ISSN: 2029-4913 / eISSN: 2029-4921

Article in press

TECHNOLOGICAL and ECONOMIC DEVELOPMENT of ECONOMY

https://doi.org/10.3846/tede.2025.23597

CLASSIFICATION AND IDENTIFICATION OF MEDICAL INSURANCE FRAUD: A CASE-BASED REASONING APPROACH

Xiaodi LIU[1,2], Shasha ZHANG[1,3], Zengwen WANG[4✉], Shitao ZHANG[1,2]

[1] *School of Microelectronics and Data Science, Anhui University of Technology, Ma'anshan, 243002 Anhui, China*
[2] *Anhui Provincial Joint Key Laboratory of Disciplines for Industrial Big Data Analysis and Intelligent Decision, Ma'anshan, 243002 Anhui, China*
[3] *Faculty of Engineering, Anhui Sanlian University, 230601 Hefei, China*
[4] *Researching Center of Social Security, Wuhan University, Wuhan, 430072 Hubei, China*

Article History:
- received 1 July 2023
- accepted 14 May 2024
- first published online 15 July 2025

Abstract. Appropriate classification of medical insurance fraud events can not only be effective in preventing and combating fraud, but also greatly improve the utilization of medical resources. Due to the uncertainty inherent in medical insurance fraud, identifying and classifying the fraud are non-trivial tasks. In addition, the selection of classification radius by traditional methods is often highly subjective. To this end, a case-based reasoning (CBR) approach in probabilistic hesitant fuzzy environment and its application to classifying the severity of medical insurance fraud events are investigated in this article. At first, the probabilistic hesitant fuzzy element (PHFE) is regarded as a discrete probability distribution, and its distribution function is defined. On this basis, a distribution discrepancy degree is proposed to make up for the shortage of existing measures between PHFEs. Then, a probabilistic hesitant fuzzy decision-making method based on CBR is proposed, which considers both decision data and the expert's own knowledge and experience. Finally, the proposed method is used to classify the severity of medical insurance fraud events, and the rationality and superiority of the method are verified by comparative analysis.

Keywords: medical insurance fraud, probabilistic hesitant fuzzy element, distribution function, distribution discrepancy degree, case-based reasoning.

JEL Classification: C44, D81, H55, I13.

✉Corresponding author. E-mail: ***wzwnjing@163.com***

1. Introduction

With the continuous improvement and development of the social medical insurance system, the number of insured personnel is also increasing. At the same time, the frequent occurrence of medical insurance fraud has not only caused the loss of medical insurance funds, but also threatened the safety of the entire social medical management institutions, which has brought serious challenges to the anti-fraud work of medical insurance. The safety of the medical insurance fund is related to the health and welfare of the people. Ensuring and maintaining the safety of the medical insurance fund require not only the strong supervision of the national regulators, but also the joint participation of the insured, medical institutions, designated pharmacies and other entities. In order to prevent and control the occurrence and spread of fraud in a timely manner, scholars from all countries have also carried out a series of studies on the detection and identification of medical insurance fraud. Settipalli and Gangadharan (2023) developed a weighted multi-tree method to analyze medical providers and detect medical fraudulent providers according to the detailed information of provider types. Jiang et al. (2021) improved the LDOF algorithm by using the proximity matrix of isolated forest for drug reselling in medical insurance fraud, and proposed a medical insurance drug anomaly detection algorithm. Li et al. (2022) constructed a theoretical model of medical insurance fraud identification, and characterized the fraud judgment variables from a three-dimensional perspective, i.e., time, quantity and cost. Kapadiya et al. (2022) proposed a security intelligent system based on blockchain and artificial intelligence to detect medical insurance fraud. Zhang et al. (2022) proposed a medical insurance fraud identification framework based on deep learning and consortium blockchain to improve the efficiency of medical insurance fraud detection. Most of the existing medical insurance fraud identification methods focus on identifying whether the behavior is fraudulent or not. In fact, after the occurrence of fraud, the measures of prevention and punishment are different according to the severity of fraud. It is extremely important for the stable development of the entire medical insurance institutions to formulate reasonable and effective prevention and control measures. Therefore, it is of great significance to classify the severity of medical insurance fraud.

On the other hand, because of the complexity and diversification of fraud means, there are often some risky and uncertain factors in the process of identifying fraudulent behaviors, and traditional methods fail to characterize the uncertain evaluations of decision makers. Hence, new tools are required to deal with the uncertainty and randomness (Han et al., 2022; Jiang et al., 2023; Zhu et al., 2023a). In 1965, Zadeh proposed the concept of fuzzy sets, which is an effective technique in handling uncertainty. Subsequently, many scholars expanded fuzzy sets into intuitionistic fuzzy sets (Atanassov, 1986), interval fuzzy sets (Gorzałczany, 1987), hesitant fuzzy sets (HFSs) (Torra, 2010) and probabilistic hesitant fuzzy sets (PHFS) (Zhu, 2014) to adapt to different decision-making environments. Among them, the HFS proposed by Torra (2010) is composed of hesitant fuzzy elements (HFE). It allows several possible values in [0, 1] to represent the membership degree, which can well describe the preferences of different decision makers (Liu et al., 2021; Zhan et al., 2023). However, with the progress of research, HFS gradually shows its drawbacks (Han & Zhan, 2023; Zhu et al., 2023b). For example, three decision makers give 0.3 as the membership degree of an alternative with respect to an attribute and one decision maker gives 0.6. The evaluation result can be represented by HFE $\{0.3, 0.6\}$. At this time, the occurrence probability of 0.3 is the same as that of 0.6, which is obviously inconsistent with the reality. In order to describe the preferences of decision-makers accurately, Zhu (2014) proposed the concept of PHFS, which consists of probabilistic hesitant fuzzy element (PHFE), i.e., the membership degree and its occurrence probability. If a PHFE is used in the above example, it can be expressed as $\{0.3(0.75), 0.6(0.25)\}$. Therefore, PHFS can more fully express decision information and is widely used in decision analysis (Liao et al., 2022; Jiang et al., 2024), emergency management (Liu et al., 2022), venture capital (Wu et al., 2021), and other fields (Cao et al., 2020).

Distance, similarity and correlation coefficient are important information measures between PHFSs (Zhang et al., 2023). Song and Chen (2021) proposed a generalized distance measure from the perspectives of membership degree, probability and length. Divsalar et al.

(2022) defined a probabilistic hesitant fuzzy distance based on the geometric score function and geometric variance function of PHFEs. Sha et al. (2021) extended Lance distance to the probabilistic hesitant fuzzy environment and defined the probabilistic hesitant fuzzy Lance distance. Wang et al. (2022) defined a hesitant degree formula, based on which a new similarity degree between PHFEs was proposed. Wang and Li (2017) introduced the northwest corner rule to obtain the expected mean value of two PHFEs, and presented the correlation coefficient between PHFEs on this basis. Song et al. (2019) gave a new probabilistic hesitant fuzzy correlation coefficient and applied it to clustering analysis and risk assessment. Although the above information measures have been applied in several fields, they all have corresponding shortcomings. The distances proposed have the following drawbacks: (1) They dissatisfy the axiomatic definition of distance (Song & Chen, 2021; Divsalar et al., 2022; Sha et al., 2021). (2) The PHFEs with different numbers of elements cannot be calculated directly (Song & Chen, 2021; Sha et al., 2021). The similarity formula defined by Wang et al. (2022) is essentially obtained by expanding on the basis of distance, which still suffers from the aforementioned defects. The correlation coefficients are essentially a kind of mean correlation coefficient, which do not take into account the probability distribution of membership degrees (Wang & Li, 2017; Song et al., 2019). In other words, as long as the means of the corresponding PHFEs are the same, the results derived by the correlation coefficients are equal, which is prone to counter-intuitive situations. Therefore, it is crucial to establish a more ideal information measure.

The probabilistic hesitant fuzzy multi-attribute decision making has attracted extensive attention from research scholars. Krishankumar et al. (2021) extended the VIKOR method to the probabilistic hesitant fuzzy environment and provided a new ranking method. Naeem et al. (2021) proposed an improved TOPSIS decision-making method. Wu et al. (2021) constructed a multi-attribute decision making method based on generalized probabilistic hesitant fuzzy Bonferroni mean operator. Most of the existing probabilistic hesitant fuzzy decision methods are based on the current decision information to determine the attribute weights and rank the alternatives. However, for some complex decision-making problems, decision makers need to rely on past successful cases to assist decision-making for deriving more reliable results. The CBR approach takes into account both the current decision information and the successful experience of experts on previous decision cases, which achieves the combination of subjectivity and objectivity in the decision-making process (Schank, 1983). The process of CBR involves five steps (Fan et al., 2014; Liao et al., 1998), which is shown in Figure 1. Li and Wei (2018) defined a new distance between hesitant fuzzy linguistic term sets and put forward a hesitant fuzzy linguistic decision model based on CBR. Wang et al. (2020) constructed different similarity algorithms for precise numerical, fuzzy semantic and symbolic data and raised an emergency decision-making model based on CBR. Li et al. (2020) combined decision data and expert experience to give a CBR-based method for ranking and clustering probabilistic linguistic information. Löw et al. (2019) put forward a CBR framework for handling incomplete databases. A dynamic case retrieval method by combining subjective and objective information to facilitate emergency decision-making (Zheng et al., 2018). Xu et al. (2019) utilized similar events to deduct the severity level of current emergencies. Thus, historical cases have proven useful in supporting emergency decision-making (Yu et al., 2018). However, there

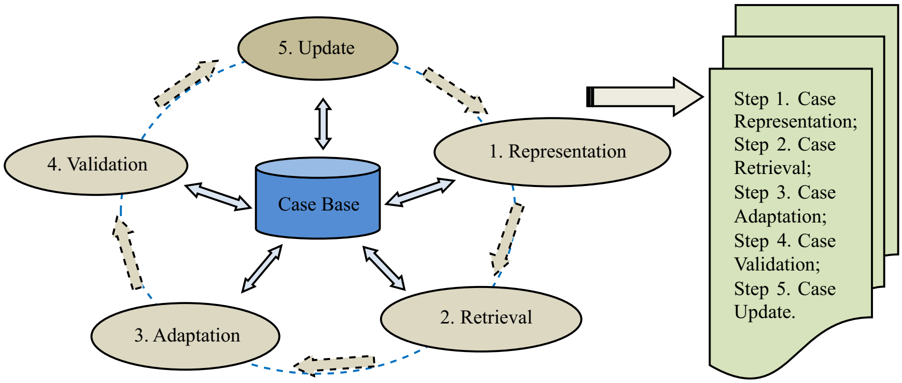

Figure 1. Decision process by CBR

are still no reports on the decision method based on CBR in the probabilistic hesitant fuzzy environment. To do so, a new CBR method within the context of PHFSs is presented in this paper and applied to the classification of medical insurance fraud.

As discussed, much effort has been devoted to fuzzy decision-making and medical insurance fraud, but it still suffers from the challenges below.

(1) Some information measures between PHFSs are put forward, such as distance, similarity and correlation coefficient, and they are widely used in decision-making and pattern recognition. However, some defects still remain, and it is of great urgency to construct a new information measure between PHFSs.

(2) As an analogical reasoning method, CBR has been expanded to accommodate different types of fuzzy sets, but little research has been conducted on PHFSs-based CBR, which limits its further application.

(3) Classifying the events of medical insurance fraud is useful in preventing and combating fraud. Hence, different measures should be taken according to the severity of fraud. Much effort has been dedicated to identifying whether the behavior is fraudulent or not, but research on classifying the severity of fraud is very little.

This paper analyzes the shortcomings of existing information measures between PHFSs, and the main innovations and contributions are as follows:

(1) The probability distribution function of PHFE is defined, based on which the distribution discrepancy degree between PHFEs is proposed. It remedies the shortcomings of existing probabilistic hesitant fuzzy information measures.

(2) A new case-based decision model is developed by extending the CBR method to probabilistic hesitant fuzzy environment for the first time. It takes into account both decision data and the expert's own knowledge and experience, and is able to classify alternatives into different categories.

(3) Taking the medical insurance fraud as the background, this paper provides a method to divide the severity of fraud and provides a basis for formulating corresponding prevention and combating means.

The rest of this article is organized as follows. Section 2 reviews relevant definitions and concepts. Section 3 defines the probabilistic hesitant fuzzy distribution function and the distribution discrepancy degree, and proposes a CBR classification method based on the distribution discrepancy degree. Section 4 gives a numerical example concerning social medical insurance fraud. Section 5 illustrates the effectiveness and superiority of the proposed method through comparison with other methods. Section 6 offers a discussion. Some conclusions are given in Section 7.

2. Preliminaries

2.1. PHFS

Definition 1 (Zhu, 2014; Xu & Zhou, 2017). Given any non-empty set *X*, the PHFS *A* defined on the set *X* can be expressed as:

$$A=\left\{\left\langle x,h(p)\right\rangle \middle| x\in X\right\}. \tag{1}$$

Here, $h(p)=\left\{\gamma_i \middle| p_i, i=1,2,\ldots,l\right\}$ is called a PHFE. $\gamma_i\in[0,1]$ is the membership degree, which indicates the degree of possibility that the element $x\in X$ belongs to *A*. $p_i\in[0,1]$ denotes the occurrence probability of the corresponding membership degree γ_i and $\sum_{i=1}^{l}p_i=1$. l is the number of elements in $h(p)$.

Definition 2 (Xu & Zhou, 2017). Let $h(p)$, $h_1(p)$ and $h_2(p)$ be three PHFEs, $\lambda>0$, then:

(1) $h^c(p)=\cup_{\gamma_i\in h(p)}\left\{[1-\gamma_i](p_i)\right\}$;

(2) $\alpha h(p)=\cup_{\gamma_i\in h(p)}\left\{[1-(1-\gamma_i)^\alpha](p_i)\right\}$;

(3) $h^\alpha(p)=\cup_{\gamma_i\in h(p)}\left\{[(\gamma_i)^\alpha](p_i)\right\}$;

(4) $h_1(p)\oplus h_2(p)=\cup_{\gamma_i^1\in h_1(p),\gamma_j^2\in h_2(p)}\left\{[\gamma_i^1+\gamma_j^2-\gamma_i^1\gamma_j^2](p_i^1p_j^2)\right\}$;

(5) $h_1(p)\otimes h_2(p)=\cup_{\gamma_i^1\in h_1(p),\gamma_j^2\in h_2(p)}\left\{[\gamma_i^1\gamma_j^2](p_i^1p_j^2)\right\}$.

Definition 3 (Xu & Zhou, 2017). *h*(*p*) is a PHFE, then its score function is:

$$s(h(p))=\sum_{i=1}^{l}\gamma_i p_i. \tag{2}$$

2.2. Distribution function of discrete random variable

Definition 4 (Loève, 2017). Let *X* be a discrete random variable, and all its possible values are $x_1,x_2,\cdots,x_n$. Suppose that the probability of *X* taking x_k is:

$$P(X=x_k)=p_k\geq 0, \tag{3}$$

and $\sum_{k=1}^{n}p_k=1$, then the set of probabilities $\left\{p_1,p_2,\cdots,p_n\right\}$ is called the distribution sequence of the random variable *X*. By the probability additivity, the distribution function of *X* is as below:

$$F(x)=P(X\leq x)=\sum_{X\leq x_k}p_k. \tag{4}$$

3. A new classification method for medical insurance fraud

3.1. Probabilistic hesitant fuzzy distribution function

Based on Definition 4, the distribution function in the probabilistic hesitant fuzzy environment is given in this Section.

Definition 5. Let $h(p)=\left\{\gamma_1(p_1),\gamma_2(p_2),\gamma_3(p_3),\cdots,\gamma_n(p_n)\right\}$ be a PHFE, $X=(\gamma_1,\gamma_2,\cdots,\gamma_n)$ be a random variable, and the distribution sequence of X be $P(X=\gamma_k)=p_k,(k=1,2,\cdots,n)$. Here, $\gamma_1<\gamma_2<\cdots<\gamma_n$, and p_i is the occurrence probability of γ_i, then the distribution function of $h(p)$ is defined as:

$$F(\gamma)=P(X\leq\gamma)=\begin{cases}0 & 0\leq\gamma<\gamma_1\\ p_1 & \gamma_1\leq\gamma<\gamma_2\\ p_1+p_2 & \gamma_2\leq\gamma<\gamma_3\\ & \vdots\\ \sum_{i=1}^{n-1}p_i & \gamma_{n-1}\leq\gamma<\gamma_n\\ \sum_{i=1}^{n}p_i=1 & \gamma_n\leq\gamma\leq 1\end{cases}. \tag{5}$$

3.2. Distribution discrepancy degree

The distribution function of PHFE reflects its probability distribution, and based on it, this section defines an information measure to characterize the difference between two PHFEs, i.e., the distribution discrepancy degree.

Definition 6. Let $h_1(p)=\left\{\gamma_1^1(p_1^1),\gamma_2^1(p_2^1),\gamma_3^1(p_3^1),\cdots,\gamma_n^1(p_n^1)\right\}$ and $h_2(p)=\left\{\gamma_1^2(p_1^2),\gamma_2^2(p_2^2),\gamma_3^2(p_3^2),\cdots,\gamma_n^2(p_n^2)\right\}$ be two PHFEs, and their distribution functions are $F_1(\gamma)$ and $F_2(\gamma)$ respectively. Then the distribution discrepancy degree between them is defined as:

$$D(h_1(p),h_2(p))=\int_0^1\left|F_1(\gamma)-F_2(\gamma)\right|d\gamma. \tag{6}$$

For the distribution discrepancy degree $D(h_1(p),\ h_2(p))$, it is easy to verify that it satisfies the following properties.

Property 1. Let $h_1(p)$ and $h_2(p)$ be two PHFEs, then their distribution discrepancy degree satisfies the following conditions.

(1) $0\leq D(h_1(p),h_2(p))\leq 1$;

(2) $D(h_1(p),h_2(p))=D(h_2(p),h_1(p))$;

(3) $D(h_1(p),h_2(p))=0\Leftrightarrow h_1(p)=h_2(p)$;

(4) $D(h_1(p),h_2^c(p))=D(h_1^c(p),h_2(p))$.

Proof. (1) From $0\leq\left|F_1(\gamma)-F_2(\gamma)\right|\leq 1$, we can get $0=\int_0^1 0\,d\gamma\leq\int_0^1\left|F_1(\gamma)-F_2(\gamma)\right|d\gamma\leq\int_0^1 1\,d\gamma=1$, so it is proved.

(2) Since $D(h_1(p),h_2(p))=\int_0^1 \left|F_1(\gamma)-F_2(\gamma)\right| d\gamma=\int_0^1 \left|F_2(\gamma)-F_1(\gamma)\right| d\gamma=D(h_2(p),h_1(p))$, it is proved.

(3) if $h_1(p) = h_2(p)$, then their distribution functions are the same, i.e., $F_1(\gamma) = F_2(\gamma)$ and thus $D(h_1(p),h_2(p))=0$. On the other hand, if $D(h_1(p),h_2(p))=0$, there is $\int_0^1 \left|F_1(\gamma)-F_2(\gamma)\right| d\gamma=0$, which is easy to get $F_1(\gamma) = F_2(\gamma)$. It is noted that the distribution sequence of discrete random variable and its distribution function are uniquely determined by each other, that is, the distribution based on the membership degree and its probability in PHFE $h_1(p)$ is the same as that in PHFE $h_2(p)$, therefore $h_1(p) = h_2(p)$.

(4) Assume that $F_i(\gamma)$ denotes the distribution function for $h_i(p)\left(i=1,2\right)$.

$$F_1(\gamma)=\begin{cases} 0 & 0\le\gamma<\gamma_1^1 \\ p_1^1=t_1 & \gamma_1^1\le\gamma<\gamma_2^1 \\ p_1^1+p_2^1=t_2 & \gamma_2^1\le\gamma<\gamma_3^1 \\ \vdots & \\ \sum_{i=1}^{n-2}p_i^1=t_{n-2} & \gamma_{n-2}^1\le\gamma<\gamma_{n-1}^1 \\ \sum_{i=1}^{n-1}p_i^1=t_{n-1} & \gamma_{n-1}^1\le\gamma<\gamma_n^1 \\ 1 & \gamma_n^1\le\gamma\le1 \end{cases},\quad F_2(\gamma)=\begin{cases} 0 & 0\le\gamma<\gamma_1^2 \\ p_1^2=m_1 & \gamma_1^2\le\gamma<\gamma_2^2 \\ p_1^2+p_2^2=m_2 & \gamma_2^2\le\gamma<\gamma_3^2 \\ \vdots & \\ \sum_{i=1}^{n-2}p_i^2=m_{n-2} & \gamma_{n-2}^2\le\gamma<\gamma_{n-1}^2 \\ \sum_{i=1}^{n-1}p_i^2=m_{n-1} & \gamma_{n-1}^2\le\gamma<\gamma_n^2 \\ 1 & \gamma_n^2\le\gamma\le1 \end{cases}.$$

Then the distribution function $\tilde{F}_i(\gamma)$ for $h_i^c(p)\left(i=1,2\right)$ can be expressed as follows:

$$\tilde{F}_1(\gamma)=\begin{cases} 0 & 0\le\gamma<1-\gamma_n^1 \\ 1-t_{n-1} & 1-\gamma_n^1\le\gamma<1-\gamma_{n-1}^1 \\ 1-t_{n-2} & 1-\gamma_{n-1}^1\le\gamma<1-\gamma_{n-2}^1 \\ \vdots & \\ 1-t_2 & 1-\gamma_3^1\le\gamma<1-\gamma_2^1 \\ 1-t_1 & 1-\gamma_2^1\le\gamma<1-\gamma_1^1 \\ 1 & 1-\gamma_1^1\le\gamma\le1 \end{cases},\quad \tilde{F}_2(\gamma)=\begin{cases} 0 & 0\le\gamma<1-\gamma_n^2 \\ 1-m_{n-1} & 1-\gamma_n^2\le\gamma<1-\gamma_{n-1}^2 \\ 1-m_{n-2} & 1-\gamma_{n-1}^2\le\gamma<1-\gamma_{n-2}^2 \\ \vdots & \\ 1-m_2 & 1-\gamma_3^2\le\gamma<1-\gamma_2^2 \\ 1-m_1 & 1-\gamma_2^2\le\gamma<1-\gamma_1^2 \\ 1 & 1-\gamma_1^2\le\gamma\le1 \end{cases}.$$

From Figure 2 and Figure 3, we can find that the subsection integration intervals for $D(h_1(p),h_2^c(p))$ and $D(h_2(p),h_1^c(p))$ are of the same length and the integrand functions are always the same. The area of the shaded region in Figure 2 shows the value of the distribution discrepancy degree between $h_1(p)$ and $h_2^c(p)$, and the area of the shaded region in Figure 3 shows the value of the distribution discrepancy degree between $h_2(p)$ and $h_1^c(p)$.

$$D(h_1(p),h_2^c(p))=\int_{1-\gamma_n^2}^{\gamma_1^1}|1-m_{n-1}|\,d\gamma+\int_{\gamma_1^1}^{1-\gamma_{n-1}^2}|1-m_{n-1}-t_1|\,d\gamma+\int_{1-\gamma_{n-1}^2}^{\gamma_2^1}|1-m_{n-2}-t_1|\,d\gamma+$$
$$\int_{\gamma_2^1}^{1-\gamma_{n-2}^2}|1-m_{n-2}-t_2|\,d\gamma+\cdots+\int_{\gamma_{n-2}^1}^{1-\gamma_2^2}|1-m_2-t_{n-2}|\,d\gamma+\int_{1-\gamma_2^2}^{\gamma_{n-1}^1}|1-m_1-t_{n-2}|\,d\gamma+$$
$$\int_{\gamma_{n-1}^1}^{1-\gamma_1^2}|1-m_1-t_{n-1}|\,d\gamma+\int_{1-\gamma_1^2}^{\gamma_n^1}|1-t_{n-1}|\,d\gamma,$$

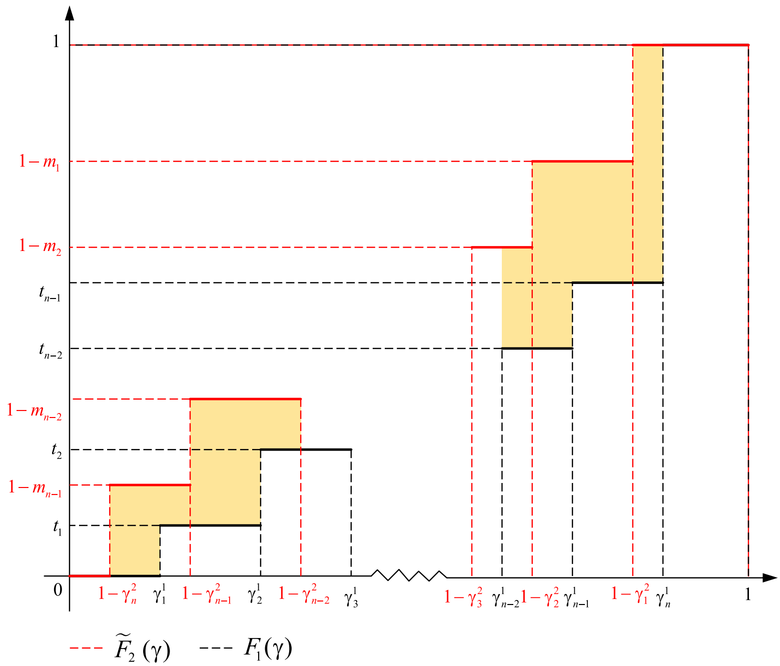

Figure 2. $F_1(\gamma)$ and $\tilde{F}2(\gamma)$

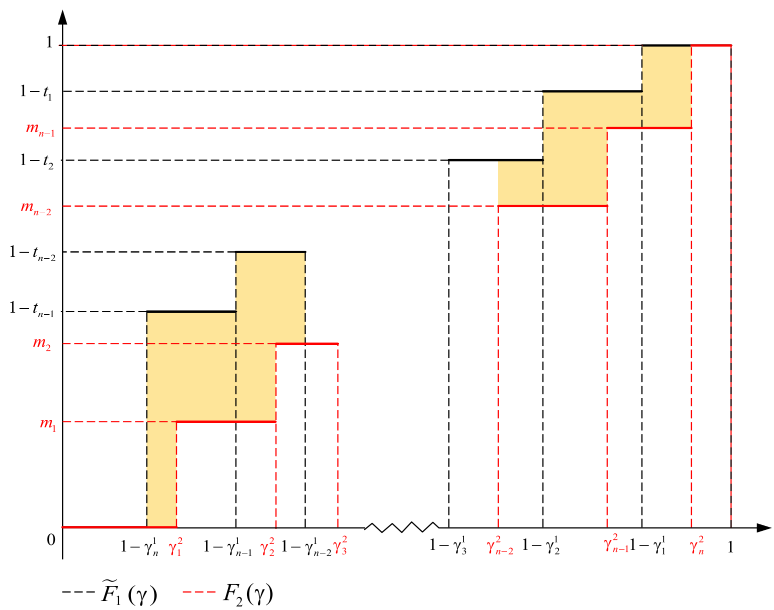

Figure 3. $F_2(\gamma)$ and $\tilde{F}1(\gamma)$

$$D(h_2(p),h_1^c(p)) = \int_{1-\gamma_n^1}^{\gamma_1^2} |1-t_{n-1}|\, d\gamma + \int_{\gamma_1^2}^{1-\gamma_{n-1}^1} |1-t_{n-1}-m_1|\, d\gamma + \int_{1-\gamma_{n-1}^1}^{\gamma_2^2} |1-t_{n-2}-m_1|\, d\gamma +$$
$$\int_{\gamma_2^2}^{1-\gamma_{n-2}^1} |1-t_{n-2}-m_2|\, d\gamma + \cdots + \int_{\gamma_{n-2}^2}^{1-\gamma_2^1} |1-t_2-m_{n-2}|\, d\gamma + \int_{1-\gamma_2^1}^{\gamma_{n-1}^2} |1-t_1-m_{n-2}|\, d\gamma +$$
$$\int_{\gamma_{n-1}^2}^{1-\gamma_1^1} |1-t_1-m_{n-1}|\, d\gamma + \int_{1-\gamma_1^1}^{\gamma_n^2} |1-m_{n-1}|\, d\gamma.$$

Therefore, $D(h_1(p),h_2^c(p)) = D(h_2(p),h_1^c(p))$.

The specific calculation process of distribution discrepancy degree between PHFEs is shown by the following example.

Example 1. Let $h_1(p)=\{0.1(0.2),0.2(0.5),0.3(0.3)\}$ and $h_2(p)=\{0.1(0.3),0.4(0.7)\}$ be two PHFEs, and their distribution sequences are shown in Table 1.

Table 1. Distribution sequence of the random variable X

	h_1(p)			h_2(p)	
X	0.1	0.2	0.3	0.1	0.4
P	0.2	0.5	0.3	0.3	0.7

Then the distribution functions $F_3(\gamma)$ for $h_1(p)$ and $F_4(\gamma)$ for $h_2(p)$ can be derived.

$$F_3(\gamma)=\begin{cases} 0 & 0\le\gamma<0.1 \\ 0.2 & 0.1\le\gamma<0.2 \\ 0.7 & 0.2\le\gamma<0.3 \\ 1 & 0.3\le\gamma\le 1 \end{cases},\quad F_4(\gamma)=\begin{cases} 0 & 0\le\gamma<0.1 \\ 0.3 & 0.1\le\gamma<0.4 \\ 1 & 0.4\le\gamma\le 1 \end{cases}.$$

The distribution discrepancy degree between $h_1(p)$ and $h_2(p)$ is obtained as below:

$$D(h_1(p),h_2(p)) = \int_0^1 \left|F_1(\gamma)-F_2(\gamma)\right| d\gamma =$$
$$\int_0^{0.1} |0-0|\, d\gamma + \int_{0.1}^{0.2} |0.2-0.3|\, d\gamma + \int_{0.2}^{0.3} |0.7-0.3|\, d\gamma +$$
$$\int_{0.3}^{0.4} |1-0.3|\, d\gamma + \int_{0.4}^{1} |1-1|\, d\gamma = 0.12.$$

3.3. A CBR classification method based on the distribution discrepancy degree

In the actual decision-making process, it is often difficult for experts to quickly rank and classify all the alternatives because of time pressure and information uncertainty. In addition, the selection of classification radius in traditional methods is often highly subjective, and thus using CBR for classification is an effective method, which can utilize typical successful cases from the past to rank and classify alternatives (Aamodt & Plaza, 1994).

3.3.1. Problem description

For the multi-attribute decision problem, let $X=\{X_1,X_2,\ldots,X_m\}$ be the set of alternatives, $C=\{C_1,C_2,\ldots,C_n\}$ be the set of attributes, and $w=(w_1,w_2,\ldots,w_n)$ be the vector of attribute weights, which is unknown and satisfies $0\le w_j\le 1$ and $\sum_{j=1}^{n} w_j=1$. The decision matrix $Z=(d_{ij})_{m\times n}$ is obtained, where d_{ij} is a PHFE and indicates the degree of alternative $X_i(i=1,2,\ldots,m)$ with respect to attribute $C_j(j=1,2,\ldots,n)$. The purpose of this paper is to rank the alternatives and classify them into pre-set categories.

3.3.2. Decision-making process

To rank the alternatives and classify them into different categories, the specific process of the algorithm is offered as below.

Step 1. Given a typical set of past cases $A=\{A_1,A_2,\ldots,A_s\}$, the experts evaluate them to obtain a decision matrix $T=(t_{ij})_{s\times n}$, where t_{ij} takes the form of PHFE.

Step 2. The experts classify the typical cases into categories $V_1,V_2,\ldots,V_q$ and determine the optimal case A^*.

According to their own knowledge and experience, experts divide the typical case set into q categories $V_1,V_2,\ldots,V_q$, where $V_p=\left\{A_1^p,A_2^p,\ldots,A_{n_p}^p\right\}$, $p=1,2,\ldots,q$, $V_1\cup V_2\cup\cdots\cup V_q=A$ and $V_i\cap V_j=\Phi(i\ne j)$. Assuming that the priority relationship is $V_1\succ V_2\succ\cdots\succ V_q$, we call it a test decision problem. The experts need to find the optimal case A^*, and obviously $A^*\in V_1$. If A^* is determined, then $A^*=A_{n_l}^1$. If A^* cannot be determined, then take the average of elements in V_1.

Step 3. Calculate the distribution discrepancy degree $D_j(A^*,A_k)$ between A^* and A_k under attribute C_j according to Equation (6).

Step 4. Determine the optimal attribute weight vector $w^*=(w_1^*,w_2^*,\cdots,w_n^*)$ and classification radius $R^*=(R_1^*,R_2^*,\cdots,R_q^*)$.

Let the classification radius be $R_1,R_2,\ldots,R_q$ and $0\le R_1<R_2<\cdots<R_q\le 1$. For any case A_k, if $A_k\in V_1$, then $0\le D(A^*,A_k)<R_1$. If $A_k\in V_q$, then $R_{q-1}\le D(A^*,A_k)\le 1$. If $A_k\in V_p(1<p<q)$, then $R_{p-1}\le D(A^*,A_k)\le R_p$. According to the distribution discrepancy degree between each case and the optimal case, an optimization model is established to determine the classification radius and the optimal attribute weights (Assuming that $R_0=0$ and $R_q=1$). Here, $0\le\alpha_k^p\le 1$ and $0\le\beta_k^p\le 1$ are relaxation variables.

$$\min\sum_{p=1}^{q}\sum_{k}^{n_p}\left((\alpha_k^p)^2+(\beta_k^p)^2\right)$$

$$\begin{aligned}\text{s.t. }&\sum_{j=1}^{n} w_j D_j(A^*,A_k^p)-\beta_k^p<R_p,\\&\sum_{j=1}^{n} w_j D_j(A^*,A_k^p)+\alpha_k^p>R_{p-1},\\&0\le R_1<R_2<\cdots<R_q\le 1,\\&0\le\alpha_k^p\le 1,0\le\beta_k^p\le 1,\end{aligned}$$

$$0 \le w_j \le 1, \sum_{j=1}^{n} w_j = 1,$$
$$k = 1, 2, \ldots, n_p, p = 1, 2, \ldots, q. \quad \text{(M-1)}$$

Obviously, the feasible domain formed by the constraints of model (M-1) is a convex set, and the objective function is quadratic. Therefore, there exists an optimal solution for the model. That is, the optimal weights $w^* = (w_1^*, w_2^*, \cdots, w_n^*)$ and the classification radius $R^* = (R_1^*, R_2^*, \cdots, R_q^*)$ can be derived by model (M-1).

Step 5. Calculate the distribution discrepancy degree $D(X_i, A^*)$ between alternative $X_i (i = 1, 2, \cdots, m)$ and A^*.

$$D(X_i, A^*) = \sum_{j=1}^{n} w_j^* D_j(X_i, A^*),\ i = 1,\ 2,\ \ldots\ldots,\ m. \quad (7)$$

Step 6. Rank the alternatives $X = \{X_1, X_2, \ldots, X_m\}$ and classify them into the corresponding categories.

To compare the two alternatives X_u and X_v, we calculate the distribution discrepancy degrees between the alternatives and the optimal case A^*. If $D(X_u, A^*) < D(X_v, A^*)$, then $X_u \succ X_v$. If $D(X_u, A^*) = D(X_v, A^*)$, then $X_u \sim X_v$. If $D(X_u, A^*) > D(X_v, A^*)$, then $X_u \prec X_v$.

Therefore, if $0 \le D(X_i, A^*) < R_1^*$, then $X_i \in V_1$. If $R_{p-1}^* \le D(X_i, A^*) < R_p^* (1 < p < q)$, then $X_i \in V_p$. If $R_{q-1}^* \le D(X_i, A^*) \le 1$, then $X_i \in V_q$.

4. Case study

In this Section, the proposed approach is used to address the classification problem on social medical insurance fraud.

Social medical insurance is an insurance that compensates people for the medical costs associated with the treatment of diseases. It is characterized by numerous operation links, complex structure of insured personnel, strong concealment of fraud and great difficulty in identification. In recent years, there have been frequent incidents of social medical insurance fraud. In addition, due to the imperfection of the regulatory system, the backwardness of the regulatory means and the lack of supervisory power and experience, some cases have involved amounts of up to one million yuan. For some suspicious medical insurance claims, it is bound to spend a lot of human and material resources on investigating, and thus increases the operating costs of social medical insurance institutions, which causes the whole social medical insurance system to operate inefficiently and poses a serious threat to its healthy and orderly development. Therefore, it is of great significance to attach great importance to preventing and combating medical insurance fraud. However, the harms caused by different levels of medical insurance fraud incidents are different, and thus the means to prevent and combat them are different. In this paper, the typical types of medical insurance fraud were extracted from the existing literatures, as shown in Table 2. Since there are many forms of medical insurance fraud, the 27 typical types of medical insurance fraud in Table 2 are considered as decision-making attributes $C_j (j = 1, 2, \cdots, 27)$, which can comprehensively reflect the different medical insurance fraud behaviors. Assume that the attribute weights are partially known.

$0.05 \le w_1 \le 0.1, 0.01 \le w_2 \le 0.05, 0.04 \le w_3 \le 0.15, 0.1 \le w_4 \le 0.25, 0.05 \le w_5 \le 0.12, 0.04 \le w_6 \le 0.1,$
$0.09 \le w_7 \le 0.15, 0.01 \le w_8 \le 0.06, 0.01 \le w_9 \le 0.06, 0.05 \le w_{10} \le 0.12, 0.06 \le w_{11} \le 0.15, 0.01 \le w_{12} \le 0.08,$
$0.01 \le w_{13} \le 0.1, 0.04 \le w_{14} \le 0.1, 0.01 \le w_{15} \le 0.05, 0.01 \le w_{16} \le 0.15, 0.01 \le w_{17} \le 0.07, 0.08 \le w_{18} \le 0.16,$
$0.01 \le w_{19} \le 0.1, 0.03 \le w_{20} \le 0.1, 0.04 \le w_{21} \le 0.09, 0.01 \le w_{22} \le 0.1, 0.01 \le w_{23} \le 0.1, 0.05 \le w_{24} \le 0.15,$
$0.01 \le w_{25} \le 0.1, 0.03 \le w_{26} \le 0.1, 0.01 \le w_{27} \le 0.08$.

In this case, there are 8 incidents of medical insurance fraud to be evaluated, which are recorded as alternatives $X_1, X_2, \cdots, X_8$. Experts with research experience in medical treatment, data management and other fields are invited to evaluate the 8 alternatives with respect to attributes $C_j (j = 1, 2, \cdots, 27)$. The decision matrix $Z = (h_{ij}(p))_{m \times n}$ is obtained as shown in Table 3, where $h_{ij}(p)$ denotes the PHFE. The aim is to classify these 8 alternatives into 4 categories: V_1 (mild fraud), V_2 (moderate fraud), V_3 (severe fraud), and V_4 (extreme fraud), and rank them according to the severity of fraud.

Table 2. Types of medical insurance fraud

Major categories	Forms of insurance fraud	Literature sources
Medical institution	Forging medical records C_1 Lowering the hospitalization threshold C_2 Increasing service in disguised form C_3 Prescribing "yin and yang prescriptions" C_4 Decommissioning hospitalization C_5	Ataabadi et al. (2022) Li et al. (2022) Villegas-Ortega et al. (2021) Villegas-Ortega et al. (2021) Villegas-Ortega et al. (2021)
Patient	Counterfeiting others' medical insurance cards C_6 Swapping card to withdraw cash or purchase non-medical insurance catalog items C_7 Repeating visits for off-site medical treatment or false opening of settlement vouchers C_8 Over-dispensing drugs and selling at a discount C_9	Li et al. (2022) Zhang (2021) Ataabadi et al. (2022) Li et al. (2022)
Doctor-patient conspiracy	Swapping non-Medicare payment categories C_{10} Hospitalization by bed C_{11} Exaggeration the illness C_{12} Fiction of medical treatment facts C_{13}	Kapadiya et al. (2022) Li et al. (2022) Li et al. (2022) Kapadiya et al. (2022)
Medical insurance designated pharmacy	Raising drug prices C_{14} Rebate promotion C_{15} Selling drugs to people with other's medical insurance cards C_{16} Refunding after empty swiping or card swiping C_{17} Uploading false drug sales records C_{18} Transferring information terminal to others C_{19}	Zhang (2021) Zhang (2021) Zhang (2021) Zhang (2021) Villegas-Ortega et al. (2021) Chen et al. (2021)
Social medical insurance institutions	Fraudulent bill reimbursement irregularities C_{20} Misappropriating social medical insurance funds C_{21} Irregular payments to the medical insurance fund C_{22} False medical security treatment procedures C_{23}	Villegas-Ortega et al. (2021) Zhou & Su (2021) Zhang (2021) Villegas-Ortega et al. (2021)
Medical insurance fraud professional groups	Leasing, borrowing or acquiring others' medical insurance cards for use C_{24} Forging or selling false materials C_{25} Reselling drugs C_{26} Tampering social insurance information C_{27}	Li et al. (2022) Li et al. (2022) Zhang et al. (2022) Guida (2021)

Step 1. Based on the historical data, the experts give the decision matrix T consisted of 12 past typical cases $(A_1, A_2, \cdots, A_{12})$, which is shown in Table 4.

Step 2. According to the successful experience of typical cases in the past, experts classify the cases $A = (A_1, A_2, \cdots, A_{12})$ into four categories: $V_1 = (A_3, A_{12})$, $V_2 = (A_7, A_9, A_{10}, A_{11})$, $V_3 = (A_1, A_2, A_4, A_8)$, $V_4 = (A_5, A_6)$, where the best case is $A^* = A_{12}$.

Step 3. Calculate the distribution discrepancy degree $D_j(A^*, A_k)$ between A^* and A_k under attribute C_j, as shown in Table 5.

Table 3. Decision matrix $Z = (h_{ij}(p))_{m\times n}$

	C_1	C_2	C_3	C_4
X_1	{0.6(0.4),0.9(0.6)}	{0.7(0.7),0.8(0.3)}	{0.5(0.5),0.6(0.5)}	{0.7(0.4),0.9(0.6)}
X_2	{0.4(0.6),0.6(0.4)}	{0.3(0.4),0.6(0.6)}	{0.6(1)}	{0.6(0.3),0.7(0.7)}
X_3	{0.8(0.3),0.9(0.7)}	{0.5(1)}	{0.8(0.3),0.9(0.7)}	{0.8(0.3),0.9(0.7)}
X_4	{0.1(1)}	{0.2(1)}	{0.2(0.6),0.5(0.4)}	{0.2(0.5),0.3(0.5)}
X_5	{0.1(0.7),0.2(0.3)}	{0.1(0.8),0.2(0.2)}	{0.7(1)}	{0.5(0.5),0.6(0.5)}
X_6	{0.1(0.4),0.3(0.6)}	{0.1(1)}	{0.2(1)}	{0.2(0.4),0.3(0.6)}
X_7	{0.6(0.5),0.7(0.5)}	{0.2(0.3),0.4(0.7)}	{0.2(0.4),0.4(0.6)}	{0.7(0.6),0.9(0.4)}
X_8	{0.3(1)}	{0.3(0.4),0.4(0.6)}	{0.4(0.3),0.5(0.7)}	{0.3(0.3),0.6(0.7)}

	C_5	C_6	C_7	C_8
X_1	{0.6(1)}	{0.1(0.8),0.3(0.2)}	{0.8(1)}	{0.2(0.8),0.3(0.2)}
X_2	{0.4(0.3),0.5(0.7)}	{0.3(0.85),0.4(0.15)}	{0.5(0.6),0.7(0.4)}	{0.5(1)}
X_3	{0.8(0.5),0.9(0.5)}	{0.8(1)}	{0.7(0.6),0.9(0.4)}	{0.7(0.3),0.9(0.7)}
X_4	{0.1(0.4),0.3(0.5),0.6(0.1)}	{0(1)}	{0.5(0.65),0.7(0.35)}	{0.2(0.3),0.4(0.7)}
X_5	{0.1(0.8),0.3(0.2)}	{0.4(0.4),0.5(0.6)}	{0.2(0.5),0.3(0.5)}	{0.5(0.25),0.6(0.75)}
X_6	{0.2(0.25),0.4(0.75)}	{0.1(0.4),0.2(0.6)}	{0.7(1)}	{0.8(0.9),0.9(0.1)}
X_7	{0.1(0.9),0.3(0.1)}	{0.7(1)}	{0.8(0.2),0.9(0.8)}	{0.3(0.2),0.4(0.8)}
X_8	{0.2(1)}	{0.3(0.6),0.5(0.4)}	{0.2(0.5),0.4(0.5)}	{0.7(0.6),0.9(0.4)}

	C_9	C_{10}	C_{11}	C_{12}
X_1	{0.1(0.3),0.3(0.7)}}	{0.7(1)}	{0.2(0.5),0.4(0.5)}	{0.4(0.5),0.7(0.5)}
X_2	{0.5(0.3),0.6(0.7)}	{0.3(0.6),0.4(0.4)}	{0.1(1)}	{0.3(0.5),0.5(0.5)}
X_3	{0.8(0.85),0.9(0.15)}	{0.2(0.4),0.4(0.6)}	{0.6(0.2),0.8(0.8)}	{0.7(1)}
X_4	{0.4(0.3),0.5(0.7)}	{0.2(0.1),0.4(0.9)}	{0.1(1)}	{0.2(0.3),0.6(0.7)}
X_5	{0.3(0.8),0.4(0.2)}	{0.1(0.7),0.2(0.3)}	{0.1(0.5),0.3(0.5)}	{0.2(0.65),0.3(0.35)}
X_6	{0.6(1)}	{0.2(0.1),0.3(0.5),0.4(0.4)}	{0.1(0.8),0.2(0.2)}	{0.7(0.85),0.8(0.15)}
X_7	{0.3(0.6),0.6(0.4)}	{0.3(0.5),0.6(0.5)}	{0.8(1)}	{0.2(0.3),0.5(0.7)}
X_8	{0.4(0.6),0.5(0.4)}	{0.2(0.6),0.5(0.4)}	{0.3(0.2),0.5(0.8)}	{0.8(0.9),0.9(0.1)}

End of Table 3

	C_{13}	C_{14}	C_{15}	C_{16}
X_1	{0.7(1)}	{0.1(0.9),0.3(0.1)}	{0.2(1)}	{0.1(0.8),0.2(0.2)}
X_2	{0.3(0.3),0.4(0.7)}	{0.1(0.3),0.2(0.7)}	{0.1(0.3),0.3(0.7)}	{0.6(0.4),0.9(0.6)}
X_3	{0.7(0.3),0.9(0.7)}	{0.7(0.6),0.8(0.4)}	{0.3(0.5),0.5(0.5)}	{0.6(0.2),0.8(0.8)}
X_4	{0.2(0.1),0.4(0.9)}	{0.1(0.7),0.2(0.3)}	{0.3(0.55),0.4(0.45)}	{0(1)}
X_5	{0.1(0.85),0.2(0.15)}	{0.1(0.3),0.3(0.7)}	{0.1(0.8),0.2(0.2)}	{0.3(1)}
X_6	{0.8(1)}	{0.1(0.8),0.2(0.2)}	{0(1)}	{0.1(0.6),0.2(0.4)}
X_7	{0.3(0.8),0.5(0.2)}	{0.6(1)}	{0.1(0.4),0.2(0.6)}	{0.8(0.4),0.9(0.6)}
X_8	{0.1(0.5),0.2(0.5)}	{0.1(0.2),0.2(0.8)}	{0.1(0.5),0.3(0.5)}	{0.3(0.5),0.5(0.5)}

	C_{17}	C_{18}	C_{19}	C_{20}
X_1	{0.1(1)}	{0.7(0.9),0.8(0.1)}	{0.1(0.9),0.2(0.1)}	{0.6(0.5),0.8(0.5)}
X_2	{0.3(0.5),0.5(0.5)}	{0.2(0.8),0.3(0.2)}	{0.1(0.4),0.2(0.6)}	{0.3(1)}
X_3	{0.5(0.4),0.7(0.6)}	{0.8(1)}	{0.3(1)}	{0.6(0.1),0.8(0.9)}
X_4	{0.2(0.8),0.3(0.2)}	{0.3(0.6),0.4(0.4)}	{0.2(0.75),0.3(0.25)}	{0.2(0.5),0.6(0.5)}
X_5	{0.2(0.85),0.3(0.15)}	{0.1(0.2),0.2(0.8)}	{0(1)}	{0.1(0.6),0.2(0.4)}
X_6	{0.1(1)}	{0.2(0.45),0.4(0.55)}	{0.1(1)}	{0.2(1)}
X_7	{0.2(0.3),0.3(0.7)}	{0.9(1)}	{0.2(0.4),0.4(0.6)}	{0.6(0.65),0.8(0.35)}
X_8	{0.2(0.5),0.3(0.5)}	{0.2(0.3),0.4(0.7)}	{0.1(0.8),0.2(0.2)}	{0.2(0.3),0.4(0.7)}

	C_{21}	C_{22}	C_{23}	C_{24}
X_1	{0.1(1)}	{0.4(0.3),0.6(0.7)}	{0.2(0.8),0.3(0.2)}	{0.1(0.7),0.2(0.3)}
X_2	{0(1)}	{0.5(1)}	{0.8(1)}	{0.2(0.6),0.3(0.4)}
X_3	{0.4(0.8),0.5(0.2)}	{0.8(0.2),0.9(0.8)}	{0.6(0.6),0.9(0.4)}	{0.8(0.3),0.9(0.7)}
X_4	{0.2(0.9),0.3(0.1)}	{0.2(0.35),0.6(0.65)}	{0.2(0.4),0.4(0.6)}	{0(1)}
X_5	{0.1(1)}	{0.2(1)}	{0.2(1)}	{0.3(0.4),0.5(0.6)}
X_6	{0.1(0.8),0.2(0.2)}	{0.3(0,3),0.5(0.7)}	{0.2(0.2),0.5(0.8)}	{0.2(0.5),0.4(0.5)}
X_7	{0.1(0.7),0.2(0.2),0.3(0.1)}	{0.2(0.6),0.4(0.4)}	{0.7(1)}	{0.7(0.7),0.8(0.3)}
X_8	{0.2(1)}	{0.5(0.7),0.6(0.3)}	{0.3(0.2),0.4(0.8)}	{0.4(0.2),0.5(0.8)}

	C_{25}	C_{26}	C_{27}
X_1	{0.1(0.8),0.2(0.2)}	{0.1(0.9),0.2(0.1)}	{0.1(1)}
X_2	{0.3(1)}	{0.5(0.35),0.7(0.65)}	{0.2(0.8),0.3(0.2)}
X_3	{0.9(1)}	{0.8(0.8),0.9(0.2)}	{0.3(1)}
X_4	{0.1(0.2),0.3(0.8)}	{0.1(0.6),0.4(0.4)}	{0(1)}
X_5	{0.1(0.8),0.2(0.2)}	{0.3(1)}	{0.1(0.8),0.2(0.2)}
X_6	{0.1(1)}	{0.1(0.3),0.2(0.7)}	{0(1)}
X_7	{0.8(0.6),0.9(0.4)}	{0.7(0.2),0.8(0.8)}	{0.1(0.7),0.3(0.3)}
X_8	{0.3(1)}	{0.4(0.4),0.5(0.6)}	{0.1(1)}

Table 4. Decision matrix of historical typical cases

	C_1	C_2	C_3	C_4
A_1	{0.7(0.5),0.9(0.5)}	{0.8(0.75),0.9(0.25)}	{(0.5(1)}	{0.9(1)}
A_2	{0.3(1)}	{0.3(0.9),0.4(0.1)}	{0.4(0.3),0.6(0.7)}	{0.6(0.3),0.8(0.7)}
A_3	{0.4(0.2),0.5(0.8)}	{0.2(1)}	{0.6(0.9),0.7(0.1)}	{0.2(0.1),0.3(0.9)}
A_4	{0.6(0.7),0.8(0.3)}	{0.3(0.8),0.5(0.2)}	{0.4(1)}	{0.2(0.9),0.3(0.1)}
A_5	{0.8(0.85),0.9(0.15)}	{0.7(0.5),0.9(0.5)}	{0.8(0.4),0.9(0.6)}	{0.8(0.9),0.9(0.1)}
A_6	{0.9(1)}	{0.4(0.3),0.5(0.7)}	{0.8(0.8),0.9(0.2)}	{0.8(0.3),0.9(0.7)}
A_7	{0.5(0.6),0.7(0.4)}	{0.5(1)}	{0.4(0.8),0.6(0.2)}	{0.6(1)}
A_8	{0.6(0.5),0.8(0.5)}	{0.5(0.6),0.6(0.4)}	{0.3(0.8),0.5(0.2)}	{0.6(0.3),0.9(0.7)}
A_9	{0.5(0.35),0.6(0.65)}	{0.4(1)}	{0.5(0.85),0.7(0.15)}	{0.6(1)}
A_{10}	{0.5(1)}	{0.4(0.8),0.5(0.2)}	{0.4(0.9),0.5(0.1)}	{0.4(0.9),0.6(0.1)}
A_{11}	{0.3(0.3),0.5(0.7)}	{0.2(0.6),0.3(0.4)}	{0.4(0.6),0.5(0.4)}	{0.5(0.8),0.7(0.2)}
A_{12}	{0.1(1)}	{0.1(0.6),0.2(0.4)}	{0.6(0.9),0.7(0.1)}	{0.4(0.5),0.6(0.5)}

	C_5	C_6	C_7	C_8
A_1	{0.5(0.4),0.6(0.6)}	{0.8(1)}	{0.5(0.6),0.8(0.4)}	{0.7(1)}
A_2	{0.5(1)}	{0.8(0.9),0.9(0.1)}	{0.7(0.3),0.8(0.7)}	{0.5(0.6),0.6(0.4)}
A_3	{0.1(1)}	{0.3(1)}	{0.2(0.7),0.3(0.3)}	{0.2(0.9),0.3(0.1)}
A_4	{0(1)}	{0.7(0.6),0.8(0.4)}	{0.6(0.5),0.9(0.5)}	{0.8(1)}
A_5	{0.4{0.5),0.5(0.5)}	{0.8(1)}	{0.7(0.6),0.8(0.4)}	{0.7(0.9),0.8(0.1)}
A_6	{0.8(0.7),0.9(0.3)}	{0.7(1)}	{0.7(0.1),0.8(0.7),0.9(0.2)}	{0.8(0.5),0.9(0.5)}
A_7	{0.2(1)}	{0.4(0.8),0.6(0.2)}	{0.4(0.2),0.6(0.8)}	{0.4(0.6),0.5(0.4)}
A_8	{0.4(0.6),0.7(0.4)}	{0.6(0.7),0.8(0.3)}	{0.3(1)}	{0.5(0.4),0.7(0.6)}
A_9	{0.4(1)}	{0.8(0.3),0.9(0.7)}	{0.5(0.4),0.6(0.6)}	{0.6(0.7),0.7(0.3)}
A_{10}	{0(1)}	{0.6(0.15),0.9(0.85)}	{0.4(0.2),0.6(0.8)}	{0.5(0.6),0.7(0.4)}
A_{11}	{0.5(0.95),0.6(0.05)}	{0.6(0.6),0.8(0.4)}	(0.4(0.5),0.5(0.5)}	{0.4(0.6),0.7(0.4)}
A_{12}	{0.2(0.8),0.3(0.2)}	{0.5(1)}	{0.1(0.5),0.2(0.5)}	{0.1(0.7),0.2(0.3)}

	C_9	C_{10}	C_{11}	C_{12}
A_1	{0.6(1)}	{0.9(1)}	{0.7(0.5),0.8(0.5)}	{0.4(0.7),0.7(0.3)}
A_2	{0.5(1)}	{0.5(0.6),0.6(0.4)}	{0.6(0.6),0.7(0.4)}	{0.4(0.9),0.5(0.1)}
A_3	{0.3(0.8),0.4(0.2)}	{0.1(0.9),0.2(0.1)}	{0.1(0.7),0.2(0.3)}	{0.1(0.8),0.2(0.2)}
A_4	{0.6(0.5),0.7(0.5)}	{0.7(0.3),0.8(0.7)}	{0.6(0.8),0.7(0.2)}	{0.8(0.1)}
A_5	{0.7(1)}	{0.6(0.2),0.7(0.1),0.8(0.7)}	{0.8(0.7),0.9(0.3)}	{0.9(1)}
A_6	{0.8(0.25),0.9(0.75)}	{0.7(0.3),0.9(0.7)}	{0.6(0.8),0.7(0.2)}	{0.7(0.8),0.8(0.2)}
A_7	{0.3(0.8),0.4(0.2)}	{0.4(1)}	{0.6(0.8),0.8(0.2)}	{0.3(0.6),0.4(0.4)}
A_8	{0.5(0.2),0.7(0.8)}	{0.6(0.4),0.7(0.6)}	0.5(0.4),0.8(0.6)}	{0.7(1)}
A_9	{0.4(1)}	{0(1)}	{0.4(0.9),0.5(0.1)}	{0.4(0.9),0.6(0.1)}
A_{10}	{0.5(1)}	{(0.5(0.8),0.7(0.2)}	{0.5(1)}	{0.5(0.7),0.6(0.3)}
A_{11}	{0.4(0.8),0.5(0.2)}	{0.3(0.6),0.5(0.4)}	{0.4(0.7),0.6(0.3)}	{0.5(0.8),0.7(0.2)}
A_{12}	{0.3{0.8),0.5(0.2)}	{0.1(1)}	{0.1(0.8),0.2(0.2)}	{0.1(0.6),0.2(0.4)}

Continue of Table 4

	C_{13}	C_{14}	C_{15}	C_{16}
A_1	{0.7(0.5),0.8(0.5)}	{0.1(0.9),0.2(0.1)}	{0.1(0.8),0.2(0.2)}	{0.7(0.8),0.8(0.2)}
A_2	{0.7(1)}	{0.1(1)}	{0.1(0.8),0.2(0.2)}	{0.9(1)}
A_3	{0.2(0.5),0.3(0.5)}	{0.1(0.4),0.2(0.6)}	{0.1(1)}	{0.1(1)}
A_4	{0.8(0.7),0.9(0.3)}	{0.7(0.5),0.9(0.5)}	{0.8(0.55),0.9(0.45)}	{0.3(0.4),0.6(0.6)}
A_5	{0.7(0.4),0.8(0.6)}	{0.7(0.5),0.8(0.5)}	{0.2(0.4),0.4(0.6)}	{0.8(1)}
A_6	{0.7(0.4),0.9(0.6)}	{0.8(0.7),0.9(0.3)}	{0.3(0.5),0.4(0.5)}	{0.4(0.5),0.6(0.5)}
A_7	{0.3(0.5),0.5(0.5)}	{0.3(1)}	{0.3(1)}	{0.3(0.7),0.5(0.3)}
A_8	{0.4(0.6),0.7(0.4)}	{0.6(0.8),0.8(0.2)}	{0.3(0.5),0.5(0.5)}	{0.8(0.15),0.9(0.85)}
A_9	{0.4(0.6),0.5(0.4)}	{0.4(0.9),0.5(0.1)}	{0.5(0.9),0.6(0.1)}	{0.6(0.1),0.9(0.9)}
A_{10}	{0.5(1)}	{0.6(1)}	{0.3(0.9),0.5(0.1)}	{0.7(0.2),0.9(0.8)}
A_{11}	{0.2(0.2),0.5(0.8)}	{0.4(0.9),0.6(0.1)}	{0.4(0.6),0.5(0.4)}	{0.3(0.6),0.4(0.4)}
A_{12}	{0.1(1)}	{0.1(0.5),0.2(0.5)}	{0.1(1)}	{0.5(1)}

	C_{17}	C_{18}	C_{19}	C_{20}
A_1	{0.3(0.9),0.4(0.1)}	{0.3(1)}	{0.1(1)}	{0.6(0.4),0.7(0.6)}
A_2	{0.2(0.9),0.3(0.1)}	{0.6(0.5).0.7(0.5)}	{0.7(1)}	{0.6(0.9),0.7(0.1)}
A_3	{0.1(1)}	{0.1(0.85),0.2(0.15)}	{0(1)}	{0.2(1)}
A_4	{0.5(0.6),0.6(0.4)}	{0.7(0.7),0.8(0.3)}	{0.4(0.7),0.5(0.3)}	{0.8(0.5),0.9(0.5)}
A_5	{0.6(0.7),0.7(0.3)}	{0.7(0.6),0.8(0.4)}	{0.3(0.4),0.5(0.6)}	{0.7(0.7),0.8(0.3)}
A_6	{0.6(0.65),0.8(0.35)}	{0.5(0.6),0.7(0.4)}	{0.2(0.3),0.4(0.7)}	{0.6(0.65),0.8(0.35)}
A_7	{0.1(0.1),0.3(0.9)}	{0.5(0.6),0.6(0.4)}	{0.1(1)}	{0.4(0.4),0.5(0.6)}
A_8	{0.3(0.6),0.4(0.4)}	{0.3(0.3),0.6(0.7)}	{0.3(0.3),0.5(0.7)}	{0.6(0.8),0.8(0.2)}
A_9	{0.3(0.3),0.5(0.7)}	{0.4(0.5),0.6(0.5)}	{0.3(0.45),0.4(0.55)}	{0.4(0.35),0.5(0.65)}
A_{10}	{0.4(0.8),0.6(0.2)}	{0.3(0.7),0.5(0.3)}	{0.3(0.6),0.4(0.4)}	{0.4(0.4),0.5(0.6)}
A_{11}	{0.4(1)}	{0.3(0.2),0.5(0.8)}	{0.2(0.5),0.4(0.5)}	{0.2(0.4),0.6(0.6)}
A_{12}	{0.1(0.9),0.2(0.1)}	{0.1(0.4),0.2(0.6)}	{0(1)}	{0.2(0.8),0.3(0.2)}

	C_{21}	C_{22}	C_{23}	C_{24}
A_1	{0.1(1)}	{0.6(0.4},0.8(0.6)}	{0.6(0.45),0.7(0.55)}	{0.5(1)}
A_2	{0(1)}	{0.2(0.7),0.4(0.3)}	{0.9(1)}	{1(1)}
A_3	{0(1)}	{0.3(0.5),0.4(0.5)}	{0.1(0.9),0.3(0.1)}	{0.1(0.5),0.2(0.5)}
A_4	{0.6(0.8),0.7(0.2)}	{0.9(1)}	{0.7(0.6),0.9(0.4)}	{0.8(0.8),0.9(0.2)
A_5	{0.6(1)}	{0.4(0.3),0.7(0.7)}	{0.5(1)}	{0.9(1)}
A_6	{0.4(0.4),0.5(0.6)}	{0.8(0.35),0.9(0.65)}	{0.7(0.45),0.9(0.55)}	{0.7(0.4),0.8(0.6)}
A_7	{0(1)}	{0.5(0.9),0.6(0.1)}	{0.4(0.5),0.5(0.5)}	{0.2(0.15),0.4(0.85)}
A_8	(0.3(0.5),0.5(0.5)}	{0.5(0.6),0.7(0.4)}	{0.3(0.5),0.5(0.5)}	{0.4(0.6),0.6(0.4)}
A_9	{0.1(1)}	{0.4(0.9),0.5(0.1)}	{0.5(0.8),0.7(0.2)}	{0.5(0.1),0.6(0.9)}
A_{10}	{0(1)}	{0.2(0.8),0.3(0.2)}	{{0.3(0.7),0.5(0.3)}	{0.6(0.8),0.7(0.2)}
A_{11}	{0.1(0.9),0.2(0.1)}	{0.4(0.7),0.6(0.3)}	{0.6(1)}	{0.6(0.3),0.8(0.7)}
A_{12}	{0(1)}	{0.1(0.1),0.2(0.9)}	{0.1(0.7),0.2(0.3)}	{0.4(0.8),0.5(0.2)}

End of Table 4

	C_{25}	C_{26}	C_{27}
A_1	{0.7(0.9),0.9(0.1)}	{0.6(0.8),0.8(0.2)}	{0.5(1)}
A_2	{0.8(0.9),0.9(0.1)}	{0.7(1)}	{0.7(0.7),0.8(0.3)}
A_3	{0.1(1)}	{0.1(1)}	{0(1)}
A_4	{0.8(0.95),0.9(0.05)}	{0.7(0.8),0.8(0.2)}	{0.1(0.8),0.2(0.2)}
A_5	{0.8(1)}	{0.7(0.8),0.8(0.2)}	{0.4(1)}
A_6	{0.8(0.3),0.9(0.7)}	{0.7(0.4),0.9(0.6)}	{0.3(0.75),0.4(0.25)
A_7	{0.3(0.9),0.4(0.1)}	{0.2(0.9),0.3(0.1)}	{0.2(0.9),0.3(0.1)}
A_8	{0.6(0.7),0.8(0.3)}	{0.4(0.2),0.6(0.8)}	{0.3(1)}
A_9	{0.6(0.2),0.7(0.8)}	{0.6(1)}	{0.3(0.8),0.4(0.2)}
A_{10}	{0.6(0.7),0.7(0.3)}	{0.4(0.9),0.5(0.1)}	{0.2(0.9),0.3(0.1)}
A_{11}	{0.5(0.3),0.7(0.7)}	{0.3(0.7),0.4(0.3)}	{0.2(0.5),0.3(0.5)}
A_{12}	{0.1(1)}	{0.1(0.6),0.2(0.4)}	{0(1)}

Table 5. Distribution discrepancy degree $D_j(A^*, A_k)$

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	C_{11}	C_{12}	C_{13}	C_{14}
A_1	0.7	0.685	0.02	0.4	0.34	0.3	0.47	0.57	0.26	0.8	0.63	0.35	0.65	0.04
A_2	0.2	0.17	0.07	0.24	0.28	0.31	0.62	0.41	0.16	0.44	0.52	0.27	0.6	0.05
A_3	0.38	0.06	0	0.21	0.12	0.2	0.08	0.08	0.1	0.01	0.01	0.02	0.15	0.01
A_4	0.56	0.2	0.21	0.29	0.22	0.24	0.6	0.67	0.31	0.67	0.5	0.66	0.73	0.65
A_5	0.715	0.66	0.25	0.31	0.23	0.3	0.59	0.58	0.36	0.65	0.71	0.76	0.66	0.6
A_6	0.8	0.33	0.2	0.37	0.61	0.2	0.66	0.72	0.535	0.74	0.5	0.58	0.72	0.68
A_7	0.48	0.36	0.17	0.1	0.02	0.1	0.41	0.31	0.02	0.3	0.52	0.2	0.3	0.15
A_8	0.6	0.4	0.27	0.31	0.3	0.16	0.15	0.49	0.32	0.56	0.56	0.56	0.42	0.49
A_9	0.465	0.26	0.09	0.1	0.18	0.37	0.41	0.5	0.1	0.1	0.29	0.28	0.34	0.26
A_{10}	0.4	0.28	0.2	0.08	0.22	0.355	0.41	0.45	0.16	0.44	0.38	0.39	0.4	0.45
A_{11}	0.34	0.04	0.17	0.1	0.285	0.18	0.3	0.39	0.08	0.28	0.34	0.4	0.34	0.27
A_{12}	0	0	0	0	0	0	0	0	0	0	0	0	0	0

	C_{15}	C_{16}	C_{17}	C_{18}	C_{19}	C_{20}	C_{21}	C_{22}	C_{23}	C_{24}	C_{25}	C_{26}	C_{27}
A_1	0.02	0.22	0.2	0.14	0.1	0.44	0.1	0.53	0.525	0.08	0.62	0.5	0.5
A_2	0.02	0.4	0.1	0.49	0.7	0.39	0	0.07	0.77	0.58	0.71	0.56	0.73
A_3	0	0.4	0.01	0.045	0	0.02	0	0.16	0.03	0.27	0	0.04	0
A_4	0.745	0.14	0.43	0.57	0.43	0.63	0.62	0.71	0.65	0.4	0.705	0.58	0.12
A_5	0.22	0.3	0.52	0.58	0.42	0.51	0.6	0.42	0.37	0.48	0.7	0.58	0.4
A_6	0.25	0.1	0.56	0.42	0.34	0.45	0.46	0.675	0.68	0.34	0.77	0.68	0.325
A_7	0.2	0.14	0.17	0.38	0.1	0.24	0	0.32	0.32	0.05	0.21	0.07	0.21
A_8	0.3	0.385	0.23	0.35	0.44	0.42	0.4	0.39	0.27	0.06	0.56	0.42	0.3
A_9	0.41	0.37	0.33	0.34	0.355	0.245	0.1	0.22	0.41	0.17	0.58	0.46	0.32
A_{10}	0.22	0.36	0.33	0.2	0.34	0.24	0	0.03	0.23	0.2	0.53	0.27	0.21
A_{11}	0.34	0.16	0.29	0.3	0.3	0.22	0.11	0.27	0.47	0.32	0.54	0.19	0.25
A_{12}	0	0	0	0	0	0	0	0	0	0	0	0	0

Step 4. Determine the optimal attribute weight w^* and classification radius R^* according to model (M-1).

$$w^* = (0.1, 0.01, 0.04, 0.1, 0.0535, 0.04, 0.09, 0.01, 0.01, 0.05, 0.0631, 0.0374, 0.01, 0.04, 0.017, 0.01, 0.01, 0.0878, 0.01, 0.03, 0.04, 0.0192, 0.0192, 0.0528, 0.01, 0.03, 0.01),$$
$$R^* = (0.1146, 0.2802, 0.5146).$$

Step 5. Calculate the distribution discrepancy degrees between alternatives $X = \{X_1, X_2, \ldots, X_m\}$ and A^* according to Equation (7) as follows.

$$D(X_1, A^*) = 0.3674, D(X_2, A^*) = 0.2153, D(X_3, A^*) = 0.4719, D(X_4, A^*) = 0.2130, D(X_5, A^*) = 0.0687, D(X_6, A^*) = 0.2164, D(X_7, A^*) = 0.4214, D(X_8, A^*) = 0.1807.$$

Step 6. Rank and classify the alternatives $X = \{X_1, X_2, \ldots, X_m\}$.

Since $D(X_5, A^*) < D(X_8, A^*) < D(X_4, A^*) < D(X_2, A^*) < D(X_6, A^*) < D(X_1, A^*) < D(X_7, A^*) < D(X_3, A^*)$, the ranking result of the alternatives is obtained as $X_5 \succ X_8 \succ X_4 \succ X_2 \succ X_6 \succ X_1 \succ X_7 \succ X_3$. $D(X_1, A^*)$, $D(X_3, A^*)$ and $D(X_7, A^*)$ are greater than R_2^* and less than R_3^*. $D(X_5, A^*)$ is less than R_1^*. $D(X_2, A^*)$, $D(X_4, A^*)$, $D(X_6, A^*)$ and $D(X_8, A^*)$ are greater than R_1^* and less than R_2^*. Therefore, $X_1, X_3, X_7 \in V_3$, $X_5 \in V_1$, and $X_2, X_4, X_6, X_8 \in V_2$.

5. Comparison analysis

5.1. Comparison with Xu et al.'s method (2022)

Xu et al. (2022) defined the fuzzy entropy and hesitancy entropy for PHFEs and combined them to propose an overall entropy as follows.

$$E_O(h(p)) = \begin{cases} 1 & h(p) = h^* \\ \dfrac{A+B}{1+AB} & h(p) \neq h^* \end{cases}, \tag{8}$$

where $h^* = \{0(p_1), 1(p_2) \mid p_1 + p_2 = 1, p_1, p_2 \neq 0\}$, $A = 1 - \sum_{i=1}^{m} 2p_i \mid \gamma_i - 0.5 \mid$, $B = \sum_{i=1}^{m} (p_i(\gamma_i - s(h(p))))^2$, and $s(h(p))$ is calculated from Equation (2).

Based on the overall entropy, Xu et al. (2022) presented a new probabilistic hesitant fuzzy multi-attribute decision making method.

Table 6. Overall entropy matrix $\widehat{Z}$

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9
X_1	1.429	1.5388	1.8967	1.3561	0.8	1.2787	0.4	1.4395	1.4758
X_2	1.7898	1.7001	0.8	1.6585	1.9376	1.6295	1.8291	1	1.8578
X_3	1.2595	1	1.2595	1.2595	1.2991	0.4	1.4351	1.3174	1.3697
X_4	0.2	0.4	1.6228	1.4984	2.4328	0	1.8499	1.6736	1.9376
X_5	1.2595	1.2397	0.6	1.8967	1.2787	1.9169	1.4984	1.8483	1.6391
X_6	1.4351	0.2	0.4	1.5185	1.6947	1.3191	0.6	1.3799	0.8
X_7	1.6976	1.6736	1.6323	1.4351	1.2397	0.6	1.2397	1.7589	1.6614
X_8	0.6	1.7177	1.9376	1.7241	0.4	1.7504	1.5923	1.4351	1.8771

End of Table 6

	C_{10}	C_{11}	C_{12}	C_{13}	C_{14}	C_{15}	C_{16}	C_{17}	C_{18}
X_1	0.6	1.5923	1.6791	0.6	1.2397	0.4	1.2397	0.2	1.5798
X_2	1.6779	0.2	1.7889	1.7382	1.3393	1.4758	1.429	1.7889	1.4395
X_3	1.6323	1.4776	0.6	1.3174	1.5185	1.7889	1.4776	1.7504	0.4
X_4	1.7586	0.2	1.6549	1.7586	1.2595	1.6877	0	1.4395	1.6779
X_5	1.2595	1.3952	1.4688	1.2298	1.4758	1.2397	0.6	1.4297	1.3596
X_6	2.6529	1.2397	1.5695	0.4	1.2397	0	1.2793	0.2	1.6121
X_7	1.6791	0.4	1.8019	1.6763	0.8	1.3191	1.2793	1.5388	0.2
X_8	1.6228	1.9145	1.3799	1.2991	1.3596	1.3952	1.7889	1.4984	1.6736

	C_{19}	C_{20}	C_{21}	C_{22}	C_{23}	C_{24}	C_{25}	C_{26}	C_{27}
X_1	1.2199	1.5923	0.2	1.7921	1.4395	1.2595	1.2397	1.2199	0.2
X_2	1.3191	0.6	0	1	0.4	1.4786	0.6	1.7317	1.4395
X_3	0.6	1.4393	1.8388	1.2397	1.5454	1.2595	0.2	1.3596	0.6
X_4	1.4492	1.5698	1.4198	1.6317	1.6323	0	1.5173	1.429	0
X_5	0	1.2793	0.2	0.4	0.4	1.8291	1.2397	0.6	1.2397
X_6	0.2	0.4	1.2397	1.8711	1.8684	1.5923	0.2	1.3393	0
X_7	1.6323	1.6528	2.275	1.5534	0.6	1.5388	1.3191	1.4395	1.3174
X_8	1.2397	1.6736	0.4	1.9376	1.7589	1.9585	0.6	1.9169	0.2

Step 1. Calculate the overall entropy of each PHFE in matrix Z according to Equation (8), and matrix $\widehat{Z}$ is obtained as shown in Table 6.

Step 2. Calculate the deviation degree between the overall entropies under each attribute according to Equation (9).

$$d_j = \frac{2}{m(m-1)} \sum_{l=k+1}^{m} \sum_{k=1}^{m-1} | E_O(h_{lj}(p)) - E_O(h_{kj}(p)) |. \tag{9}$$

Hence, we derive

$d_1 = 0.6183, d_2 = 0.6944, d_3 = 0.7181, d_4 = 0.2497, d_5 = 0.7608, d_6 = 0.8321, d_7 = 0.6220, d_8 = 0.3225, d_9 = 0.4167,$
$d_{10} = 0.6100, d_{11} = 0.7983, d_{12} = 0.3991, d_{13} = 0.5914, d_{14} = 0.2375, d_{15} = 0.7034, d_{16} = 0.6242, d_{17} = 0.6879, d_{18} = 0.6290,$
$d_{19} = 0.7089, d_{20} = 0.5446, d_{21} = 1.0221, d_{22} = 0.6019, d_{23} = 0.7241, d_{24} = 0.6291, d_{25} = 0.6206, d_{26} = 0.4338, d_{27} = 0.7208,$

Step 3. According to equation (10), the weights of each attribute can be determined as follows.

$$w_j = \frac{d_j}{\sum_{j=1}^{n} d_j}, \tag{10}$$

$w = (0.0374, 0.042, 0.0435, 0.0151, 0.046, 0.0504, 0.0376, 0.0195, 0.0252, 0.0369, 0.0483, 0.0242, 0.0358, 0.0144,$
$0.0426, 0.0378, 0.0416, 0.0381, 0.0429, 0.033, 0.0619, 0.0364, 0.0438, 0.0381, 0.0376, 0.0263, 0.0436),$

Step 4. Aggregate the decision information for each alternative according to Equation (11).

$$PHFOWG(h_1(p), h_2(p), \cdots, h_n(p)) = \overset{n}{\underset{j=1}{\otimes}} (\hat{h}_j(p))^{w_j} = \left\{ \prod_{j=1}^{n} \left(\hat{\gamma}^{j}_{\sigma(t)} \right)^{w_j} \left(\sum_{j=1}^{n} w_j \hat{p}^{j}_{\sigma(t)} \right) \middle| t = 1, 2, \cdots, k \right\}. \tag{11}$$

Here, $\hat{h}_j(p)$ is derived by normalizing $h_j(p)$. $\hat{\gamma}^{j}_{\sigma(t)}$ denotes the tth largest membership degree in $\hat{h}_j(p)$, and $\hat{p}^{j}_{\sigma(t)}$ is its occurrence probability. The aggregation results are as follows.

$PHFOWG(h_{11}(p), h_{12}(p), \cdots, h_{1,27}(p)) = \{0.3436(0.5691), 0.2397(0.4309)\}$,
$PHFOWG(h_{21}(p), h_{22}(p), \cdots, h_{2,27}(p)) = \{0(1)\}$,
$PHFOWG(h_{31}(p), h_{32}(p), \cdots, h_{3,27}(p)) = \{0.6983(0.6902), 0.5809(0.3098)\}$,
$PHFOWG(h_{41}(p), h_{42}(p), \cdots, h_{4,27}(p)) = \{0(1)\}$, $PHFOWG(h_{51}(p), h_{52}(p), \cdots, h_{5,27}(p)) = \{0(1)\}$,
$PHFOWG(h_{61}(p), h_{62}(p), \cdots, h_{6,27}(p)) = \{0(1)\}$,
$PHFOWG(h_{71}(p), h_{72}(p), \cdots, h_{7,27}(p)) = \{0.5233(0.5742), 0.3458(0.3825), 0.3313(0.0433)\}$,
$PHFOWG(h_{81}(p), h_{82}(p), \cdots, h_{8,27}(p)) = \{0.3545(0.6464), 0.2409(0.3536)\}$.

Step 5. Calculate the score of each alternative based on Equation (2).
$s(X_1) = 0.2988, s(X_2) = 0, s(X_3) = 0.6619, s(X_4) = 0, s(X_5) = 0, s(X_6) = 0, s(X_7) = 0.4471, s(X_8) = 0.3143$.

Step 6. According to the score values, the severity of 8 medical fraud events is ranked as below.

$$X_2 \sim X_4 \sim X_5 \sim X_6 \prec X_1 \prec X_8 \prec X_7 \prec X_3.$$

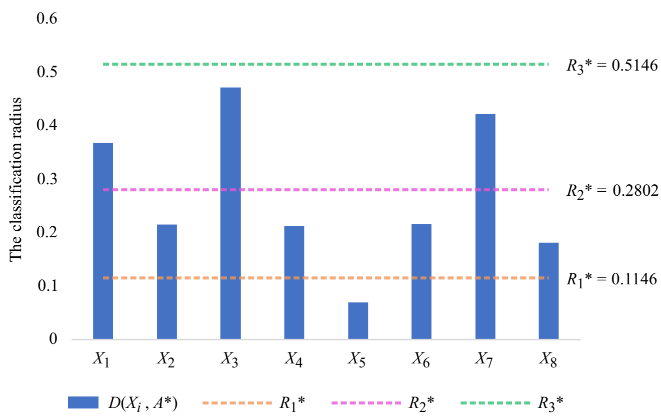

Figure 4. The classification results by the proposed method

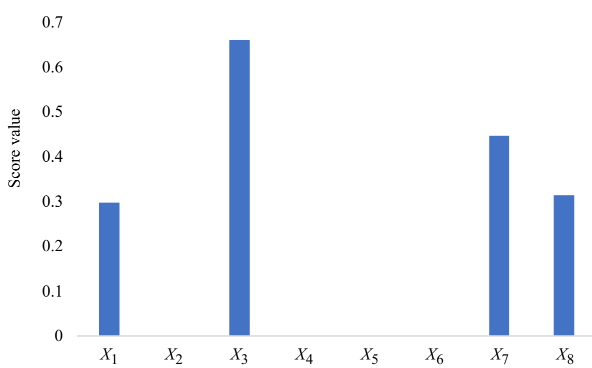

Figure 5. The decision results by Xu et al method (2022)

The decision results obtained by different methods are shown in Figures 4 and 5. In Figure 4, the height of the bar chart represents the severity of the corresponding medical insurance fraud event, so we can easily give a ranking on the severity of the fraud event. Furthermore, according to the distribution discrepancy degree $D(X_i, A^*)$ and the classification radius R^*, we can intuitively determine the classification of the alternatives. However, Xu et al method in Figure 5 cannot achieve the above two points.

5.2. Comparison with Liu and Guo's method (2022)

Liu and Guo (2022) proposed a distance measure integrating hesitancy degree, incompleteness degree and improved difference measure, based on which a probabilistic hesitant fuzzy multi-attribute decision-making method is presented. The specific process is as follows:

Step 1. Determine the positive and negative ideal solution.

$$X^+ = \left\{\begin{array}{l} \langle 0.8(0.3), 0.9(0.7)\rangle, \langle 0.7(0.7), 0.8(0.3)\rangle, \langle 0.8(0.3), 0.9(0.7)\rangle, \langle 0.8(0.3), 0.9(0.7)\rangle, \langle 0.8(0.5), 0.9(0.5)\rangle, \\ \langle 0.8(1)\rangle, \langle 0.8(0.2), 0.9(0.8)\rangle, \langle 0.7(0.3), 0.9(0.7)\rangle, \langle 0.8(0.85), 0.9(0.15)\rangle, \langle 0.7(1)\rangle, \langle 0.8(1)\rangle, \\ \langle 0.8(0.9), 0.9(0.1)\rangle, \langle 0.7(0.3), 0.9(0.7)\rangle, \langle 0.7(0.6), 0.8(0.4)\rangle, \langle 0.3(0.5), 0.5(0.5)\rangle, \langle 0.8(0.4), 0.9(0.6)\rangle, \\ \langle 0.5(0.4), 0.7(0.6)\rangle, \langle 0.9(1)\rangle, \langle 0.2(0.4), 0.4(0.6)\rangle, \langle 0.6(0.1), 0.8(0.9)\rangle, \langle 0.4(0.8), 0.5(0.2)\rangle, \\ \langle 0.8(0.2), 0.9(0.8)\rangle, \langle 0.8(1)\rangle, \langle 0.8(0.3), 0.9(0.7)\rangle, \langle 0.9(1)\rangle, \langle 0.8(0.8), 0.9(0.2)\rangle, \langle 0.3(1)\rangle \end{array}\right\},$$

$$X^- = \left\{\begin{array}{l} \langle 0.1(1)\rangle, \langle 0.1(1)\rangle, \langle 0.2(1)\rangle, \langle 0.2(0.5), 0.3(0.5)\rangle, \langle 0.1(0.9), 0.3(0.1)\rangle, \langle 0(1)\rangle, \langle 0.2(0.5), 0.3(0.5)\rangle, \\ \langle 0.2(0.8), 0.3(0.2)\rangle, \langle 0.1(0.3), 0.3(0.7)\rangle, \langle 0.1(0.7), 0.2(0.3)\rangle, \langle 0.1(1)\rangle, \langle 0.2(0.65), 0.3(0.35)\rangle, \\ \langle 0.1(0.85), 0.2(0.15)\rangle, \langle \{0.1(0.9), 0.3(0.1)\}\rangle, \langle 0(1)\rangle, \langle 0(1)\rangle, \langle 0.1(1)\rangle, \langle 0.1(0.2), 0.2(0.8)\rangle, \langle 0(1)\rangle, \\ \langle 0.1(0.6), 0.2(0.4)\rangle, \langle 0(1)\rangle, \langle 0.2(1)\rangle, \langle 0.2(1)\rangle, \langle 0(1)\rangle, \langle 0.1(1)\rangle, \langle 0.1(0.9), 0.2(0.1)\rangle, \langle 0(1)\rangle \end{array}\right\}.$$

Step 2. Obtain the attribute weights as follows (Liu & Guo, 2022).

$$w_1 = 0.0276, w_2 = 0.0326, w_3 = 0.0306, w_4 = 0.0346, w_5 = 0.0425, w_6 = 0.0347, w_7 = 0.0431, w_8 = 0.0348,$$
$$w_9 = 0.0436, w_{10} = 0.0540, w_{11} = 0.0429, w_{12} = 0.0416, w_{13} = 0.0454, w_{14} = 0.0497, w_{15} = 0.0214,$$
$$w_{16} = 0.0265, w_{17} = 0.0279, w_{18} = 0.0451, w_{19} = 0.0286, w_{20} = 0.0428, w_{21} = 0.0358, w_{22} = 0.0288,$$
$$w_{23} = 0.0355, w_{24} = 0.0242, w_{25} = 0.0424, w_{26} = 0.0414, w_{27} = 0.0421.$$

Step 3. The alternatives are ranked according to the compromise ratio method, and the weighted distances of each medical fraud event to the positive ideal solution X^+ and negative ideal solution X^- are calculated as shown in Table 7.

Table 7. Distances from medical fraud event x_i to X^+ and X^-

X_i	$d(X_i, X^+)$	$d(X_i, X^-)$
X_1	0.3459	0.3029
X_2	0.3304	0.2365
X_3	0.1444	0.4305
X_4	0.3719	0.1567
X_5	0.4093	0.1665
X_6	0.4496	0.2398
X_7	0.2213	0.3490
X_8	0.3318	0.2380

The compromise ratio for each medical fraud event x_i is derived as below: $R\left(X_1\right)=0.4369$, $R\left(X_2\right)=0.3410$, $R\left(X_3\right)=1$, $R\left(X_4\right)=0.1273$, $R\left(X_5\right)=0.0839$, $R\left(X_6\right)=0.1518$, $R\left(X_7\right)=0.7252$, $R\left(X_8\right)=0.3415$. Hence, the severity of medical fraud events is ranked as below:

$X_3 \succ X_7 \succ X_1 \succ X_8 \succ X_2 \succ X_6 \succ X_4 \succ X_5$.

The decision results obtained by different methods are shown in Figures 4–6. In Figure 4, we can find that the proposed method can not only produce a ranking on the severity of the fraud event, but also determine the classification of the alternatives. However, we cannot derive the classification of the alternatives by Liu and Guo's method (2022) (See Figure 6). By comparison, we can find that there are the following differences among the three methods.

(1) The attribute weights are determined in different ways. In this paper, we mainly use CBR to determine the attribute weights according to the test decision matrix, while Xu et al. (2022) adopted the overall entropy method based on the original decision matrix for deriving the attribute weights. Liu and Guo (2022) developed a mathematical programming model based on the improved distance to determine the attribute weights. The method proposed in this paper takes into account the decision information and experts' own experience, while the other two methods only considered the decision information and ignored the subjective will of decision makers. Hence, the results by their method may not be consistent with the facts.

(2) The aggregation techniques among them are different. In the test decision problem, the proposed method aggregates the decision information based on the distribution discrepancy degree between the alternative and historical optimal case. Furthermore, it is not necessary to add elements into the shorter PHFEs when using the proposed method, and thus can avoid the change of decision information. In contrast, when Xu et al. (2022) adopted the aggregation operator for aggregating decision information, they needed to add elements into the shorter PHFEs. Moreover, as long as there exists an evaluation value $\{0(1)\}$, then the aggregation result is still $\{0(1)\}$, which makes it unable to distinguish between X_2, X_4, X_5 and X_6. That's to say, when experts believe that the medical insurance fraud events do not involve any one of these types of behavior, these fraud events will be regarded as the same in the severity, which is obviously unreasonable. In Liu and Guo's method (2022), additional elements are added into the shorter PHFEs when calculating the distance, which will inevitably affect the decision results.

(3) The ability to classify alternatives into different categories is different. The method by Xu et al. (2022) is not only unable to rank the social medical insurance fraud events reasonably, but also cannot classify them into different categories. Likewise, the classification of the alternatives cannot be obtained by Liu and Guo's method (2022). In fact, for different types of medical insurance fraud events, the means of prevention and combating are also different. If the severity of fraud events cannot be determined, various social departments cannot provide rapid and effective response measures. Therefore, the other two methods cannot solve the problem of assessing the severity of medical insurance fraud well in practical applications. The proposed method in this paper is able to classify and rank the alternatives simultaneously, thus assisting decision makers to make rational decisions.

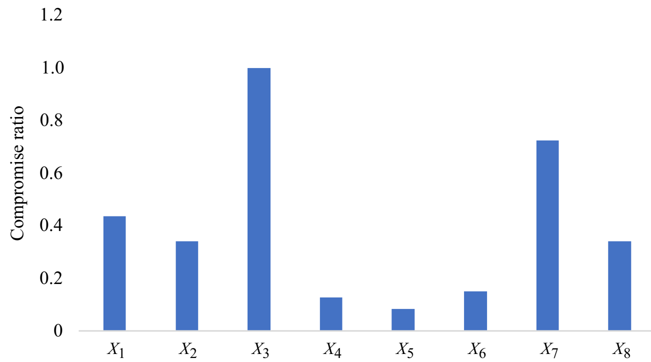

Figure 6. The decision results by Liu and Guo's method (2022)

6. Discussion

Considering the complexity and diversification of fraud means and many other uncertain factors in the process of identifying fraudulent behaviors, this paper adopts the PHFSs to address the uncertainty and randomness, based on which a new CBR method is presented and applied to the classification of medical insurance fraud.

The proposed method not only realizes the prioritization of the severity of fraud in medical insurance, but also can classify the fraud cases into different categories, which is beneficial for government department to perform its own responsibilities, focus on key areas and conduct classified strikes. In addition, the utilization of medical resources can be improved greatly. For example, for mild fraud events, the integrity of insurance publicity and education for all types of fraud subjects can be strengthened to raise the awareness of participants about the seriousness of fraud and prevent the recurrence of such incidents. Moderate fraud and severe fraud events can be deeply analyzed and exposed through newspapers, the network, radio and other media, so as to guide wide attention and supervision of all sectors of society, and establish different levels of punishment mechanisms for medical insurance fraud. For extreme fraud events, the various units and individuals who defrauded the medical insurance fund should be blacklisted and even sentenced and convicted in order to put an end to such phenomena. Therefore, the proposed method can classify medical insurance fraud events into several categories according to different fraud levels, and provide a basis for formulating corresponding prevention and combating strategies to ensure the safe operation of medical insurance funds.

7. Conclusions

This paper defines the distribution function for PHFE, based on which the formula of the distribution discrepancy degree between PHFEs is proposed. It does not need the same number of elements in the compared PHFEs, i.e., no additional information is added when using the proposed method. In addition, the distribution discrepancy degree is introduced to CBR for classifying and identifying the severity of medical insurance fraud, which overcomes the defect of traditional methods that the knowledge and experience of experts are neglected.

The decision method in this paper is conducted using previous relevant cases, which can well deal with the medical insurance fraud. Compared with the traditional decision-making methods, the proposed method can not only maximize the advantages of experts' own knowledge and experience, but also make full use of objective decision-making data, and thus the derived results are more reasonable and closer to reality. In the future work, we will continue to investigate the uncertainty measures between PHFEs, such as distance, similarity and entropy, and apply the proposed method to other fields, such as machine learning and pattern recognition.

Acknowledgements

This work was supported by the National Social Science Foundation of China (Nos. 22BGL211 and 22AJL002), the National Natural Science Foundation of China (No. 72074001) and the Anhui Provincial Natural Science Foundation (No. 2108085MG240).

Disclosure statement

The authors have no competing interests to declare that are relevant to the content of this article.

References

Aamodt, A., & Plaza, E. (1994). Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI communications*, *7*(1), 39–59. https://doi.org/10.3233/AIC-1994-7104

Ataabadi, P. E., Neysiani, B. S., Nogorani, M. Z., & Mehraby, N., (2022, May 11–12). Semi-supervised medical insurance fraud detection by predicting indirect reductions rate using machine learning generalization capability. In *Proceedings of the 2022 8th International Conference on Web Research* (*ICWR*) (pp. 176–182). Tehran, Iran, Islamic Republic of. IEEE. https://doi.org/10.1109/ICWR54782.2022.9786251

Atanassov, K. T. (1986). Intuitionistic fuzzy sets. *Fuzzy Sets and Systems*, *20*(1), 87–96. https://doi.org/10.1016/S0165-0114(86)80034-3

Cao, Q., Liu, X. D., Wang, Z. W., Zhang, S. T., & Wu, J. (2020). Recommendation decision-making algorithm for sharing accommodation using probabilistic hesitant fuzzy sets and bipartite network projection. *Complex & Intelligent Systems*, *6*, 431–445. https://doi.org/10.1007/s40747-020-00142-7

Chen, H. Y., Wu, Z. Y., Chen, T. L., Huang, Y. M., & Liu, C. H. (2021). Security privacy and policy for cryptographic based electronic medical information system. *Sensors*, *21*(3), Article 713. https://doi.org/10.3390/s21030713

Divsalar, M., Ahmadi, M., Ebrahimi, E., & Ishizaka, A. (2022). A probabilistic hesitant fuzzy Choquet integral-based TODIM method for multi-attribute group decision-making. *Expert Systems with Applications*, *191*, Article 116266. https://doi.org/10.1016/j.eswa.2021.116266

Fan, Z. P., Li, Y. H., Wang, X., & Liu, Y. (2014). Hybrid similarity measure for case retrieval in CBR and its application to emergency response towards gas explosion. *Expert Systems with Applications*, *41*(5), 2526–2534. https://doi.org/10.1016/j.eswa.2013.09.051

Gorzałczany, M. B. (1987). A method of inference in approximate reasoning based on interval-valued fuzzy sets. *Fuzzy Sets and Systems*, *21*(1), 1–17. https://doi.org/10.1016/0165-0114(87)90148-5

Guida, S. (2021). Ransomware attacks are data breaches and finally started to be reported properly by major companies. *European Journal of Privacy Law & Technologies*, *7*, 1–3. https://universitypress.unisob.na.it/ojs/index.php/ejplt/article/view/1374

Han, X. R., & Zhan, J. M. (2023). A sequential three-way decision-based group consensus method under probabilistic linguistic term sets. *Information Sciences*, *624*, 567–589. https://doi.org/10.1016/j.ins.2022.12.111

Han, X. R., Zhang, C., & Zhan, J. M. (2022). A three-way decision method under probabilistic linguistic term sets and its application to Air Quality Index. *Information Sciences*, *617*, 254–276. https://doi.org/10.1016/j.ins.2022.10.108

Jiang, J. C., Liu, X. D., Garg, H., & Zhang, S. T. (2023). Large group decision-making based on interval rough integrated cloud model. *Advanced Engineering Informatics*, *56*, Article 101964. https://doi.org/10.1016/j.aei.2023.101964

Jiang, J. C., Liu, X. D., Wang, Z. W., Ding, W. P., & Zhang, S. T. (2024). Large group emergency decision-making with bi-directional trust in social networks: A probabilistic hesitant fuzzy integrated cloud approach. *Information Fusion*, *102*, Article 102062. https://doi.org/10.1016/j.inffus.2023.102062

Jiang, X. S., Lin, K. B., Zeng, Y. F., & Yang, F., (2021, August, 17–21). Medical insurance medication anomaly detection based on isolated forest proximity matrix. In *Proceedings of the 2021 16th International Conference on Computer Science & Education* (*ICCSE*) (pp. 512–517). Lancaster, United Kingdom. IEEE. https://doi.org/10.1109/ICCSE51940.2021.9569723

Kapadiya, K., Patel, U., Gupta, R., Alshehri, M. D., Tanwar, S., Sharma, G., & Bokoro, P. N. (2022). Blockchain and AI-empowered healthcare insurance fraud detection: An analysis, architecture, and future prospects. *IEEE Access*, *10*, 79606–79627. https://doi.org/10.1109/ACCESS.2022.3194569

Krishankumar, R., Ravichandran, K. S., Liu, P. D., Kar, S., & Gandomi, A. H. (2021). A decision framework under probabilistic hesitant fuzzy environment with probability estimation for multi-criteria decision making. *Neural Computing and Applications*, *33*(14), 8417–8433. https://doi.org/10.1007/s00521-020-05595-y

Li, J., Lan, Q. L., Zhu, E. Y., Xu, Y., & Zhu, D. (2022). A study of health insurance fraud in China and recommendations for fraud detection and prevention. *Journal of Organizational and End User Computing* (*JOEUC*), *34*(4), 1–19. https://doi.org/10.4018/JOEUC.301271

Li, P., Liu, J., Yang, Y. J., & Wei, C. P. (2020). Evaluation of poverty-stricken families in rural areas using a novel case-based reasoning method for probabilistic linguistic term sets. *Computers & Industrial Engineering*, *147*, Article 106658. https://doi.org/10.1016/j.cie.2020.106658

Li, P., & Wei, C. P. (2018). A case-based reasoning decision-making model for hesitant fuzzy linguistic information. *International Journal of Fuzzy Systems*, *20*, 2175–2186. https://doi.org/10.1007/s40815-017-0391-1

Liao, T. W., Zhang, Z., & Mount, C. R. (1998). Similarity measures for retrieval in case-based reasoning systems. *Applied Artificial Intelligence*, *12*(4), 267–288. https://doi.org/10.1080/088395198117730

Liao, N. N., Wei, G. W., & Chen, X. D. (2022). TODIM method based on cumulative prospect theory for multiple attributes group decision making under probabilistic hesitant fuzzy setting. *International Journal of Fuzzy Systems*, *24*(1), 322–339. https://doi.org/10.1007/s40815-021-01138-2

Liu, X. D., Wang, Z. W., Zhang, S. T., & Garg, H. (2021). Novel correlation coefficient between hesitant fuzzy sets with application to medical diagnosis. *Expert Systems with Applications*, *183*, Article 115393. https://doi.org/10.1016/j.eswa.2021.115393

Liu, X. D., Wu, J., Zhang, S. T., Wang, Z. W., & Garg, H. (2022). Extended cumulative residual entropy for emergency group decision-making under probabilistic hesitant fuzzy environment. *International Journal of Fuzzy Syst*ems, *24*(1), 159–179. https://doi.org/10.1007/s40815-021-01122-w

Liu, S. J., & Guo, Z. X. (2022). Probabilistic hesitant fuzzy multi-attribute decision-making method based on improved distance measurement. *Journal of Intelligent & Fuzzy Systems*, *43*(5), 5953–5964. https://doi.org/10.3233/JIFS-213427

Loève, M. (2017). *Probability theory*. Courier Dover Publications.

Löw, N., Hesser, J., & Blessing, M. (2019). Multiple retrieval case-based reasoning for incomplete datasets. *Journal of Biomedical Informatics*, *92*, Article 103127. https://doi.org/10.1016/j.jbi.2019.103127

Naeem, M., Khan, M. A., Abdullah, S., Qiyas, M., & Khan, S. (2021). Extended TOPSIS method based on the entropy measure and probabilistic hesitant fuzzy information and their application in decision support system. *Journal of Intelligent & Fuzzy Systems*, *40*(6), 11479–11490. https://doi.org/10.3233/JIFS-202700

Schank, R. C. (1983). *Dynamic memory: A theory of reminding and learning in computers and people*. Cambridge University Press.

Settipalli, L., & Gangadharan, G. R. (2023). Provider profiling and labeling of fraudulent health insurance claims using Weighted MultiTree. *Journal of Ambient Intelligence and Humanized Computing*, *14*, 3487–3508. https://doi.org/10.1007/s12652-021-03481-6

Sha, X. Y., Yin, C. C., Xu, Z. S., & Zhang, S. (2021). Probabilistic hesitant fuzzy TOPSIS emergency decision-making method based on the cumulative prospect theory. *Journal of Intelligent & Fuzzy Systems*, *40*(3), 4367–4383. https://doi.org/10.3233/JIFS-201119

Song, C. Y., Xu, Z. S., & Zhao, H. (2019). New correlation coefficients between probabilistic hesitant fuzzy sets and their applications in cluster analysis. *International Journal of Fuzzy Systems*, *21*(2), 355–368. https://doi.org/10.1007/s40815-018-0578-0

Song, H. F., & Chen, Z. C. (2021). Multi-attribute decision-making method based distance and COPRAS method with probabilistic hesitant fuzzy environment. *International Journal of Computational Intelligence Systems*, *14*(1), 1229–1241. https://doi.org/10.2991/ijcis.d.210318.001

Torra, V. (2010). Hesitant fuzzy sets. *International Journal of Intelligent Systems*, *25*(6), 529–539. https://doi.org/10.1002/int.20418

Villegas-Ortega, J., Bellido-Boza, L., & Mauricio, D. (2021). Fourteen years of manifestations and factors of health insurance fraud, 2006–2020: A scoping review. *Health & Justice*, *9*, 1–23. https://doi.org/10.1186/s40352-021-00149-3

Wang, D. L., Wan, K. D., & Ma, W. X. (2020). Emergency decision-making model of environmental emergencies based on case-based reasoning method. *Journal of Environmental Management*, *262*, Article 110382. https://doi.org/10.1016/j.jenvman.2020.110382

Wang, Y. B., Jia, X. L., & Zhang, L. X. (2022). Evaluation of the survival of Yangtze finless porpoise under probabilistic hesitant fuzzy environment. *International Journal of Intelligent Systems*, *37*(10), 7665–7684. https://doi.org/10.1002/int.22898

Wang, Z. X., & Li, J. (2017). Correlation coefficients of probabilistic hesitant fuzzy elements and their applications to evaluation of the alternatives. *Symmetry*, *9*(11), Article 259. https://doi.org/10.3390/sym9110259

Wu, W. Y., Ni, Z. W., Jin, F. F., Wu, J., Li, Y., & Li, P. (2021). Investment selection based on Bonferroni mean under generalized probabilistic hesitant fuzzy environments. *Mathematics*, *9*(1), Article 107. https://doi.org/10.3390/math9010107

Xu, T. T., Zhang, H., & Li, B. Q. (2022). Fuzzy entropy and hesitancy entropy in probabilistic hesitant fuzzy information and their applications. *Soft Computing*, *26*(18), 9101–9115. https://doi.org/10.1007/s00500-022-07309-z

Xu, X. H., Huang, Y. X., & Chen, K. (2019). Method for large group emergency decision making with complex preferences based on emergency similarity and interval consistency. *Natural Hazards*, *97*, 45–64. https://doi.org/10.1007/s11069-019-03624-1

Xu, Z. S., & Zhou, W. (2017). Consensus building with a group of decision makers under the hesitant probabilistic fuzzy environment. *Fuzzy Optimization and Decision Making*, *16*, 481–503. https://doi.org/10.1007/s10700-016-9257-5

Yu, F., Li, X. Y., & Han, X. S. (2018). Risk response for urban water supply network using case-based reasoning during a natural disaster. *Safety Science*, *106*, 121–139. https://doi.org/10.1016/j.ssci.2018.03.003

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, *8*(3), 338–353. https://doi.org/10.1016/S0019-9958(65)90241-X

Zhan, J. M., Wang, J. J., Ding, W. P., & Yao, Y. Y. (2022). Three-way behavioral decision making with hesitant fuzzy information systems: Survey and challenges. *IEEE/CAA Journal of Automatica Sinica*, *10*(2), 330–350. https://doi.org/10.1109/JAS.2022.106061

Zhang, G. M., Zhang, X. Y., Bilal, M., Dou, W. C., Xu, X. L., & Rodrigues, J. J. (2022). Identifying fraud in medical insurance based on blockchain and deep learning. *Future Generation Computer Systems*, *130*, 140–154. https://doi.org/10.1016/j.future.2021.12.006

Zhang, J. J. (2021). Behavior types and legal governance of medical insurance fraud. *Social Sciences Review*, *36*(4), 123–129. https://doi.org/10.16745/j.cnki.cn62-1110/c.2021.04.021

Zhang, S. S., Liu, X. D., Garg, H., & Zhang, S. T. (2023). Investment decision making in the fuzzy context: An integrated model approach. *Journal of Intelligent & Fuzzy Systems*, *44*(3), 3763–3786. https://doi.org/10.3233/JIFS-223059

Zhang, W. M., Liu, X. Y., Zhang, X. Y., Hu, W. H., Zhang, J. C., & Shao, W. Y., (2022, May 6–8). Medicare fraud gang discovery based on community discovery algorithms. In *Proceedings of the 2022 IEEE 8th Intl Conference on Big Data Security on Cloud* (*BigDataSecurity*), *IEEE Intl Conference on High Performance and Smart Computing*, (*HPSC*) *and IEEE Intl Conference on Intelligent Data and Security* (*IDS*) (pp. 206–211). Jinan, China. IEEE. https://doi.org/10.1109/BigDataSecurityHPSCIDS54978.2022.00047

Zheng, J., Wang, Y. M., & Chen, S. Q. (2016). Dynamic case retrieval method with subjective preferences and objective information for emergency decision making. *IEEE/CAA Journal of Automatica Sinica*, *5*(3), 749–757. https://doi.org/10.1109/JAS.2016.7510232

Zhou, X. Y., & Su, W. H. (2021). Exploring the audit supervision of China's medical insurance fund based on the 2019 medical insurance fund audit data. *Accounting & Finance Research*, *10*, 78–86. https://doi.org/10.5430/AFR.V10N2P78

Zhu, B. (2014). *Decision method for research and application based on preference relation*. Southeast University, Nanjing, China.

Zhu, J. X., Ma, X. L., Kou, G., Herrera-Viedma, E., & Zhan, J. M. (2023a). A three-way consensus model with regret theory under the framework of probabilistic linguistic term sets. *Information Fusion*, *95*, 250–274. https://doi.org/10.1016/j.inffus.2023.02.029

Zhu, J. X., Ma, X. L., Martínez, L., & Zhan, J. M. (2023b). A probabilistic linguistic three-way decision method with regret theory via fuzzy c-means clustering algorithm. *IEEE Transactions on Fuzzy Systems*, *31*(8), 2821–2835. https://doi.org/10.1109/TFUZZ.2023.3236386