Serdica J. Computing **10** (2016), No 1, 13–30

**Serdica**
Journal of Computing

Bulgarian Academy of Sciences
Institute of Mathematics and Informatics

# A METHOD TO CONSTRUCT AN EXTENSION OF FUZZY INFORMATION GRANULARITY BASED ON FUZZY DISTANCE*

Nguyen Van Thien, Janos Demetrovics, Vu Duc Thi, Nguyen Long Giang, Nguyen Nhu Son

Abstract. In fuzzy granular computing, a fuzzy granular structure is the collection of fuzzy information granules and fuzzy information granularity is used to measure the granulation degree of a fuzzy granular structure. In general, the fuzzy information granularity characterizes discernibility ability among fuzzy information granules in a fuzzy granular structure. In recent years, researchers have proposed some concepts of fuzzy information granularity based on partial order relations. However, the existing forms of fuzzy information granularity have some limitations when evaluating the fineness/coarseness between two fuzzy granular structures. In this paper, we propose an extension of fuzzy information granularity based on a fuzzy distance measure. We prove theoretically and experimentally that the proposed fuzzy information granularity is the best one to distinguish fuzzy granular structures.

*ACM Computing Classification System* (1998): I.5.2, I.2.6.

*Key words:* granular computing, fuzzy granular structure, fuzzy information granule, fuzzy information granularity, fuzzy distance.

*This research has been funded by the research projects VAST01.08/16–17 (Vietnam Academy of Science and Technology) and QG.15.41 (Vietnam National University).

**1. Introduction.** Granular computing (GrC), which was proposed by Zadeh in 1996 [13], is an important research direction in information processing and widely applied in many computer science fields such as approximate reasoning, artificial intelligence, data mining, machine learning. In GrC, a binary relation divides a data set into some information granules. Each information granule is a set of objects that are indistinguishability, similarity, and proximity of functionality.

In Pawlak's rough set theory [8], a given equivalence relation divides a data set into equivalence classes or concepts. According to the granular computing approach, the set of those equivalence classes is called a granular structure and each equivalence class is called an information granule. For a tolerance rough set model on an incomplete information system [4], each tolerance relation determines a covering on the object set, in which each element is a tolerance class. That covering is called a tolerance granular structure and each tolerance class is called a tolerance granule. In a fuzzy rough set model [1, 2], the equivalence relation in Pawlak's rough set is extended to a fuzzy similarity relation. Each fuzzy similarity relation determines a fuzzy partition on the object set. The fuzzy partition is called a fuzzy granular structure and each fuzzy similarity class in the fuzzy partition is called a fuzzy information granule.

In granular computing, information granularity is a measure of the granulation degree of a universe based on a given binary relation. Information granularity represents the fineness or roughness of granular structure. The smaller the information granularity is, the finer a granular structure is. In general, the information granularity characterizes discernibility ability or the difference among information granules in a granular structure and it has been used effectively in approximation problems, data mining and machine learning. In the past two decades, information granularity has attracted the attention of many granular computing researchers. Many concepts of information granularity were proposed by many different approaches [5, 6, 7, 11, 12]. Wierman [11] introduced the concept of information granularity based on Shannon's entropy in complete information systems. Liang et al. [5, 6] proposed the concept of information granularity based on Liang's entropy in complete and incomplete information systems. Qian and Liang [7] introduced the concept of information granularity based on combination entropy. Xu et al. [12] presented information granularity based on the improvement of roughness in rough set theory.

Fuzzy information granularity is the information granularity of a fuzzy granular structure. Qian et al. [9] proposed two fuzzy information granularities in order to evaluate the fineness/roughness of a fuzzy granular structure based on three partial order relations. The first fuzzy information granularity is the extension of information granularity in [3], the second fuzzy information granularity is the extension of rough entropy in [5]. However, both fuzzy information granularities based on partial order relations experience limitations in assessing the fineness or roughness of a fuzzy granular structure in some special cases.

In this paper, we first build a fuzzy distance between two fuzzy granular structures and discover some properties. Based on the fuzzy distance, we propose extension fuzzy information granularity to overcome the limitations of fuzzy information granularity in [9]. Finally, some experiments are performed on data sets from UCI to evaluate the effectiveness of extension fuzzy information granularity. The paper is organized as follows. Section 2 reviews some basic concepts of fuzzy granular structure space and partial order relations in [9]. Section 3 constructs a fuzzy distance between two fuzzy granular structures and discovers some properties of the fuzzy distance. Section 4 proposes the concept of extension fuzzy information granularity. Section 5 presents the results of experiments on some data sets from UCI. Finally, section 6 gives some conclusions and subsequent developments.

**2. Some basic concepts.** Based on the fuzzy similarity relation in fuzzy rough set [1, 2], in this section we present some basic concepts related to fuzzy information granule, fuzzy granular structure, fuzzy granular structure space [9].

In rough set theory, an information system is denoted by $IS = (U, A)$ where $U$ is the set of objects; $A$ is the set of attributes. Each equivalence relation $IND(P)$ on attribute set $P \in A$ defines a partition on $U$, denoted by $U/IND(P)$. According to the granular computing approach, $U/IND(P)$ is an information granular structure and each equivalence class in $U/IND(P)$ is an information granule. If $a \in A$ contains a missing value then $IS$ is called an incomplete information system. In incomplete information systems, Kryszkiewicz defines a tolerance relation $SIM(P)$ on attribute set $P \in A$. The relation $SIM(P)$ determines a covering on $U$, denoted by $U/SIM(P)$. $U/SIM\left(P\right) = \{S_P\left(x\right)|x \in U\}$ where $S_P\left(x\right)$ is a tolerance class that contains object $x \in U$. $U/SIM(P)$ is called a tolerance granular structure and $S_P\left(x\right)$ is called a tolerance granule. The information granules and tolerance granules are a basic calculation unit to build efficient measures to solve attribute reduction and rule extraction problems. However, the equivalence relation or tolerance relation cannot characterize the similarity among objects in real problems. For that reason, researchers use a fuzzy similarity relation to replace the equivalence relation in the fuzzy rough set [1, 2].

Let $IS = (U, A)$ be an information system, a fuzzy similarity relation $\widetilde{P}$ is defined on the attribute set $P \in A$. Then, $\widetilde{P}$ has properties: reflexive $\widetilde{P}(x, x) = 1$, symmetric $\widetilde{P}(x, y) = \widetilde{P}(y, x)$, max-min transitive $\widetilde{P}(x, z) \geq \min\{\widetilde{P}(x, y), \widetilde{P}(y, z)\}$ for any $x, y, z \in U$. The fuzzy similarity relation $\widetilde{P}$ on attribute set $P \in A$ is represented by the following relation matrix:

$$M(\widetilde{P}) = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix}$$

where $p_{ij} = \widetilde{P}\left(x_i, x_j\right)$ is the value of the relation between object $x_i$ and $x_j$ on attribute set $P$, $r_{ij} \in [0, 1]$.

Given two fuzzy similarity relations $\widetilde{P}$ and $\widetilde{Q}$ on attribute set $P$ and $Q$, for any $x, y \in U$ we have:

1. $\widetilde{P} = \widetilde{Q} \Leftrightarrow \widetilde{P}\left(x, y\right) = \widetilde{Q}\left(x, y\right)$
2. $\widetilde{R} = \widetilde{P} \cup \widetilde{Q} \Leftrightarrow \widetilde{R}\left(x, y\right) = \max\left\{\widetilde{P}\left(x, y\right), \widetilde{Q}\left(x, y\right)\right\}$
3. $\widetilde{R} = \widetilde{P} \cap \widetilde{Q} \Leftrightarrow \widetilde{R}\left(x, y\right) = \min\left\{\widetilde{P}\left(x, y\right), \widetilde{Q}\left(x, y\right)\right\}$
4. $\widetilde{P} \subseteq \widetilde{Q} \Leftrightarrow \widetilde{P}\left(x, y\right) \leq \widetilde{Q}\left(x, y\right)$

According to the granular computing approach, a fuzzy similarity relation $\widetilde{P}$ defines a fuzzy granular structure $K(\widetilde{P})$ in $U$. $K(\widetilde{P}) = \left(S_{\widetilde{P}}(x_1), S_{\widetilde{P}}(x_2), \ldots, S_{\widetilde{P}}(x_n)\right)$ where $S_{\widetilde{P}}(x_i) = p_{i1}/x_1 + p_{i2}/x_2 + \cdots + p_{in}/x_n$ is a fuzzy information granule of object $x_i$. The cardinality of the fuzzy information granule $S_{\widetilde{P}}(x_i)$ is calculated by

$$\left|S_{\widetilde{P}}(x_i)\right| = \sum_{j=1}^{n} p_{ij}$$

Let $K(U)$ be a set of all fuzzy granular structures in $U$ determined by fuzzy similarity relations on attribute sets. $K(U)$ is called a fuzzy granulation structure space. Given a fuzzy granular structure

$$K(\widetilde{P}) = (S_{\widetilde{P}}(x_1), S_{\widetilde{P}}(x_2), \ldots, S_{\widetilde{P}}(x_n)),$$

$$S_{\widetilde{P}}(x_i) = p_{i1}/x_1 + p_{i2}/x_2 + \cdots + p_{in}/x_n.$$

In particular, if $p_{ij} = 0$, $i, j \le n$ then $\left|S_{\widetilde{P}}(x_i)\right| = 0$, $i \le n$ and the fuzzy granular structure $K(\widetilde{P})$ is the finest one, denoted as $K(\widetilde{\omega})$, i.e.,

$$K(\widetilde{\omega}) = (S_{\widetilde{\omega}}(x_1), S_{\widetilde{\omega}}(x_2), \ldots, S_{\widetilde{\omega}}(x_n)),$$

where $S_{\widetilde{\omega}}(x_i) = \sum_{j=1}^{n} \omega_{ij}/x_j$, $\forall i, j \le n$, $\omega_{ij} = 0$. If $p_{ij} = 1$, $i, j \le n$ then $\left|S_{\widetilde{P}}(x_i)\right| = |U|$, $i \le n$, and the fuzzy granular structure $K(\widetilde{P})$ is the coarsest one, denoted as $K(\widetilde{\delta})$, i.e., $K(\widetilde{\delta}) = \left(S_{\widetilde{\delta}}(x_1), S_{\widetilde{\delta}}(x_2), \ldots, S_{\widetilde{\delta}}(x_n)\right)$ where

$$S_{\widetilde{\delta}}(x_i) = \sum_{j=1}^{n} \delta_{ij}/x_j, \ \forall i, j \le n, \ \delta_{ij} = 1.$$

In [9], Qian et al. present three definitions of partial order relations $\preceq_1$, $\preceq_2$, $\preceq_3$ to characterize the uncertainty of a fuzzy granular structure.

**Definition 1.** *Given two fuzzy granular structures $K(\widetilde{P}), K(\widetilde{Q}) \in K(U)$ where*

$$K(\widetilde{P}) = \left(S_{\widetilde{P}}(x_1), S_{\widetilde{P}}(x_2), \ldots, S_{\widetilde{P}}(x_n)\right), \ K(\widetilde{Q}) = \left(S_{\widetilde{Q}}(x_1), S_{\widetilde{Q}}(x_2), \ldots, S_{\widetilde{Q}}(x_n)\right).$$

*The partial order relations are defined as*

1. $K(\widetilde{P}) \preceq_1 K(\widetilde{Q}) \Leftrightarrow S_{\widetilde{P}}(x_i) \subseteq S_{\widetilde{Q}}(x_i),\ i \le n \Leftrightarrow p_{ij} \le q_{ij},\ i, j \le n$, *just* $\widetilde{P} \preceq_1 \widetilde{Q}$. *The equality* $K(\widetilde{P}) = K(\widetilde{Q}) \Leftrightarrow S_{\widetilde{P}}(x_i) = S_{\widetilde{Q}}(x_i),\ i \le n \Leftrightarrow p_{ij} = q_{ij},\ i, j \le n$ *can be written as* $\widetilde{P} = \widetilde{Q}$. $K(\widetilde{P}) \prec_1 K(\widetilde{Q}) \Leftrightarrow K(\widetilde{P}) \preceq_1 K(\widetilde{Q})$ *and* $K(\widetilde{P}) \ne K(\widetilde{Q})$, *which is denoted by* $\widetilde{P} \prec_1 \widetilde{Q}$.

2. $K(\widetilde{P}) \preceq_2 K(\widetilde{Q}) \Leftrightarrow \left|S_{\widetilde{P}}(x_i)\right| \le \left|S_{\widetilde{Q}}(x_i)\right|,\ i \le n$ *where* $\left|S_{\widetilde{P}}(x_i)\right| = \sum_{i=1}^{n} p_{ij}$, $\left|S_{\widetilde{Q}}(x_i)\right| = \sum_{i=1}^{n} q_{ij}$, , *just* $\widetilde{P} \preceq_2 \widetilde{Q}$. *The special case,* $K(\widetilde{P}) \simeq K(\widetilde{Q}) \Leftrightarrow \left|S_{\widetilde{P}}(x_i)\right| = \left|S_{\widetilde{Q}}(x_i)\right|,\ i \le n$ *can be written as* $\widetilde{P} \simeq \widetilde{Q}$. $K(\widetilde{P}) \prec_2 K(\widetilde{Q}) \Leftrightarrow K(\widetilde{P}) \preceq_2 K(\widetilde{Q})$ *and* $K(\widetilde{P}) \ne K(\widetilde{Q})$ *can be written as* $\widetilde{P} \prec_2 \widetilde{Q}$.

3. $K(\widetilde{P}) \preceq_3 K(\widetilde{Q}) \Leftrightarrow$ *for* $K(\widetilde{P})$, *there exists a sequence* $K^1(\widetilde{Q})$ *of* $K(\widetilde{Q})$ *such that* $\left|S_{\widetilde{P}}(x_i)\right| \le \left|S_{\widetilde{Q}}\left(x_i^1\right)\right|,\ i \le n$ *where* $K^1(\widetilde{Q}) = \left(S_{\widetilde{Q}}(x_1^1), S_{\widetilde{Q}}(x_2^1), \ldots, S_{\widetilde{Q}}(x_n^1)\right)$, *in short* $\widetilde{P} \preceq_3 \widetilde{Q}$. *Special case,* $K(\widetilde{P}) \approx K(\widetilde{Q}) \Leftrightarrow \left|S_{\widetilde{P}}(x_i)\right| =$

$\left|S_{\widetilde{Q}}\left(x_i^1\right)\right|$, $i \leq n$, *in short* $\widetilde{P} \approx \widetilde{Q}$. $K(\widetilde{P}) \prec_3 K(\widetilde{Q}) \Leftrightarrow K(\widetilde{P}) \preceq_3 K(\widetilde{Q})$ *and* $K(\widetilde{P}) \neq K(\widetilde{Q})$, *in short* $\widetilde{P} \prec_3 \widetilde{Q}$.

*For the three partial order relations defined, Qian et al.* [9] *draw a conclusion that the partial order relation* $\preceq_3$ *is the best one for characterizing the coarseness/fineness between two fuzzy granular structures. Based on the relation* $\preceq_3$, *Qian et al.* [9] *proposed a definition of fuzzy information granularity.*

**Definition 2.** *Let* $K(U)$ *be the set of all fuzzy granular structures on* $U$. *For any* $K(\widetilde{P}) \in K(U)$ *there exists a real number* $g(\widetilde{P})$ *satisfying the following properties:*

1. $g(\widetilde{P}) \geq 0$ *(non-negativity)*
2. *For any* $K(\widetilde{P}), K(\widetilde{Q}) \in K(U)$, *if* $K(\widetilde{P}) \approx K(\widetilde{Q})$ *then* $g(\widetilde{P}) = g(\widetilde{Q})$ *(invariability)*
3. *For any* $K(\widetilde{P}), K(\widetilde{Q}) \in K(U)$, *if* $K(\widetilde{P}) \prec_3 K(\widetilde{Q})$ *then* $g(\widetilde{P}) < g(\widetilde{Q})$ *(monotonicity);*

*then* $g$ *is called a fuzzy information granularity.*

In order to find a fuzzy information granularity that satisfies Definition 2, we need to check the relation $\preceq_3$ between two fuzzy granular structures. However, there does not exist any relation $\preceq_3$ between two fuzzy granular structures in the following Example 1.

**Example 1.** Given $U = \{x_1, x_2, x_3, x_4\}$, $K(\widetilde{P}) = \left(S_{\widetilde{P}}(x_1), S_{\widetilde{P}}(x_2), S_{\widetilde{P}}(x_3), S_{\widetilde{P}}(x_4)\right) \in K(U)$ and $K(\widetilde{Q}) = \left(S_{\widetilde{Q}}(x_1), S_{\widetilde{Q}}(x_2), S_{\widetilde{Q}}(x_3), S_{\widetilde{Q}}(x_4)\right) \in K(U)$ are two fuzzy granular structures on $U$ where $S_{\widetilde{P}}(x_1) = 1/x_1 + 0/x_2 + 0/x_3 + 0/x_4$, $S_{\widetilde{P}}(x_2) = 0.3/x_1 + 0.6/x_2 + 0/x_3 + 0/x_4$, $S_{\widetilde{P}}(x_3) = 0/x_1 + 0/x_2 + 0.4/x_3 + 0/x_4$, $S_{\widetilde{P}}(x_4) = 0/x_1 + 0/x_2 + 0/x_3 + 0.1/x_4$ and $S_{\widetilde{Q}}(x_1) = 1/x_1 + 0.6/x_2 + 0/x_3 + 0.7/x_4$, $S_{\widetilde{Q}}(x_2) = 0.3/x_1 + 0.7/x_2 + 0.8/x_3 + 0/x_4$, $S_{\widetilde{Q}}(x_3) = 0/x_1 + 0/x_2 + 0/x_3 + 0/x_4$, $S_{\widetilde{Q}}(x_4) = 0/x_1 + 0/x_2 + 0.7/x_3 + 0.4/x_4$. It is easy to see that there does not exist any $\preceq_3$ such that $K(\widetilde{P}) \preceq_3 K(\widetilde{Q})$ or $K(\widetilde{Q}) \preceq_3 K(\widetilde{P})$. Therefore, we need to find new information measures to replace $\preceq_3$ for characterizing the coarseness/fineness between two fuzzy granular structures. In this paper, we propose a new definition of fuzzy information granularity based on fuzzy distance.

**3. Fuzzy distance and properties.** In this section, we will construct a fuzzy distance between two fuzzy granular structures. The fuzzy distance is constructed based on basic units which are the fuzzy distance between two fuzzy granules or two fuzzy sets.

***3.1. Fuzzy distance between two fuzzy sets.***

**Lemma 1.** *Given three real numbers $a$, $b$, $m$ where $a \geq b$. Then, we have $a - b \geq \min(a,\ m) - \min(b,\ m)$.*

Proof. It is easy to see that $a - b \geq \min(a,\ m) - \min(b,\ m)$ satisfies all three cases $m \geq a,\ b \leq m < a,\ m < b$. This completes the proof. □

**Lemma 2.** *Given three fuzzy sets $\widetilde{A}, \widetilde{B}, \widetilde{C}$ on the same universe $U$. Then we have the following properties:*

1. *If* $\widetilde{A} \subseteq \widetilde{B}$ *then* $|\widetilde{B}| - \left|\widetilde{B} \cap \widetilde{C}\right| \geq |\widetilde{A}| - \left|\widetilde{A} \cap \widetilde{C}\right|$.
2. *If* $\widetilde{A} \subseteq \widetilde{B}$ *then* $|\widetilde{C}| - |\widetilde{C} \cap \widetilde{A}| \geq |\widetilde{C}| - |\widetilde{C} \cap \widetilde{B}|$.
3. $|\widetilde{A}| - |\widetilde{A} \cap \widetilde{B}| + |\widetilde{C}| - |\widetilde{C} \cap \widetilde{A}| \geq |\widetilde{C}| - |\widetilde{C} \cap \widetilde{B}|$.

Proof. 1) From $\widetilde{A} \subseteq \widetilde{B}$, for any $x_i \in U$ we have $\mu_{\widetilde{B}}(x_i) \geq \mu_{\widetilde{A}}(x_i)$. By using Lemma 1 we have

$$\mu_{\widetilde{B}}(x_i) - \mu_{\widetilde{A}}(x_i) \geq \min\left(\mu_{\widetilde{B}}(x_i), \mu_{\widetilde{C}}(x_i)\right) - \min\left(\mu_{\widetilde{A}}(x_i), \mu_{\widetilde{C}}(x_i)\right).$$

$$\Leftrightarrow \sum_{i=1}^{|U|} \mu_{\widetilde{B}}(x_i) - \sum_{i=1}^{|U|} \mu_{\widetilde{A}}(x_i) \geq \sum_{i=1}^{|U|} \min\left(\mu_{\widetilde{B}}(x_i), \mu_{\widetilde{C}}(x_i)\right) - \sum_{i=1}^{|U|} \min\left(\mu_{\widetilde{A}}(x_i), \mu_{\widetilde{C}}(x_i)\right)$$

$$\rightarrow |\widetilde{B}| - |\widetilde{A}| \geq |\widetilde{B} \cap \widetilde{C}| - |\widetilde{A} \cap \widetilde{C}| \rightarrow |\widetilde{B}| - |\widetilde{B} \cap \widetilde{C}| \geq |\widetilde{A}| - |\widetilde{A} \cap \widetilde{C}|$$

2) From $\widetilde{A} \subseteq \widetilde{B}$, for any $x_i \in U$ we have $\mu_{\widetilde{B}}(x_i) \geq \mu_{\widetilde{A}}(x_i)$

$$\Leftrightarrow \min\left(\mu_{\widetilde{B}}(x_i), \mu_{\widetilde{C}}(x_i)\right) \geq \min\left(\mu_{\widetilde{A}}(x_i), \mu_{\widetilde{C}}(x_i)\right)$$

$$\Leftrightarrow \mu_{\widetilde{C}}(x_i) - \min\left(\mu_{\widetilde{A}}(x_i), \mu_{\widetilde{C}}(x_i)\right) \geq \mu_{\widetilde{C}}(x_i) - \min\left(\mu_{\widetilde{B}}(x_i), \mu_{\widetilde{C}}(x_i)\right)$$

$$\Leftrightarrow \sum_{i=1}^{|U|} \mu_{\widetilde{C}}(x_i) - \sum_{i=1}^{|U|} \min\left(\mu_{\widetilde{A}}(x_i), \mu_{\widetilde{C}}(x_i)\right) \geq \sum_{i=1}^{|U|} \mu_{\widetilde{C}}(x_i) - \sum_{i=1}^{|U|} \min\left(\mu_{\widetilde{B}}(x_i), \mu_{\widetilde{C}}(x_i)\right)$$

$$\Leftrightarrow |\widetilde{C}| - \left|\widetilde{C} \cap \widetilde{A}\right| \geq |\widetilde{C}| - \left|\widetilde{C} \cap \widetilde{B}\right|$$

3) From $\widetilde{A} \cap \widetilde{C} \subseteq \widetilde{A}$, by using property 1) we have

$$|\widetilde{A}| - |\widetilde{A} \cap \widetilde{B}| \geq |\widetilde{A} \cap \widetilde{C}| - |\widetilde{A} \cap \widetilde{C} \cap \widetilde{B}| \tag{1}$$

From $\widetilde{A} \cap \widetilde{B} \subseteq \widetilde{B}$, by using property 2) we have

$$|\widetilde{C}| - |\widetilde{C} \cap \widetilde{A} \cap \widetilde{B}| \geq |\widetilde{C}| - |\widetilde{C} \cap \widetilde{B}| \tag{2}$$

From (1) and (2) we have $|\widetilde{A}| - |\widetilde{A} \cap \widetilde{B}| + |\widetilde{C}| - |\widetilde{C} \cap \widetilde{A}| \geq |\widetilde{A} \cap \widetilde{C}| - |\widetilde{A} \cap \widetilde{C} \cap \widetilde{B}| + |\widetilde{C}| - |\widetilde{C} \cap \widetilde{A}| = |\widetilde{C}| - |\widetilde{A} \cap \widetilde{B} \cap \widetilde{C}| \geq |\widetilde{C}| - |\widetilde{C} \cap \widetilde{B}|$ □

**Proposition 1.** *Given two fuzzy sets $\widetilde{A}, \widetilde{B}$ in the same universe $U$. Then $d(\widetilde{A}, \widetilde{B}) = |\widetilde{A}| + |\widetilde{B}| - 2|\widetilde{A} \cap \widetilde{B}|$ is a fuzzy distance between $\widetilde{A}$ and $\widetilde{B}$.*

Proof.

1) It is clear that $|\widetilde{A}| \geq |\widetilde{A} \cap \widetilde{B}|$ and $|\widetilde{B}| \geq |\widetilde{A} \cap \widetilde{B}|$, so $d(\widetilde{A}, \widetilde{B}) \geq 0$.

2) It is clear that $d(\widetilde{A}, \widetilde{B}) = d(\widetilde{B}, \widetilde{A})$.

3) We need to prove the triangle inequality: $d(\widetilde{A}, \widetilde{B}) + d(\widetilde{A}, \widetilde{C}) \geq d(\widetilde{B}, \widetilde{C})$. By using Lemma 2 property 3) we have:

$|\widetilde{A}| - |\widetilde{A} \cap \widetilde{B}| + |\widetilde{C}| - |\widetilde{C} \cap \widetilde{A}| \geq |\widetilde{C}| - |\widetilde{C} \cap \widetilde{B}|$ (1)

$|\widetilde{A}| - |\widetilde{A} \cap \widetilde{C}| + |\widetilde{B}| - |\widetilde{B} \cap \widetilde{A}| \geq |\widetilde{B}| - |\widetilde{B} \cap \widetilde{C}|$ (2)

Adding (1) with (2) we have

$$(|\widetilde{A}| + |\widetilde{B}| - 2|\widetilde{A} \cap \widetilde{B}|) + (|\widetilde{A}| + |\widetilde{C}| - 2|\widetilde{A} \cap \widetilde{C}|) \geq |\widetilde{B}| + |\widetilde{C}| - 2|\widetilde{B} \cap \widetilde{C}|, \text{ or}$$

$$d(\widetilde{A}, \widetilde{B}) + d(\widetilde{A}, \widetilde{C}) \geq d(\widetilde{B}, \widetilde{C}).$$

From 1), 2), 3) we can draw a conclusion that $d(\widetilde{A}, \widetilde{B})$ is a fuzzy distance between two fuzzy sets $\widetilde{A}$ and $\widetilde{B}$. Based on this fuzzy distance, in the next subsection we will construct the fuzzy distance between two fuzzy granular structures. □

### *3.2. Fuzzy distance between two fuzzy granular structures and its properties.*

**Theorem 1.** *Let*

$$K(\widetilde{P}) = \left(S_{\widetilde{P}}(x_1), S_{\widetilde{P}}(x_2), \ldots, S_{\widetilde{P}}(x_n)\right), \ K(\widetilde{Q}) = \left(S_{\widetilde{Q}}(x_1), S_{\widetilde{Q}}(x_2), \ldots, S_{\widetilde{Q}}(x_n)\right)$$

*be two fuzzy granular structures on $K(U)$ where $U = \{x_1, x_2, \ldots, x_n\}$. Then,*

$$D\left(K(\widetilde{P}), K(\widetilde{Q})\right) = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{\left|S_{\widetilde{P}}(x_i)\right| + \left|S_{\widetilde{Q}}(x_i)\right| - 2\left|S_{\widetilde{P}}(x_i) \cap S_{\widetilde{Q}}(x_i)\right|}{n} \right)$$

*is a fuzzy distance between $K(\widetilde{P})$ and $K(\widetilde{Q})$.*

Proof. It is clear that $D(K(\widetilde{P}), K(\widetilde{Q})) \geq 0$ and $D(K(\widetilde{P}), K(\widetilde{Q})) = D(K(\widetilde{Q}), K(\widetilde{P}))$. We need to prove the triangle inequality. For any $K(\widetilde{P})$, $K(\widetilde{Q})$, $K(\widetilde{R}) \in K(U)$ we have to prove that

$$D(K(\widetilde{P}), K(\widetilde{Q})) + D(K(\widetilde{P}), K(\widetilde{R})) \geq D(K(\widetilde{Q}), K(\widetilde{R})).$$

From Proposition 1, for any $x_i \in U$ we have

$$d(S_{\widetilde{P}}(x_i), S_{\widetilde{Q}}(x_i)) + d(S_{\widetilde{P}}(x_i), S_{\widetilde{R}}(x_i)) \geq d(S_{\widetilde{Q}}(x_i), S_{\widetilde{R}}(x_i)).$$

Thus $D(K(\widetilde{P}), K(\widetilde{Q})) + D(K(\widetilde{P}), K(\widetilde{R}))$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\left|S_{\widetilde{P}}(x_i)\right| + \left|S_{\widetilde{Q}}(x_i)\right| - 2\left|S_{\widetilde{P}}(x_i) \cap S_{\widetilde{Q}}(x_i)\right|}{n}\right)$$

$$+\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\left|S_{\widetilde{P}}(x_i)\right| + \left|S_{\widetilde{R}}(x_i)\right| - 2\left|S_{\widetilde{P}}(x_i) \cap S_{\widetilde{R}}(x_i)\right|}{n}\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\frac{d\left(S_{\widetilde{P}}(x_i), S_{\widetilde{Q}}(x_i)\right)}{n} + \frac{1}{n}\sum_{i=1}^{n}\frac{d\left(S_{\widetilde{P}}(x_i), S_{\widetilde{R}}(x_i)\right)}{n}$$

$$\geq \frac{1}{n}\sum_{i=1}^{n}\frac{d\left(S_{\widetilde{Q}}(x_i), S_{\widetilde{R}}(x_i)\right)}{n} = D\left(K(\widetilde{Q}), K\left(\widetilde{R}\right)\right).$$

It is easy to see that $D\left(K(\widetilde{P}), K(\widetilde{Q})\right)$ gets the minimum value 0 if and only if $K(\widetilde{P}) = K(\widetilde{Q})$ and $D\left(K(\widetilde{P}), K(\widetilde{Q})\right)$ gets the maximum value 1 if and only if $K(\widetilde{P}) = K(\widetilde{\omega})$ and $K(\widetilde{Q}) = K(\widetilde{\delta})$ (or $K(\widetilde{P}) = K(\widetilde{\delta})$ and $K(\widetilde{Q}) = K(\widetilde{\omega})$). Therefore, $0 \leq D\left(K(\widetilde{P}), K(\widetilde{Q})\right) \leq 1$. □

**Example 2.** Given $U = \{x_1, x_2\}$, $K(\widetilde{P}) = (S_{\widetilde{P}}(x_1), S_{\widetilde{P}}(x_2))$, $K(\widetilde{Q}) = (S_{\widetilde{Q}}(x_1), S_{\widetilde{Q}}(x_2))$, $K(\widetilde{R}) = (S_{\widetilde{R}}(x_1), S_{\widetilde{R}}(x_2))$ where $S_{\widetilde{P}}(x_1) = 0.1/x_1 + 0.2/x_2$, $S_{\widetilde{P}}(x_2) = 0.2/x_1 + 0.4/x_2$, $S_{\widetilde{Q}}(x_1) = 0.2/x_1 + 0.2/x_2$, $S_{\widetilde{Q}}(x_2) = 0.1/x_1 + 0.6/x_2$, $S_{\widetilde{R}}(x_1) = 0/x_1 + 0.2/x_2$, $S_{\widetilde{R}}(x_2) = 0.1/x_1 + 0.1/x_2$. From Theorem 1 we have $D(K(\widetilde{P}), K(\widetilde{Q})) = 0.1$, $D(K(\widetilde{Q}), K(\widetilde{R})) = 0.175$, $D(K(\widetilde{P}), K(\widetilde{R})) = 0.125$. It is clear that $D(K(\widetilde{P}), K(\widetilde{Q})) + D(K(\widetilde{P}), K(\widetilde{R})) > D(K(\widetilde{Q}), K(\widetilde{R}))$.

**Proposition 2.** *Let $K(\widetilde{P}), K(\widetilde{Q}), K(\widetilde{R}) \in K(U)$ be three fuzzy granular structures on $K(U)$. If $K(\widetilde{P}) \preceq_1 K(\widetilde{Q}) \preceq_1 K(\widetilde{R})$ or $K(\widetilde{R}) \preceq_1 K(\widetilde{Q}) \preceq_1 K(\widetilde{P})$ then $D(K(\widetilde{P}), K(\widetilde{Q})) + D(K(\widetilde{Q}), K(\widetilde{R})) = D(K(\widetilde{P}), K(\widetilde{R}))$.*

P r o o f. Suppose that $K(\widetilde{P}) \preceq_1 K(\widetilde{Q}) \preceq_1 K(\widetilde{R})$. Then for any $x_i \in U$ we have $S_{\widetilde{P}}(x_i) \subseteq S_{\widetilde{Q}}(x_i) \subseteq S_{\widetilde{R}}(x_i)$ and $|S_{\widetilde{P}}(x_i) \cap S_{\widetilde{Q}}(x_i)| = |S_{\widetilde{P}}(x_i)|$, $|S_{\widetilde{Q}}(x_i) \cap S_{\widetilde{R}}(x_i)| = |S_{\widetilde{Q}}(x_i)|$, $|S_{\widetilde{P}}(x_i) \cap S_{\widetilde{R}}(x_i)| = |S_{\widetilde{P}}(x_i)|$, $|S_{\widetilde{Q}}(x_i)| \le |S_{\widetilde{R}}(x_i)|$. Therefore,

$$D(K(\widetilde{P}), K(\widetilde{Q})) + D(K(\widetilde{Q}), K(\widetilde{R}))$$

$$= \frac{1}{|U|} \sum_{i=1}^{|U|} \left( \frac{\left|S_{\widetilde{P}}(x_i)\right| + \left|S_{\widetilde{Q}}(x_i)\right| - 2\left|S_{\widetilde{P}}(x_i) \cap S_{\widetilde{Q}}(x_i)\right|}{|U|} \right)$$

$$+ \frac{1}{|U|} \sum_{i=1}^{|U|} \left( \frac{\left|S_{\widetilde{Q}}(x_i)\right| + \left|S_{\widetilde{R}}(x_i)\right| - 2\left|S_{\widetilde{Q}}(x_i) \cap S_{\widetilde{R}}(x_i)\right|}{|U|} \right)$$

$$= \frac{1}{|U|} \sum_{i=1}^{|U|} \left( \frac{\left|S_{\widetilde{Q}}(x_i)\right| - \left|S_{\widetilde{P}}(x_i)\right|}{|U|} + \frac{\left|S_{\widetilde{R}}(x_i)\right| - \left|S_{\widetilde{Q}}(x_i)\right|}{|U|} \right)$$

$$= \frac{1}{|U|} \sum_{i=1}^{|U|} \left( \frac{\left|S_{\widetilde{R}}(x_i)\right| - \left|S_{\widetilde{P}}(x_i)\right|}{|U|} \right) = D\left(K(\widetilde{P}), K\left(\widetilde{R}\right)\right).$$

□

**Example 3.** Given $U = \{x_1, x_2\}$, $K(\widetilde{P}) = (S_{\widetilde{P}}(x_1), S_{\widetilde{P}}(x_2))$, $K(\widetilde{P}) = (S_{\widetilde{P}}(x_1), S_{\widetilde{P}}(x_2))$, $K(\widetilde{R}) = (S_{\widetilde{R}}(x_1), S_{\widetilde{R}}(x_2))$ and $K(\widetilde{P}) \preceq_1 K(\widetilde{Q}) \preceq_1 K(\widetilde{R})$ where $S_{\widetilde{P}}(x_1) = 0.1/x_1 + 0.2/x_2$, $S_{\widetilde{P}}(x_2) = 0.2/x_1 + 0.3/x_2$, $S_{\widetilde{Q}}(x_1) = 0.2/x_1 + 0.3/x_2$, $S_{\widetilde{Q}}(x_2) = 0.3/x_1 + 0.4/x_2$, $S_{\widetilde{R}}(x_1) = 0.3/x_1 + 0.4/x_2$, $S_{\widetilde{R}}(x_2) = 0.4/x_1 + 0.6/x_2$. From Theorem 1 we have

$$D(K(\widetilde{P}), K(\widetilde{Q})) = \frac{0.4}{4},\ D(K(\widetilde{Q}), K(\widetilde{R})) = \frac{0.5}{4},\ D(K(\widetilde{P}), K(\widetilde{R})) = \frac{0.9}{4}.$$

It is clear that $D(K(\widetilde{P}), K(\widetilde{Q})) + D(K(\widetilde{Q}), K(\widetilde{R})) = D(K(\widetilde{P}), K(\widetilde{R}))$. From Proposition 2 we have the following Corollary 1 and Corollary 2.

**Corollary 1.** *Let $K(U)$ be a fuzzy granular structures space on $U$ and $K(\widetilde{P}), K(\widetilde{Q}) \in K(U)$. If $K(\widetilde{P}) \preceq_1 K(\widetilde{Q})$ then $D(K(\widetilde{P}), K(\widetilde{\omega})) \leq D(K(\widetilde{Q}), K(\widetilde{\omega}))$.*

**Corollary 2.** *Let $K(U)$ be a fuzzy granular structures space on $U$ and $K(\widetilde{P}), K(\widetilde{Q}) \in K(U)$. If $K(\widetilde{P}) \preceq_1 K(\widetilde{Q})$ then $D(K(\widetilde{P}), K(\widetilde{\delta})) \geq D(K(\widetilde{Q}), K(\widetilde{\delta}))$.*

**Proposition 3.** *Let $K(\widetilde{P}) \in K(U)$ be a fuzzy granular structure in $K(U)$, then we have $D(K(\widetilde{P}), K(\widetilde{\delta})) + D(K(\widetilde{P}), K(\widetilde{\omega})) = 1$.*

Proof. Suppose that $K(\widetilde{P}) = (S_{\widetilde{P}}(x_1), S_{\widetilde{P}}(x_2), \ldots, S_{\widetilde{P}}(x_n))$. Then

$$D(K(\widetilde{P}), K(\widetilde{\omega})) = \frac{1}{n^2} \sum_{i=1}^{n} |S_{\widetilde{P}}(x_i)|, \quad D(K(\widetilde{P}), K(\widetilde{\delta})) = \frac{1}{n^2} \sum_{i=1}^{n} (n - |S_{\widetilde{P}}(x_i)|).$$

Then we have $D(K(\widetilde{P}), K(\widetilde{\delta})) + D(K(\widetilde{P}), K(\widetilde{\omega})) = 1$. □

**4. Extension fuzzy information granularity based on fuzzy distance and its properties.** In this section we will present a new definition of fuzzy information granularity based on fuzzy distance, called extension fuzzy information granularity. We also present some problems about the relation between fuzzy distance and fuzzy information entropies in [3, 9].

The following Theorem 2 shows that the fuzzy information granularity in Definition 2 is also defined based on fuzzy distance.

**Theorem 2.** *Let $K(U)$ be a fuzzy granular structure space in universe $U$. For any $K(\widetilde{P}), K(\widetilde{\omega}) \in K(U)$, $D(K(\widetilde{P}), K(\widetilde{\omega}))$ is a fuzzy information granularity.*

Proof. Suppose that $K(\widetilde{P}) = (S_{\widetilde{P}}(x_1), S_{\widetilde{P}}(x_2), \ldots, S_{\widetilde{P}}(x_n))$ and $K(\widetilde{\omega}) = (S_{\widetilde{\omega}}(x_1), S_{\widetilde{\omega}}(x_2), \ldots, S_{\widetilde{\omega}}(x_n))$.

(1) It is clear that $D$ is not negative.

(2) If $K(\widetilde{P}) \approx K(\widetilde{Q})$ then there exists $f : F(\widetilde{P}) \to F(\widetilde{Q})$ such that for any $x_i \in U$, $|S_{\widetilde{P}}(x_i)| = |f(S_{\widetilde{P}}(x_i))|$ and $f(S_{\widetilde{P}}(x_i)) = S_{\widetilde{Q}}(x_{j_i})$. Then we have

$$D(K(\widetilde{P}), K(\widetilde{\omega}))$$

$$= \frac{1}{|U|} \sum_{i=1}^{|U|} \left( \frac{\left|S_{\widetilde{P}}(x_i)\right| + |S_{\widetilde{\omega}}(x_i)| - 2\left|S_{\widetilde{P}}(x_i) \cap S_{\widetilde{\omega}}(x_i)\right|}{|U|} \right) = \frac{1}{|U|} \sum_{i=1}^{|U|} \left( \frac{\left|S_{\widetilde{P}}(x_i)\right|}{|U|} \right)$$

$$= \frac{1}{|U|}\sum_{i=1}^{|U|}\left(\frac{\left|f\left(S_{\widetilde{P}}(x_i)\right)\right|}{|U|}\right) = \frac{1}{|U|}\sum_{i=1}^{|U|}\left(\frac{\left|S_{\widetilde{Q}}\left(x_{j_i}\right)\right|}{|U|}\right) = \frac{1}{|U|}\sum_{i=1}^{|U|}\left(\frac{\left|S_{\widetilde{Q}}(x_j)\right|}{|U|}\right)$$

$$= D\left(K(\widetilde{Q}), K\left(\widetilde{\omega}\right)\right).$$

(3) We will prove that if $K(\widetilde{P}) \prec_3 K(\widetilde{Q})$ then $D(K(\widetilde{P}), K(\widetilde{\omega})) < D(K(\widetilde{Q}), K(\widetilde{\omega}))$. Suppose that $K(\widetilde{P}) = (S_{\widetilde{P}}(x_1), S_{\widetilde{P}}(x_2), \dots, S_{\widetilde{P}}(x_{|U|}))$, $K(\widetilde{Q}) = (S_{\widetilde{Q}}(x_1), S_{\widetilde{Q}}(x_2), \dots, S_{\widetilde{Q}}(x_{|U|}))$ and $K(\widetilde{P}) \prec_3 K(\widetilde{Q})$, then there exists a fuzzy granular structure $K^1(\widetilde{Q})$ of $K(\widetilde{Q})$ where $K^1(\widetilde{Q}) = (S_{\widetilde{Q}}(x_1^1), S_{\widetilde{Q}}(x_2^1), \dots, S_{\widetilde{Q}}(x_{|U|}^1))$ such that $|S_{\widetilde{P}}(x_i)| \leq S_{\widetilde{Q}}(x_i^1)$ and there exist at least one $x_s \in U$ such that $|S_{\widetilde{P}}(x_s)| < |f(S_{\widetilde{P}}(x_s))| = |S_{\widetilde{Q}}(x_s^1)|$. Therefore

$$D\left(K(\widetilde{P}), K\left(\widetilde{\omega}\right)\right)$$

$$= \frac{1}{|U|}\sum_{i=1}^{|U|}\left(\frac{\left|S_{\widetilde{P}}(x_i)\right| + |S_{\widetilde{\omega}}(x_i)| - 2\left|S_{\widetilde{P}}(x_i) \cap S_{\widetilde{\omega}}(x_i)\right|}{|U|}\right) = \frac{1}{|U|}\sum_{i=1}^{|U|}\left(\frac{\left|S_{\widetilde{P}}(x_i)\right|}{|U|}\right)$$

$$= \frac{1}{|U|}\left(\sum_{i=1, i\neq s}^{|U|}\frac{\left|S_{\widetilde{P}}(x_i)\right|}{|U|} + \frac{\left|S_{\widetilde{P}}\left(x_s\right)\right|}{|U|}\right) < \frac{1}{|U|}\left(\sum_{i=1, i\neq s}^{|U|}\frac{\left|S_{\widetilde{Q}}\left(x_i^1\right)\right|}{|U|} + \frac{\left|S_{\widetilde{Q}}\left(x_s^1\right)\right|}{|U|}\right)$$

$$= \frac{1}{|U|}\left(\sum_{i=1}^{|U|}\frac{\left|S_{\widetilde{Q}}\left(x_i^1\right)\right|}{|U|}\right) = D\left(K(\widetilde{Q}), K\left(\widetilde{\omega}\right)\right).$$

Consequently, $D\left(K(\widetilde{P}), K\left(\widetilde{\omega}\right)\right)$ is a fuzzy information granularity on $K(U)$.

For $K(\widetilde{P}), K(\widetilde{Q}) \in K(U)$, Definition 2 shows that to construct a fuzzy information granularity we need to check $K(\widetilde{P}) \prec_3 K(\widetilde{Q})$ or $K(\widetilde{Q}) \prec_3 K(\widetilde{P})$. Example 1 shows that in some cases it cannot be determined whether $K(\widetilde{P}) \prec_3 K(\widetilde{Q})$, $K(\widetilde{Q}) \prec_3 K(\widetilde{P})$ or not. Therefore, the relation $\prec_3$ could not be used to compare the fineness or coarseness between two fuzzy granular structures. Theorem 2 shows that fuzzy distance $D\left(K(\widetilde{P}), K(\widetilde{\omega})\right)$ can be used to replace $\prec_3$ to definition new fuzzy information granularity. $\square$

**Example 4** (Continue from Example 1)**.** To compare the fineness/coarseness between $K(\widetilde{P}), K(\widetilde{Q}) \in K(U)$, we calculate the fuzzy space $D(K(\widetilde{P}), K(\widetilde{\omega})) = 0.15$, $D(K(\widetilde{Q}), K(\widetilde{\omega})) = 0.325$. Therefore, $D(K(\widetilde{P}), K(\widetilde{\omega})) < D(K(\widetilde{Q}), K(\widetilde{\omega}))$ and we can conclude that $K(\widetilde{Q})$ is rougher than $K(\widetilde{P})$.

Based on the fuzzy distance $D(K(\widetilde{P}), K(\widetilde{\omega}))$, extension fuzzy information granularity is defined as

**Definition 3.** *Let $K(U)$ be a fuzzy granular structure space on $U$. For any $K(\widetilde{P}) \in K(U)$ there exists a real number $g(\widetilde{P})$ satisfying the following properties:*

1. $g(\widetilde{P}) \geq 0$ *(non-negativity)*
2. $\forall K(\widetilde{P}), K(\widetilde{Q}) \in K(U)$, *if* $D(K(\widetilde{P}), K(\widetilde{\omega})) = D(K(\widetilde{Q}), K(\widetilde{\omega}))$ *then* $g(\widetilde{P}) = g(\widetilde{Q})$ *(invariability)*
3. $\forall K(\widetilde{P}), K(\widetilde{Q}) \in K(U)$, *if* $D(K(\widetilde{P}), K(\widetilde{\omega})) < D(K(\widetilde{Q}), K(\widetilde{\omega}))$ *then* $g(\widetilde{P}) < g(\widetilde{Q})$ *(monotonicity)*

*then $g$ is called an extension fuzzy information granularity.*

Next, we present the relation between fuzzy distance $D(K(\widetilde{P}), K(\widetilde{\omega}))$ and fuzzy information entropies [3, 9].

Let $K(\widetilde{P}) = (S_{\widetilde{P}}(x_1), S_{\widetilde{P}}(x_2), \ldots, S_{\widetilde{P}}(x_n))$ be a fuzzy granular structure in universe $U$. Based on Shannon's entropy and Liang's entropy, fuzzy entropies $H(\widetilde{P})$ and $E(\widetilde{P})$ are defined as

$$H(\widetilde{P}) = -\frac{1}{n}\sum_{i=1}^{n}\log_2\frac{|S_{\widetilde{P}}(x_i)|}{n}, \quad E(\widetilde{P}) = \sum_{i=1}^{n}\frac{1}{n}\left(1 - \frac{|S_{\widetilde{P}}(x_i)|}{n}\right)$$

**Proposition 4.** *Let $K(U)$ be a fuzzy granular structure space and $K(\widetilde{P})$, $K(\widetilde{Q}) \in K(U)$. Then we have*

1) *If* $D(K(\widetilde{P}), K(\widetilde{\omega})) \geq D(K(\widetilde{Q}), K(\widetilde{\omega}))$ *then* $H(\widetilde{P}) \leq H(\widetilde{Q})$

2) *If* $D(K(\widetilde{P}), K(\widetilde{\omega})) \geq D(K(\widetilde{Q}), K(\widetilde{\omega}))$ *then* $E(\widetilde{P}) \leq E(\widetilde{Q})$.

Proof. If $D(K(\widetilde{P}), K(\widetilde{\omega})) \geq D(K(\widetilde{Q}), K(\widetilde{\omega}))$ then according to Proposition 1 we have $|S_{\widetilde{P}}(x_i)| \geq |S_{\widetilde{Q}}(x_i)|$ for any $x_i \in U$. From the formula of $H(\widetilde{P})$ and $E(\widetilde{P})$ we have $H(\widetilde{P}) \leq H(\widetilde{Q})$ and $E(\widetilde{P}) \leq E(\widetilde{Q})$.

Proposition 4 shows that the coarser the $K(\widetilde{P})$ is (the bigger the fuzzy distance $D(K(\widetilde{P}), K(\widetilde{\omega}))$ is), the smaller the entropies $H(\widetilde{P})$, $E(\widetilde{P})$ are and vice versa. This property is the same as the property of information entropies in [3]. Consequently, extension fuzzy information granularity in Definition 3 is strongly related to fuzzy information entropies and it is used to estimate the fineness/coarseness or the difference, distinguishability of fuzzy granular structures. □

**5. Experiment.** Let us consider an information system $IS = (U, A)$, suppose that $K(\widetilde{P})$ and $K(\widetilde{Q})$ are two fuzzy granular structures defined by fuzzy similarity relations $\widetilde{P}$ and $\widetilde{Q}$ on $P, Q \subseteq A$. The purpose of this experiment is to determine the ability to distinguish $K(\widetilde{P})$, $K(\widetilde{Q})$ based on partial order relations $\preceq_1$, $\preceq_2$, $\preceq_3$ (Definition 1) and fuzzy distance $D(K(\widetilde{P}), K(\widetilde{\omega}))$. For example, for the relation $\preceq_1$, if $K(\widetilde{P}) \prec_1 K(\widetilde{Q})$ or $K(\widetilde{Q}) \prec_1 K(\widetilde{P})$, $K(\widetilde{P})$ can be distinguised from $K(\widetilde{Q})$ by the relation $\preceq_1$. The relations $\preceq_1$, $\preceq_2$, $\preceq_3$ and $D(K(\widetilde{P}), K(\widetilde{\omega}))$ are denoted as QH1, QH2, QH3 and QHM respectively.

The experiment performed on six data sets which have numeric attributes from UCI [10] as in the following Table 1:

Table 1. Data sets for experiment

| No | Data sets | Objects | Condition attributes | Granular structures | Pairs of granular structures |
|---|---|---|---|---|---|
| 1 | *Ecoli* | 336 | 7 | 127 | 8001 |
| 2 | *Seeds* | 210 | 7 | 127 | 8001 |
| 3 | *Brest Cancer* | 699 | 9 | 511 | 130305 |
| 4 | *Lenses* | 24 | 4 | 15 | 105 |
| 5 | *Balloons* | 20 | 4 | 15 | 105 |
| 6 | *Hayes − Roth* | 132 | 4 | 15 | 105 |

Each data set is considered as an information system $IS = (U, A)$. We use the fuzzy similarity relation $\widetilde{p}$ where $p \in A$ in which the value of the relation between $x_i, x_j \in U$ is defined as

$$p_{ij} = \begin{cases} 1 - 4 * \dfrac{|p(x_i) - p(x_j)|}{|p_{\max} - p_{\min}|}, & \dfrac{|p(x_i) - p(x_j)|}{|p_{\max} - p_{\min}|} \leq 0.25 \\ 0, \quad \text{otherwise} \end{cases}$$

where $p(x_i)$ is the value of the attribute $p$ at the object $x_i$, the maximum and minimum value of the attribute $p$ are $p_{\max}$, $p_{\min}$ respectively. It is clear that $p_{ii} = 1$ and $0 \leq p_{ij} \leq 1$. Each attribute set $P \subseteq A$ determines a granular structure $K(\widetilde{P})$. The total number of granular structures of each data set is $2^{|A|} - 1$ where $|A|$ is the number of condition attributes (column 5 of Table 1). The total number of granular structure pairs which are compared to each other is $C_n^2 = \dfrac{n!}{2!\,(n-2)!}$ where $n$ is the total number of granular structures (column 6 of Table 1).

To evaluate the distinguishability of two fuzzy granular structures, we use an identification rate:

*Identification rate (IR) = (Total number of distinguishability granular structure pairs) / (total number of granular structures)*

The experiments run on a PC Pentium dual core 2.13 GHz CPU, 1GB of RAM, Windows 7. The values of identification rate (IR) are displayed in Table 2.

Table 2. Values of IR for QH1, QH2, QH3, QHM

| **No** | **Data sets** | **QH1** | **QH2** | **QH3** | **QHM** |
|---|---|---|---|---|---|
| 1 | *Ecoli* | 0.0125 | 0.1257 | 0.2506 | 1 |
| 2 | *Seeds* | 0.0114 | 0.3756 | 0.7502 | 1 |
| 3 | *Brest Cancer* | 0.0084 | 0.1127 | 0.2058 | 1 |
| 4 | *Lenses* | 0.0185 | 0.8096 | 0.8113 | 0.8596 |
| 5 | *Balloons* | 0.0167 | 0.7558 | 0.7652 | 0.9045 |
| 6 | *Hayes − Roth* | 0.0174 | 0.4312 | 0.6994 | 0.9256 |

Values of IR for QH1, QH2, QH3, QHM on each data set are displayed on Fig. 1:

From the results of experiments as Table 2 and Fig. 1, QHM has the largest IR, following by QH3, QH2 and QH1. Moreover, QHM distinguishes all fuzzy granular structure pairs on the data set Ecoli, Seeds and Brest Cancer. For other data sets, QHM is smaller than 1 because it does not count the fuzzy granular structure pairs which have the same fuzzy distance $D(K(\widetilde{P}), K(\widetilde{\omega}))$. Consequently, the fuzzy distance $D(K(\widetilde{P}), K(\widetilde{\omega}))$ is the best measure to distinguish fuzzy granular structures. The proposed extension fuzzy information granularity defined by fuzzy distance is the best one to characterize the fineness/coarseness of a fuzzy granular structure. Determining the fineness/coarseness of a fuzzy

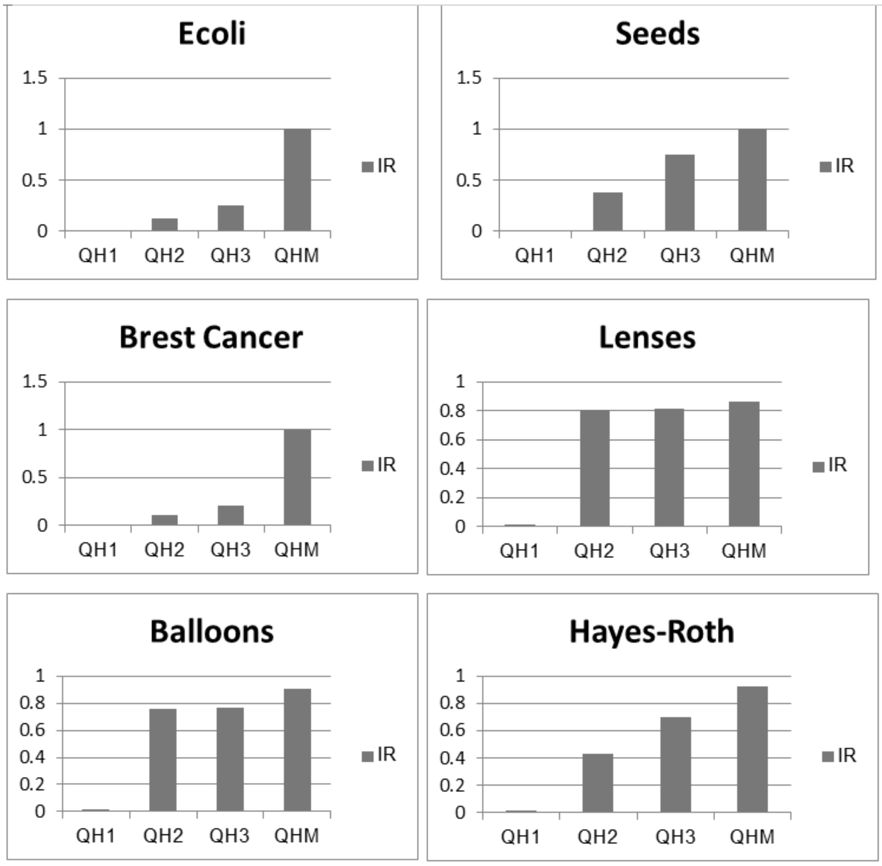


Fig. 1. Value of IR on six data sets

granular structure is the basic problem to solve the approximating, data mining, especially in attribute reduction and rule extraction based on the fuzzy rough set approach.

**6. Conclusions.** The information granularity is one of the basis measures in granular computing, and has attracted the attention of researchers for many years. According to the fuzzy similarity relation approach, in this paper we proposed an extension fuzzy information granularity based on the fuzzy distance between two fuzzy granular structures. The extension fuzzy information granularity overcomes the disadvantages of fuzzy information granularity based on three partial order relations referred in [9]. The result of experiments on some data sets shows that, the extension fuzzy information granularity is the best one to distinguish fuzzy granular structures and characterize the fineness/coarseness of a fuzzy granular structure. Our future direction is to research methodologies of attribute reduction directly on decision tables with numeric attribute domain by using this fuzzy distance.

## REFERENCES

[1] DUBOIS D., H. PRADE. Putting Rough Sets and Fuzzy Sets Together. In: Intelligent Decision Support (Ed. by R. Slowinski). Part 2. *Theory and Decision Library*, **11** (1992), 203–232.

[2] DUBOIS D., H. PRADE. Rough Fuzzy Sets and Fuzzy Rough Sets. *International Journal of General Systems*, **17** (1990), No 2–3, 191–209.

[3] HU Q. H., D. R. YU, Z. X. XIE, J. F. LIU. Fuzzy probabilistic approximation spaces and their information measures. *IEEE Trans. Fuzzy Syst.*, **14** (2006), No 2, 191–201.

[4] KRYSZKIEWICZ M. Rough set approach to incomplete information systems. *Information Science*, **112** (1998), 39–49.

[5] LIANG J. Y., Z. Z. SHI, D. Y. LI, M. J. WIERMAN. The information entropy, rough entropy and knowledge granulation in incomplete information system. *Int. J. Gen. Syst.*, **35** (2006), No 6, 641–654.

[6] LIANG J. Y., Z. Z. SHI. The information entropy, rough entropy and knowledge granulation in rough set theory. *Int. J. Unc. Fuzz. Knowl. Based Syst.*, **12** (2004), No 1, 37–46.

[7] LIANG J. Y., Y. H. QIAN. Information granules and entropy theory in information systems. *Sci. China Ser. F-Inf. Sci.*, **51** (2008), No 10, 1427–1444.

[8] PAWLAK Z. Rough sets: Theoretical Aspects of Reasoning About Data. Kluwer Academic Publishers, 1991.

[9] QIAN Y. H., J. Y. LIANG, W. Z. WU, C. Y. DANG. Information Granularity in Fuzzy Binary GrC Model. *IEEE Trans. Fuzzy Syst.*, **19** (2011), No 2, 253–264.

[10] The UCI Machine Learning Repository. `http://archive.ics.uci.edu/ml/datasets.html`, 14 December 2016.

[11] Wierman M. J. Measuring uncertainty in rough set theory. *Int. J. Gen. Syst.*, **28** (1999), No 4–5, 283–297.

[12] Xu B. W., Y. M. Zhou, H. M. Lu. An improved accuracy measure for rough sets. *J. Comput. Syst. Sci.*, **71** (2005), No 2, 163–173.

[13] Zadeh L. Fuzzy logic equals computing with words. *IEEE Trans. Fuzzy Syst.*, **4** (1996), No 2, 103–111.

*Nguyen Van Thien*
*Hanoi University of Industry, Vietnam*
*e-mail:* `nvthien1970@gmail.com`

*Demetrovics Janos*
*Institute for Computer and Control*
*(SZTAKI)*
*Hungarian Academy of Sciences*
*e-mail:* `demetrovics@sztaki.mta.hu`

*Vu Duc Thi*
*e-mail:* `vdthi@vnu.edu.vn`
*Nguyen Long Giang*
*e-mail:* `nlgiang@ioit.ac.vn`
*Nguyen Nhu Son*
*e-mail:* `nnson@ioit.ac.vn`
*Institute of Information Technology*
*VNU, Vietnam*

*Received April 11, 2016*
*Final Accepted July 4, 2016*