\mu_A(x_i) - \nu_A(x_i)\right)^2\right\}^{1/2} = K_N(A);$$

(A.1.4) Given two PFSs A and B, we have

$$K_N(A) \geq K_N(B) \Leftrightarrow \left(\mu_A(x_i) + \nu_A(x_i)\right)^2 \geq \left(\mu_B(x_i) + \nu_B(x_i)\right)^2 \text{ and } \left(\mu_A(x_i) - \nu_A(x_i)\right)^2 \geq \left(\mu_B(x_i) - \nu_B(x_i)\right)^2$$

$$\Leftrightarrow 4\mu_A(x_i)\nu_A(x_i) \geq 4\mu_B(x_i)\nu_B(x_i) \Leftrightarrow \mu_A(x_i) \geq \mu_B(x_i) \text{ and } \nu_A(x_i) \geq \nu_B(x_i).$$

This completes the proof.

Due to such constructed knowledge measure, it cannot distinguish a PFV and its complement. To distinguish the significance between positive and negative information, a knowledge-based similarity measure is proposed as follows.

Definition 7. Let $A, B, C \in PFS(X)$ be PFVs with the same significance of information, i.e. $\mu_A(x) \geq \nu_A(x)$, $\mu_B(x) \geq \nu_B(x)$ and $\mu_C(x) \geq \nu_C(x)$ or inversely, a mapping

$Sim_N : PFS(X) \longrightarrow [0,1]$ is called a similarity measure between PFSs,----``` if it satisfies the following properties: $\forall x_i \in X$,

(A2.1) $Sim_N(A, B) = 1$ iff A=B,

(A2.2) $Sim_N(A, B) = 0$ iff $\pi_A = 1$ & B is a crisp set or A is a crisp set & $\pi_B = 1$,

(A2.3) $Sim_N(A, B) = Sim_N(B, A)$,

(A2.4) If $A \succcurlyeq B \succcurlyeq C$ then $Sim_N(A, C) \leq Sim_N(A, B)$ & $Sim_N(A, C) \leq Sim_N(B, C)$.

Theorem 2. For any $A, B \in PFS(X)$ with the same significance of information, i.e. $\mu_A(x) \geq \nu_A(x)$ and $\mu_B(x) \geq \nu_B(x)$ or inversely, the function defined as:

$$Sim_N(\mathrm{A}, \mathrm{B}) = 1 - |K_N(A) - K_N(B)| \tag{13}$$

is a similarity measure between A and B.

From Definition 6, it is easily seen that $0 \leq Sim_N(\mathrm{A}, \mathrm{B}) \leq 1, \forall x_i \in X$.

Proof:

(A2.1) We have from Definition 7:

$$Sim_N(A, B) = 1 \Leftrightarrow |K_N(A) - K_N(B)| = 0 \Leftrightarrow A = B.$$

(A2.2) Having $0 \leq S_N(A) \leq 1, \forall x \in X$, we obtain:

$Sim_N(A, B) = 0 \Leftrightarrow S_N(B) = 0$ & $S_N(A) = 1$ *or* $S_N(B) = 1$ & $S_N(A) = 0$.

Thus, $Sim_N(A, B) = 0 \Leftrightarrow A$ is a crisp set & $\pi_B = 1$ or $\pi_A = 1$ & B is a crisp set.

(A2.3) It is obvious from Definition 7.

(A2.4) From (A1.4) for $\mu_A \geq \nu_A$, $\mu_B \geq \nu_B$ and $\mu_C \geq \nu_C$ we have:

if $A \succcurlyeq B \succcurlyeq C$ then $K_N(A) \geq K_N(B) \geq K_N(C)$ which implies:

$$K_N(A) - K_N(C) \geq K_N(A) - K_N(B) \geq 0 \Rightarrow 1 - |K_N(A) - K_N(C)| \leq 1 - |K_N(A) - K_N(B)|$$

$\Rightarrow Sim_N(A, C) \leq Sim_N(A, B)$, and

$$K_N(A) - K_N(C) \geq K_N(B) - K_N(C) \geq 0 \Rightarrow 1 - |K_N(A) - K_N(C)| \leq 1 - |K_N(B) - K_N(C)|$$

$\Rightarrow Sim_N(A, C) \leq Sim_N(B, C)$.

This completes the proof.

4. Numerical examples

In this section, the performance of the proposed knowledge measure will be examined based on some numerical examples.

Example 5 (Continuing Ex. 1). Let $P_1 = \langle \frac{\sqrt{5}}{3}, 2/3 \rangle$ and $P_2 = \langle \frac{2}{3}, \sqrt{3}/3 \rangle$ be two PFNs, according to (12), we have $K_N(P_1) = 0.74$ and $K_N(P_2)$=0.67, then $P_1 \succ P_2$, whereas the relation (9) cannot distinguish these two.

Example 6 (Continuing Ex. 2). Let $P_1 = \langle 0.5, 0.4\rangle$, $P_2 = \langle 0.3, 0.2\rangle$ be two PFNs and α=0.5, then using (6) we have $\alpha P_1 = \langle\sqrt{1-(1-0.5^2)^{0.5}}, (0.4)^{0.5}\rangle = \langle 0.366, 0.632\rangle$ and $\alpha P_2 = \langle\sqrt{1-(1-0.3^2)^{0.5}}, (0.2)^{0.5}\rangle = \langle 0.214, 0.447\rangle$ respectively. According to (12) we have $K_N(P_1) = 0.5$, $K_N(P_2) = 0.3$, then $P_1 \succ P_2$, whereas $K_N(\alpha P_1) = 0.63$, $K_N(\alpha P_2) = 0.45$, then $\alpha P_1 \succ \alpha P_2$. Thus in this case, the operation (6) for PFNs is monotone with respect to the proposed knowledge measure.

Example 7 (Continuing Ex. 3). Let $\boldsymbol{P_1 = \langle 0.5, 0.4\rangle}$, $\boldsymbol{P_2 = \langle 0.3, 0.2\rangle}$ be two PFNs and α=0.5, then using (7) we have $\boldsymbol{P_1{}^{\alpha} = \langle (0.5)^{0.5}, \sqrt{1-(1-0.4^2)^{0.5}}\rangle = \langle 0.707, 0.289\rangle}$ and $\boldsymbol{P_2{}^{\alpha} = \langle (0.2)^{0.5}, \sqrt{1-(1-0.3^2)^{0.5}}\rangle = \langle 0.547, 0.142\rangle}$ respectively. According to (12), we have $\boldsymbol{K_N(P_1) = 0.5}$, $\boldsymbol{K_N(P_2) = 0.3}$, then $\boldsymbol{P_1 \succ P_2}$, while $\boldsymbol{K_N(\alpha P_1) = 0.71}$, $\boldsymbol{K_N(\alpha P_2) = 0.54}$, then $\boldsymbol{\alpha P_1 \succ \alpha P_2}$. Thus in this case, the operation (7) for PFNs is monotone with respect to the proposed knowledge measure.

5. Application

In this section, the proposed similarity measure is employed for the fault diagnosis of steam turbine generator unit under Pythagorean fuzzy environment. The vibration of the steam turbine generator unit suffers the influence of varying factors, such as mechanical load, vacuum degree, fluctuation of network load, the temperature of lubricant oil and defects of mechanical structure. Interaction effects of these factors result in the vibration of the generator unit. Ten fault types in rotating machines are established as failure patterns, i.e. P_1- unbalance, P_2- pneumatic force couple, P_3- offset center, P_4- oil-membrane oscillation, P_5- radial impact friction of rotor, P_6- symbiosis looseness, P_7- damage of antithrust bearing, P_8- surge, P_9- looseness of bearing block and P_{10}- non-uniform bearing stiffness. The knowledge of fault types and symptoms in the vibration frequencies of the turbine is adopted from [12] by transforming the vague values into the PFVs. The vibration frequency of turbine is divided into 9 different frequency ranges, in which the failure patterns are represented by PFVs, as shown in Table 1.

Table 1

The knowledge of system fault pattern [12]

	Frequency range (f- operating frequency)								
Fault patterns	0.01-0.39f	0.40-0.49f	0.50f	0.51-0.99f	f	2f	3-5f	Odd times of f	High freq >5f
P1- Unbalance	⟨0,1⟩	⟨0,1⟩	⟨0,1⟩	⟨0,1⟩	⟨0.85, 0⟩	⟨0.04, 0.94⟩	⟨0.04, 0.93⟩	⟨0,1⟩	⟨0,1⟩
P2- Pneumatic force couple	⟨0,1⟩	⟨0.28, 0.69⟩	⟨0.09, 0.88⟩	⟨0.55, 0.3⟩	⟨0,1⟩	⟨0,1⟩	⟨0,1⟩	⟨0,1⟩	⟨0.08, 0.83⟩

P3- Offset center	⟨0,1⟩	⟨0,1⟩	⟨0,1⟩	⟨0,1⟩	⟨0.3, 0.42⟩	⟨0.4, 0.58⟩	⟨0.08, 0.83⟩	⟨0,1⟩	⟨0,1⟩
P4- Oil-membrane oscillation	⟨0.09, 0.11⟩	⟨0.78, 0.18⟩	⟨0,1⟩	⟨0.08, 0.89⟩	⟨0,1⟩	⟨0,1⟩	⟨0,1⟩	⟨0,1⟩	⟨0,1⟩
P5- Radial impact friction of rotor	⟨0.09, 0.88⟩	⟨0.09, 0.11⟩	⟨0.08, 0.88⟩	⟨0.09, 0.88⟩	⟨0.18, 0.79⟩	⟨0.08, 0.83⟩	⟨0.08, 0.83⟩	⟨0.08, 0.88⟩	⟨0.08, 0.88⟩
P6- Symbiosis looseness	⟨0,1⟩	⟨0,1⟩	⟨0,1⟩	⟨0,1⟩	⟨0.18, 0.78⟩	⟨0.12, 0.83⟩	⟨0.37 0.55⟩	⟨0,1⟩	⟨0.22, 0.72⟩
P7- Damage of antithrust bearing	⟨0,1⟩	⟨0,1⟩	⟨0.08, 0.88⟩	⟨0.86, 0.07⟩	⟨0,1⟩	⟨0,1⟩	⟨0,1⟩	⟨0,1⟩	⟨0,1⟩
P8- Surge	⟨0,1⟩	⟨0.27, 0.68⟩	⟨0.08, 0.88⟩	⟨0.54, 0.38⟩	⟨0,1⟩	⟨0,1⟩	⟨0,1⟩	⟨0,1⟩	⟨0,1⟩
P9- Looseness of bearing block	⟨0.85, 0.07⟩	⟨0,1⟩	⟨0,1⟩	⟨0,1⟩	⟨0,1⟩	⟨0,1⟩	⟨0,1⟩	⟨0.08, 0.88⟩	⟨0,1⟩
P10- Non-uniform bearing stiffness	⟨0,1⟩	⟨0,1⟩	⟨0,1⟩	⟨0,1⟩	⟨0,1⟩	⟨0.77, 0.17⟩	⟨0.19, 0.77⟩	⟨0,1⟩	⟨0,1⟩

Suppose that there are two fault-testing samples A and B expressed in PFVs as shown in Table 2.

Table 2

The characteristic of the fault-testing samples [12]

Fault-testing samples	Frequency range (f- operating frequency)								
	0.01-0.39f	0.40-0.49f	0.50f	0.51-0.99f	F	2f	3-5f	Odd times of f	High freq >5f
A	⟨0,1⟩	⟨0,1⟩	⟨0.1, 0.9⟩	⟨0.9, 0.1⟩	⟨0,1⟩	⟨0,1⟩	⟨0,1⟩	⟨0,1⟩	⟨0,1⟩
B	⟨0.39, 0.56⟩	⟨0.07, 0.88⟩	⟨0,0.95⟩	⟨0.06, 0.89⟩	⟨0,0.95⟩	⟨0.13, 0.82⟩	⟨0,0.95⟩	⟨0,0.95⟩	⟨0.35, 0.60⟩

Our goal in the fault diagnosis analysis is to classify the fault-testing samples into one of the known fault patterns $P_j, (j = 1,2,..,10)$, adopting the proposed similarity measure $\mathrm{Sim}_{\mathrm{N}}$ (13). The calculation results of similarity measures between the fault-testing samples and the known fault patterns are summarized in rows of Table 3.

Table 3

Results of similarity measures between the fault-testing samples and the known patterns

Fault-testing samples	Similarity measures									
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
A	0.997	0.893	0.877	0.967	0.859	0.878	0.996	0.912	0.995	0.965
B	0.846	0.949	0.966	0.876	0.984	0.964	0.847	0.931	0.848	0.877

According to the principle of maximum similarity measure, we can decide that the fault-testing sample A is most similar to the known fault pattern P_7- damage of antithrust bearing, consistent with the results obtained in [12]12. In the same manner, we derive that the fault-testing sample B is most similar to the known fault pattern P_5- radial impact friction of rotor, which is also consistent with the results obtained in [12].

6. Conclusions

The paper discusses some limitations of the existing methods for PFSs. Also, a novel similarity measure based on knowledge measure for PFSs has been proposed, which incorporates the knowledge conveyed by PFSs and significance related to positive and negative information to distinguish their meanings. Thanks to that, it can overcome the drawbacks of the existing methods. The superiority of the proposed method in dealing with uncertain information has been shown in numerical examples. The effectiveness and applicability of the proposed method have been demonstrated in applications of fault diagnosis of the turbine generator. The results obtained by the proposed methods were consistent with that of the existing methods, which though have shown to be failing in some other cases. The limitation of the paper is the lack of experimental verification by own real-life applications or statistical test. Thus, future work will be focused on these issues.

7. References

1. Bakioglu G., Atahan A.O.: AHP integrated TOPSIS and VIKOR methods with Pythagorean fuzzy sets to prioritize risks in self-driving vehicles. Appl. Soft Comput. 2021, 99, 106948.
2. Chen S.M. and Chang C.H.: A novel similarity between Atanassov's intuitionistic fuzzy sets based on transformation techniques with applications to pattern recognition, Information Sci., 291, 96-114, 2015.

3. Hoang, D.T., Kang, H.J.: A motor current signal-based bearing fault diagnosis using deep learning and information fusion. IEEE Trans. Instrum. Meas. 2019, 69, 3325–3333.
4. Hoang D.T., Tran X.T., Van, M., Kang H.J.: A Deep Neural Network-Based Feature Fusion for Bearing Fault Diagnosis. Sensors, 2021, 21, 244.
5. Jiao J., Zhao M., Lin J., Ding C.: Deep Coupled Dense Convolutional Network With Complementary Data for Intelligent Fault Diagnosis, IEEE Trans. Ind. Electron., 66 (12) (2019) 9858-9867.
6. Li C., Sanchez V., Zurita G., Lozada M.C., Cabrera D.: Rolling element bearing defect detection using the generalized synchrosqueezing transform guided by time–frequency ridge enhancement. ISA Trans. 2016, 60, 274–284.
7. Li Y.H., Olson D.L. and Qin D.: Similarity measures between intuitionistic fuzzy (vague) sets: A comparative analysis. Pattern Recognition Letters, vol. 28, 278-285, 2007.
8. Nguyen H.: A new knowledge-based measure for intuitionistic fuzzy sets and its application in multiple attribute group decision making, Expert. Syst. Appl., vol. 42, no. 22, pp. 8766–8774, 2015.
9. Papakostas G.A., Hatzimichailidis A.G. and Kaburlasos V.G.: Distance and similarity measures between intuitionistic fuzzy sets: A comparative analysis from a pattern recognition point of view, Pattern Recognit. Lett. 34, 1609-1622, 2013.
10. Peng X., Yang Y.: Some results for pythagorean Fuzzy Sets, Int. J. Intell. Syst. 30 (2015) 1133–1160.
11. Rauber T.W., de Assis Boldt F., Varejão F.M.: Heterogeneous feature models and feature selection applied to bearing fault diagnosis. IEEE Trans. Ind. Electron. 2014, 62, 637–646.
12. Shi L.L., Ye J.: Study on fault diagnosis of turbine using an improved cosine similarity measure of vague sets. Journal of Applied Sciences 13 (10), 1781-1786 (2013).
13. Xu G., Hou D., Qi H. and Bo L.: High-speed train wheel set bearing fault diagnosis and prognostics: A new prognostic model based on extendable useful life. Mech. Syst. Signal. Proc. 2021, 146, 107050.
14. Yager R.R.: Pythagorean fuzzy subsets, in: IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS), IEEE, 2013, pp. 57–61.
15. Ye Z., Yu J.: Deep morphological convolutional network for feature learning of vibration signals and its applications to gearbox fault diagnosis. Mech. Syst. Signal Process. 2021, 161, 107984.
16. Zhang H. and Yu L.: New distance measures between intuitionistic fuzzy sets and interval-valued fuzzy sets. Inform. Sci., 245, 181-196, 2013.
17. Zhang R., Wang J., Zhu X., Xia M. and Yu Y.: Some Generalized Pythagorean Fuzzy Bonferroni Mean Aggregation Operators with Their Application to Multiattribute Group Decision Making, Complexity, vol. 2017, Article ID 5937376, 16 pages, 2017.
18. Zhang X., Xu Z.: Extension of TOPSIS to multiple criteria decision making with Pythagorean fuzzy sets, Int. J. Intell. Syst. 29 (12) (2014) 1061–1078.