SOME INDEXES FOR COMPARING AND SELECTING PARTITIONS

Xuezheng Zhang, Zheng Pei, Yong Liu

School of Mathematics & Computer Engineering, Xihua University,
Chengdu, Sichuan, China
E-mail:pei.zheng28@gmail.com, pqyz@263.net

Received: 21-01-2010
Accepted: 10-02-2011

Abstract

Based on the equivalence relation, we can partition a set U, formally, different equivalence relations correspond to different partitions of U. In this paper, based on Yager's works (Some Measure Relating Partitions Useful For Computational Intelligence, International Journal of Computational Intelligence Systems. 1, 1-18 (2008)), we discuss how to compare the different partitions, moreover, we obtain some indexes to select a better partition from given partitions. By an illustrative example, we show that our proposed indexes can be used for selecting partitions, feature selection, and help us to gather more information to decision making.

Keywords: Computational Intelligence, Partitions, Data mining, Information Granulation, Measurement

1. Introduction

Our knowledge in daily life is inherently associated with a way in which they are perceived, described, and classified. More general, human beings use perceptions of direction, speed, time and other features of physical/mental objects to process information, *e.g.*, driving and cooking. Perceptions are granular (information granular), which are collections of objects arranged together based on their similarity, functional adjacency and indistinguishability [23,24], information granulation exhibits different facets of formalism and as such rely on the well established theories of interval and interval calculus, fuzzy sets, rough sets and alike [6,7,8,9,11]. From the mathematical point of view, the fundamental of information granulation is relations on the set of objects, objects can be easily managed by these relations on the set of objects [15,16,17,19]. However in Data mining, we are always faced with a great of data [5,10,13], in such case, it needs sophisticated tools of computational intelligence to manage objects by classifying a set [3,6,7,9,12,13,14,20,22,25]. On the one hand, we use features for classifying objects and obtaining partitions of the set of objects. On the other hand, features in complex systems are so different in describing local knowledge that we have to analyze importance of every feature and select important features for constructing knowledge-based systems [1,2,18].

In the reference [21], Yager provides an alternative method for dealing with the above mentioned problem. The method handles the relation between partitions and the congruent about two partitions of the same set, distinguishes partitions by measurements related to partitions. In this paper, our discussions concentrate on "How to select the best partition from those different partitions of the same set". In the real world practice, the problem is associated with knowledge extraction, feature selection, and decision making. Based on Yager's works, we compare with different partitions as well as select a better par-

Published by Atlantis Press
Copyright: the authors

tition from given partitions. The main results are:

1. Providing a new method to calculate the degree of congruence or similarity of two partitions and indexes for selecting partitions;

2. Providing Stability Entropy, accuracy rate and validity entropy to evaluate a partition.

The paper is organized as follows: In Section 2, we review the congruence and partitions. In Section 3, based on analyzing the relation of elements in two partitions of the same set U, we propose a new method to calculate the degree of congruence or similarity of two partitions. In Section 4, we discuss indexes for selecting partitions, and provide Stability Entropy, accuracy rate and validity entropy to evaluate a partition. In Section 5, we give an illustrative example to explain our method. We conclude in Section 6.

2. The Congruence of Partitions

According to the paper [21], we review some concepts of partitions as follow:

An equivalence relation E on U is a mapping $E: U \times U \longrightarrow \{0,1\}$ such that

1. Identity: $E(x,x) = 1$;

2. Symmetry: $E(x,y) = E(y,x)$;

3. Transitivity:$E(x,z) \geqslant Min_y[E(x,y),E(y,z)]$

Condition 3 implies that if $E(x,y) = 1$ and $E(y,z) = 1$, then $E(x,z) = 1$. Transitivity means that if $\{A_1,A_2,\cdots,A_n\}$ is a partition of U, then we can obtain an equivalence relation E such that $E(x,y) = 1$ if x and y in the same class and $E(x,y) = 0$ if they are in different classes. Hence, we can associate with each $x \in U$ an equivalence class A_x such that $y \in A_x$ if $E(x,y) = 1$.

Assume that P_1 and P_2 are two partitions of U, the mapping $Cong: P_1 \times P_2 \rightarrow [0,1]$ indicates the degree of congruence or similarity of two partitions if the following conditions are satisfied:

$C_1: \ Cong(P_1,P_2) = Cong(P_2,P_1)$;

$C_2: \ Cong(P_1,P_1) = 1$ if $P_1 = P_2$.

Formally, $Cong$ is similarity relation on partitions of U.

In all partitions of U, we identify the two special partitions P^* and P_*, in which, $P^* = \langle\{U\}\rangle$, *i.e.*, the universal partition where just have one set; P_* is the one in which each element in U is in a different class, *i.e.*, if $P_* = \langle A_1,A_2,\cdots,A_n\rangle$, then for every $i = 1,2,\cdots,n$, $A_i = \{x_i\}(x_i \in U)$. Formally, the above $Cong$ also satisfies the following condition:

$C_3: \ Cong(P^*,P_*) = Min\{Cong(P_1,P_2)\}$,

in which, P_1 and P_2 are two partitions of U. In the real-world practice, the following two $Cong$ are used [21]:

$$\begin{aligned} & Cong_1(P_1,P_2) = Cong_1(E_1,E_2) \\ = \ & 1 - \frac{1}{\begin{pmatrix} 2 \\ n \end{pmatrix}} \sum_U |U \cdot E_1(\langle x,y\rangle) - U \cdot E_2(\langle x,y\rangle)|, \end{aligned}$$

In $Cong_1$, $U \cdot E(\langle x,y\rangle) = E(x,y) = E(y,x)$ (E is an equivalence relation on U), $\sum_U |U \cdot E_1(\langle x,y\rangle) - U \cdot E_2(\langle x,y\rangle)|$ is the number of pairs that have different values in E_1 (the equivalence relation corresponding to P_1) and E_2 (the equivalence relation corresponding to P_2), $n = |U|$ and

$$\begin{pmatrix} 2 \\ n \end{pmatrix} = \frac{n(n-1)}{2}.$$

$$Cong_2(P_1,P_2) = Max[Score(g(P_1,P_2))],$$

in which, $P_1 = \{A_1,A_2,\cdots,A_q\}$, $P_2 = \{B_1,B_2,\cdots,B_q\}$, let $Q = \{1,2,\cdots,q\}$ and $g: Q \rightarrow Q$ be bijective, then

$$\begin{aligned} Score(g(P_1,P_2)) &= \frac{\sum_{j=1}^{q} Card(D_{g \cdot j})}{Card(U)}, \\ D_{g \cdot j} &= A_j \bigcap B_{g(j)}. \end{aligned}$$

In the reference [21], many interesting properties of the above measures of congruence of partitions are discussed. In the follows, we focus on some indexes for comparing partitions, which will help us to select a better partition from all partitions of U.

3. Indexes for Evaluating Partitions

In this section, we provide a new *Cong* of two partitions and some indexes for evaluating partitions based on Yager's works [21].

3.1. A New *Cong* of Partitions

Assume that $E_1, E_2, \cdots, E_n$ are equivalence relations of U, partitions $P_1, P_2, \cdots, P_n$ are induced by $E_1, E_2, \cdots, E_n$, respectively. Let the unordered pair $\langle x, y\rangle$ such that $x, y \in U$ and $x \neq y$.

Definition 1. $\langle x, y\rangle$ is called a double basic factor (DBF) if $E_1(x,y) = E_2(x,y) = 1$, *i.e.*, for the partitions P_1 and P_2, the double basic factor $\langle x, y\rangle$ means that x and y are in the same class.

Definition 2. $\langle x, y\rangle$ is called an $n-$multiple basic factor (nBF) if $E_1(x,y) = E_2(x,y) = \cdots = E_n(x,y) = 1$, *i.e.*, for the partitions $P_1, P_2, \cdots, P_n$, the $n-$multiple basic factor $\langle x, y\rangle$ implies that x and y are in the same class.

Definition 3. $\langle x, y\rangle$ is called a double independent factor (DIF) if $E_1(x,y) = E_2(x,y) = 0$. *i.e.*, for the partitions P_1 and P_2, the double independent factor $\langle x, y\rangle$ means that x and y are not in the same class.

Definition 4. $\langle x, y\rangle$ is called an $n-$multiple independent factor (nIF) if $E_1(x,y) = E_2(x,y) = \cdots = E_n(x,y) = 0$, *i.e.*, for the partitions $P_1, P_2, \cdots, P_n$, the $n-$multiple independent factor $\langle x, y\rangle$ implies that x and y are not in the same class.

Definition 5. $x_0 \in U$ is called an absolute independent element (AIE) of E if for any $x \in U$ and $x \neq x_0$, then $E(x, x_0) = 0$, *i.e.*, if x_0 is the absolute independent element of E, then $A_{x_0} = \{x_0\}$.

Definition 6. $x_0 \in U$ is called an $n-$multiple absolute independent element $(nAIE)$ of $E_1, E_2, \cdots, E_n$ if for any $x \in U$ and $x \neq x_0$, then $E_1(x, x_0) = E_2(x, x_0) = \cdots = E_n(x, x_0) = 0$, *i.e.*, if x_0 is the $n-$multiple absolute independent element of $E_1, E_2, \cdots, E_n$, then for every $E_j (j = 1, 2, \cdots, n)$, $A_{x_0} = \{x_0\}$.

Example 1. Let $U = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$. Assume equivalence relations E_1 and E_2, in which,

$$\begin{aligned} P_1 &= \{A_1, A_2, A_3\} \\ &= \{\{x_1, x_2, x_3\}, \{x_4, x_5, x_6\}, \{x_7\}\}, \\ P_2 &= \{B_1, B_2\} \\ &= \{\{x_2, x_4, x_6\}, \{x_1, x_3, x_5, x_7\}\}. \end{aligned}$$

In this case, according to Definitions 1-6, we know that $\langle x_1, x_3\rangle$ and $\langle x_4, x_6\rangle$ are double basic factors. $\langle x_1, x_4\rangle$, $\langle x_1, x_6\rangle$, $\langle x_2, x_5\rangle$, $\langle x_2, x_7\rangle$, $\langle x_3, x_4\rangle$, and $\langle x_3, x_6\rangle$ are double independent factors. And x_7 is the absolute independent factor of P_1.

Assume $P_1 = \{A_1, \cdots, A_n\}$, $P_2 = \{B_1, \cdots, B_n\}$ and $M_{ij} = A_i \bigcap B_j$. It is not difficult to obtain the properties as follows.

Theorem 1. *If $|M_{ij}| \geqslant 2$, then $\forall x, y \in M_{ij}$, $\langle x, y\rangle$ is a DBF.*

Proof. For any $x, y \in M_{ij}$, $x, y \in A_i$ and $x, y \in B_j$, *i.e.*, $E_1(x,y) = E_2(x,y) = 1$, according to Definition 1, $\langle x, y\rangle$ is a DBF. □

Corollary 2. *For any $x \in U$,*

1. *x is an AIE of P_*;*
2. *x is not an AIE of P^*.*

Theorem 3. *For P_* and P^*, $\forall x, y \in U$ and $x \neq y$, $\langle x, y\rangle$ is not a DBF of P_* and P^*.*

Proof. $\forall x, y \in U$, $x \neq y$, we have $E^*(x,y) = 1$ in P^* and $E_*(x,y) = 0$ in P_* , hence, $\langle x, y\rangle$ is not a DBF of P_* and P^*. □

Theorem 4. *For any partition P of U such that $P \neq P_*$, there exists a DBF $\langle x, y\rangle$ of P and P^*.*

Proof. Due to $P \neq P_*$, hence there are at least two elements x and y are in the same class in P. That is $E(x,y) = 1$. And $\forall x, y \in U$, $E^*(x,y) = 1$. This means that $\langle x, y\rangle$ is a DBF of P and P^*. □

Theorem 5. *There exists a DBF of P_1 and P_2 if and only if there exist $A_i \in P_1$ and $B_j \in P_2$ such that $|M_{ij}| = |A_i \bigcap B_j| \geqslant 2$.*

Proof. Assume that there exists a DBF of P_1 and P_2, denoted by $\langle x,y\rangle$. It means that x and y are in the same class in both P_1 and P_2, *i.e.*, there exists $A_i \in P_1$ and $B_j \in P_2$ such that $x,y \in A_i$ and $x,y \in B_j$, hence $x,y \in M_{ij} = A_i \bigcap B_j$, *i.e.*, $|M_{ij}| \geqslant 2$.

Assume that for $A_i \in P_1$ and $B_j \in P_2$ such that $|M_{ij}| = |A_i \bigcap B_j| \geqslant 2$. Then there exist $x,y \in M_{ij}$ and $x \neq y$, *i.e.*, $x,y \in A_i$ and $x,y \in B_j$, hence, $\langle x,y\rangle$ is a DBF of P_1 and P_2. □

Formally, we can calculate the degree of similarity of partitions P_1 and P_2 of U based on all DBF and AIE as follow:

$$Cong_3(P_1,P_2) = Cong_3(E_1,E_2) = \frac{|U^*|+|U_*|}{|U|}, \quad (1)$$

in which, U^* is the set of all double basic factors, *i.e.*,

$$U^* = \{x,y|x,y \in U, x \neq y, E_1(x,y) = E_2(x,y) = 1\},$$

U_* is the set of all 2-multiple absolute independent elements, *i.e.*,

$$U^* = \{x|\forall y \in U, x \neq y, E_1(x,y) = E_2(x,y) = 0\}.$$

Theorem 6. *For any partitions P_1 and P_2 of U,*

$$Cong_3(P_*,P^*) = Min[Cong_3(P_1,P_2)].$$

Proof. According to Eq. (1), we have

$$Cong_3(P_1,P_2) = \frac{|U^*|+|U_*|}{|U|}.$$

It easy to prove that for P_* and P^* of U, we have $|U^*| = 0$ and $|U_*| = 0$, *i.e.*,

$$Cong_3(P_*,P^*) = 0.$$

On the other hand, for any partitions P_1 and P_2 of U, $Min[Cong_3(P_1,P_2)] \geqslant 0$, hence,

$$Cong_3(P_*,P^*) = Min[Cong_3(P_1,P_2)].$$

□

Corollary 7. *Assume that P_1 and P_2 are two partitions of U. $Cong_3 : P_1 \times P_2 \to [0,1]$ is the degree of congruence or similarity of two partitions.*

Proof. According to the Eq. (1), $Cong_3$ satisfies that

1: $Cong_3(P_1,P_2) = Cong_3(P_2,P_1) = \frac{|U^*|+|U_*|}{|U|}$,

2: $Cong_3(P_1,P_1) = \frac{|U|}{|U|} = 1$

Based on Theorem 6, $Cong_3$ is the degree of congruence or similarity of two partitions [21]. □

$Cong_3$ can be generalized to $n-$multiple case as follows:

$$Cong_3(P_1,P_2,\cdots,P_n) = \frac{|U^{*n}|+|U_{*n}|}{|U|}. \quad (2)$$

in which, U^{*n} is the set of all $n-$multiple basic factors, *i.e.*,

$$U^{*n} = \{x,y|x,y \in U, x \neq y, E_1(x,y) = \cdots E_n(x,y) = 1\},$$

U_{*n} is the set of all $n-$multiple absolute independent, *i.e.*,

$$U_{*n} = \{x|\forall y \in U, x \neq y, E_1(x,y) = \cdots = E_n(x,y) = 0\}.$$

Compared $Cong_1$ and $Cong_2$ with $Cong_3$, $Cong_3$ is easier to calculate than $Cong_1$ and $Cong_2$. This can be shown in the following example.

Example 2. Let $U = \{x_1,x_2,x_3x_4,x_5,x_6,x_7\}$. Assume $P_1 = \{A_1\ (= \{x_1,x_2,x_3\}), A_2(= \{x_4,x_5,x_6\}), A_3(= \{x_7\})\}$, $P_2 = \{B_1 = \{x_2,x_4,x_6\},\ B_2 = \{x_1, x_3,x_5,x_7\}\}$.

The $Cong_1$ can be calculated by the follows:

$$\begin{aligned} & Cong_1(P_1,P_2) = Cong_1(E_1,E_2) \\ = & \ 1 - \frac{1}{\begin{pmatrix} 2 \\ n \end{pmatrix}} \sum_U |U \cdot E_1(\langle x,y\rangle) - U \cdot E_2(\langle x,y\rangle)| \\ = & \ 1 - \frac{11}{21} = \frac{10}{21} \end{aligned}$$

To calculate $Cong_2$, we add $B_3 = \emptyset$ in P_2. The number of bijective $g : \{1,2,3\} \to \{1,2,3\}$ is six, *i.e.*,

1. For $g(1) = 2$, $g(2) = 1$ and $g(3) = 3$, we have the pairs: $(A_1,B_2),(A_2,B_1)$ and (A_3,B_3). Then,

$$\begin{aligned} D_{g.1} &= A_1\bigcap B_2 = \{x_1,x_3\}, \\ D_{g.2} &= A_2\bigcap B_1 = \{x_4,x_6\}, \\ D_{g.3} &= A_3\bigcap B_3 = \emptyset. \end{aligned}$$

Hence, $Score(g(P_1,P_2) = \frac{4}{7}$.

2. For $g(1) = 1$, $g(2) = 2$ and $g(3) = 3$, similarly, we have $Score(g(P_1,P_2) = \frac{2}{7}$.
3. For $g(1) = 1$, $g(2) = 3$ and $g(3) = 2$, we have $Score(g(P_1,P_2) = \frac{2}{7}$.
4. For $g(1) = 2$, $g(2) = 3$ and $g(3) = 1$, we have $Score(g(P_1,P_2) = \frac{2}{7}$.
5. For $g(1) = 3$, $g(2) = 2$ and $g(3) = 1$, we have $Score(g(P_1,P_2) = \frac{1}{7}$.
6. For $g(1) = 3$, $g(2) = 1$ and $g(3) = 2$, we have $Score(g(P_1,P_2) = \frac{3}{7}$.

Based on the above conclusions, we have

$$\begin{aligned} Cong_2(P_1,P_2) &= Max[Score(g(P_1,P_2))] \\ &= Max\{\frac{4}{7},\frac{2}{7},\frac{2}{7},\frac{2}{7},\frac{1}{7},\frac{3}{7}\} \\ &= \frac{4}{7}. \end{aligned}$$

According to P_1 and P_2, we have

$E_1(x_1,x_3) = E_2(x_1,x_3) = 1$
$E_1(x_4,x_6) = E_2(x_4,x_6) = 1$

i.e., $\langle x_1,x_3\rangle$ and $\langle x_4,x_6\rangle$ are *DBF* of P_1 and P_2, hence, $U^* = \{x_1,x_3,x_4,x_6\}$ and $|U^*| = 4$. Because there are no 2-multiple absolute independent elements, *i.e.*, $U_* = \emptyset$. Finally, we have

$$Cong_3(P_1,P_2) = Cong_3(E_1,E_2) = \frac{4+0}{7} = \frac{4}{7}.$$

In addition, if we add partition $P_3 = \{C_1(=\{x_1,x_3\}),C_2(=\{x_2,x_4,x_5\}),C_3(=\{x_6,x_7\})\}$ in this example, according to Eq. (2), we can obtain

$$E_1(x_1,x_3) = E_2(x_1,x_3) = E_3(x_1,x_3) = 1.$$

It is easy to obtain $U^{*3} = \{x_1,x_3\}, U_{*3} = \emptyset$, hence,

$$\begin{aligned} Cong_3(P_1,P_2,P_3) &= Cong_3(E_1,E_2,E_3) \\ &= \frac{2+0}{7} = \frac{2}{7}. \end{aligned}$$

4. Indexes for Selecting Partitions

Formally, $Cong_i(i = 1,2,3)$ of two partitions only provides us similarity of two partitions. It is difficult to tell us which one of two partitions is best. To select a better partition from all partitions of U, we need the following indexes.

4.1. Stability Entropy of Partition

For a fixed partition P of U, stability entropy of P, denoted by SE_P, is calculated as follows:

$$SE_P = 1 - \frac{S_P^2}{(\bar{x}_P-1)^2+(n-\bar{x}_P-1)^2}, \tag{3}$$

in which, $n = |U|$, m is the number of classes of P,

$$\begin{aligned} \bar{x}_P &= \frac{n}{m}, & (4) \\ S_P^2 &= \sum_{i=1}^{m}(\bar{x}_P-a_i)^2, & (5) \end{aligned}$$

a_i is the number of the elements of the $i-$th class of P, *i.e.*, $a_i = |A_i|$.

Theorem 8. *For any partition P of U, $0 \leqslant SE_P \leqslant 1$.*
Proof. For any partition P of U,

$$\begin{aligned} SE_P &= 1-\frac{S_P^2}{(\bar{x}_P-1)^2+(n-\bar{x}_P-1)^2} \\ &= 1-\frac{S_P^2}{(\frac{n}{m}-1)^2+(n-\frac{n}{m}-1)^2} \\ &= 1-\frac{m^2S_P^2}{(n-m)^2+(m(n-1)-n)^2} \\ &= 1-\frac{\sum_{i=1}^{m}(n-ma_i)^2}{(n-m)^2+(m(n-1)-n)^2}. \end{aligned}$$

Due to $a_i \geqslant 1$, $(n-ma_i)^2 \leqslant (n-m)^2$ and

$$\sum_{i=1}^{m}(n-ma_i)^2 \leqslant m(n-m)^2.$$

On the other hand,

$$\begin{aligned}&(n-m)^2+(m(n-1)-n)^2-m(n-m)^2\\ = \;&(m-2)((m-1)n^2-m^2),\end{aligned}$$

Due to $1 \leqslant m \leqslant n$, $(m-2)((m-1)n^2-m^2) \geqslant 0$, *i.e.*, $(n-m)^2+(m(n-1)-n)^2 \geqslant m(n-m)^2$,

$$0 \leqslant \frac{\sum\limits_{i=1}^{m}(n-ma_i)^2}{(n-m)^2+(m(n-1)-n)^2} \leqslant 1,$$

hence, for any partition P of U, $0 \leqslant SE_P \leqslant 1$. □

Definition 7. A partition P is called an extreme partition of the element $a \in U$, denoted by $P_a^{\top}$, if $\{a\}$ is a class of P, and $U-\{a\}$ is a class of P, *i.e.*, $P_a^{\top} = \{\{a\}, U-\{a\}\}$.

Theorem 9. *Stability entropy SE_P satisfies the following properties:*

1. $SE_{P^*} = 1$;
2. $SE_{P_*} = 1$;
3. *For any $a \in U$, $SE_{P_a^{\top}} = 0$.*

Proof. Assume that P^*, P_* and $P_a^{\top}$ are partitions of U, respectively,

1. Due to $P^* = \{U\}$, $m = 1$. According to Eq. (4), we have

$$\bar{x}_{P^*} = |U|, a_1 = |U|,$$

according to Eq. (5), we have

$$\begin{aligned}S^2_{P^*} &= \sum_{i=1}^{1}(\bar{x}_P - a_i)^2\\ &= (\bar{x}_P - a_1)^2\\ &= (|U|-|U|)^2 = 0.\end{aligned}$$

According to the Eq.(3), we have

$$\begin{aligned}SE_{P^*} &= 1-\frac{S^2_{P^*}}{(\bar{x}_{P^*}-1)^2+(n-\bar{x}_{P^*}-1)^2}\\ &= 1-\frac{S^2_{P^*}}{(n-1)^2+(n-n-1)^2}\\ &= 1-0=1.\end{aligned}$$

2. Due to $P_* = \{\{a_1\},\{a_2\},\cdots,\{a_n\}\}$, $m = |U| = n$. We know that for any $i \in \{1,2,\cdots,n\}$,

$$\bar{x}_{P_*} = 1, a_i = 1,$$

according to Eq. (5), we have

$$\begin{aligned}S^2_{P_*} &= \sum_{i=1}^{n}(\bar{x}_{P_*} - a_i)^2\\ &= \sum_{i=1}^{n}(1-1)^2 = 0.\end{aligned}$$

According to the Eq.(3), we have

$$\begin{aligned}SE_{P_*} &= 1-\frac{S^2_{P_*}}{(\bar{x}_{P_*}-1)^2+(n-\bar{x}_{P_*}-1)^2}\\ &= 1-\frac{S^2_{P_*}}{(1-1)^2+(n-1-1)^2}\\ &= 1-0=0.\end{aligned}$$

3. For any $a \in U$, due to $P_a^{\top} = \{\{a\}, U-\{a\}\}$, $m = 2$. We know

$$\bar{x}_{P_a^{\top}} = \frac{n}{2}, a_1 = 1, a_2 = n-1,$$

according to Eq. (5), we have

$$\begin{aligned}S^2_{P_a^{\top}} &= \sum_{i=1}^{2}(\bar{x}_{P_a^{\top}} - a_i)^2\\ &= (\frac{n}{2}-1)^2+(\frac{n}{2}-n+1)^2\\ &= \frac{(n-2)^2}{2}.\end{aligned}$$

According to the Eq.(3), we have

$$\begin{aligned}SE_{P_a^{\top}} &= 1-\frac{S^2_{P_a^{\top}}}{(\bar{x}_{P_a^{\top}}-1)^2+(n-\bar{x}_{P_a^{\top}}-1)^2}\\ &= 1-\frac{\frac{(n-2)^2}{2}}{(\frac{n}{2}-1)^2+(n-\frac{n}{2}-1)^2}\\ &= 1-1=0.\end{aligned}$$

□

Intuitively, SE_P expresses stability of the number of classes of partition P, in which, we select $\bar{x}_P$ (the average number of classes) as a level value, and if $\sum_{i=1}^{m}(\bar{x}_P - a_i)^2 \to 0$, then $SE_P \to 1$.

Example 3. Continue Example 2. For P_1 and P_2, $m = 3$ and $m = 2$, respectively. According to SE_P, we have

$$\begin{aligned}
\bar{x}_{P_1} &= \frac{7}{3}, \bar{x}_{P_2} = \frac{7}{2} = 3.5, \\
S_{P_1}{}^2 &= (\bar{x}_{P_1} - 3)^2 + (\bar{x}_{P_1} - 3)^2 + (\bar{x}_{P_1} - 1)^2 = \frac{8}{3}, \\
S_{P_2}{}^2 &= (\bar{x}_{P_2} - 3)^2 + (\bar{x}_{P_2} - 4)^2 = 0.5, \\
SE_{P_1} &= 1 - \frac{S_{P_1}{}^2}{(\bar{x}_{P_1} - 1)^2 + (7 - \bar{x}_{P_1} - 1)^2} \\
&= 1 - \frac{\frac{8}{3}}{\frac{49}{3}} = \frac{41}{49} \doteq 0.84, \\
SE_{P_2} &= 1 - \frac{S_{P_2}{}^2}{(\bar{x}_{P_2} - 1)^2 + (7 - \bar{x}_{P_2} - 1)^2} \\
&= 1 - \frac{\frac{1}{2}}{\frac{25}{2}} \\
&= \frac{24}{25} = 0.96.
\end{aligned}$$

Hence, $SE_{P_1} < SE_{P_2}$. Intuitively, we consider that stability of the partition P_1 is less than stability of the partition P_2.

4.2. Accuracy rate of Partition

From the information granulations point of view [23,24], every information granulation is understood by knowledge, the more information granulations, the more knowledge. It is well known that classes of partition P are special information granulations of U, based on information granulations, in this paper, we define the following index to evaluate a partition of U, it is also called accuracy rate (AR_P) of the partition P.

$$AR_P = \frac{m}{|U|}, \tag{6}$$

in which, m is the number of classes of P. According to Eq. (6), for special partitions of U, we have

$$\begin{aligned}
AR_{P^*} &= \frac{1}{|U|}, \\
AR_{P_*} &= 1, \\
AR_{P^\top} &= \frac{2}{|U|}.
\end{aligned}$$

In the above equations, if $|U| \longrightarrow \infty$, then accuracy rate $AR_P \longrightarrow 0$. Due to $m \leqslant |U|$, hence, for any partition P,

$$0 \leqslant AR_P \leqslant 1.$$

Corollary 10. *$AR_P = 1$ if and only if $P = P_*$.*

In Example 3, we have

$$AR_{p_1} = \frac{3}{7}, AR_{p_2} = \frac{2}{7}.$$

This means that knowledge of P_1 is more than P_2.

4.3. Validity entropy of Partition

From the practical point of view, it is difficult to evaluate a partition P of U. In many cases, selecting a partition P of U is associated with many aspects, from the attributes selection point of view, selecting a partition P of U is equal to selecting attributes. In this paper, SE_P and AR_P are only partial evaluations of P, respectively. By integrating SE_P and AR_P, we propose validity entropy (VE_P) of P as follow:

$$VE_P = w \times SE_P + (1 - w) \times AR_P, \tag{7}$$

in which, $w \in [0,1]$. w and $1 - w$ are understood as weights of SE_P and AR_P, respectively.

Corollary 11. *For any $w \in [0,1]$, $VE_{P_*} = 1$.*

As a special case, if $w = 0.5$, then VE_P is the average of SE_P and AR_P, *e.g.*, in Example 3, let

$w = 0.5$, then

$$\begin{aligned} VE_{P_1} &= \frac{SE_{P_1} + AP_{P_1}}{2} \\ &= \frac{\frac{41}{49} + \frac{3}{7}}{2} = \frac{31}{49} \doteq 0.63, \\ VE_{P_2} &= \frac{SE_{P_2} + AP_{P_2}}{2} \\ &= \frac{\frac{24}{25} + \frac{2}{7}}{2} = \frac{109}{175} \doteq 0.62. \end{aligned}$$

Example 4. Let $U = \{x_1, x_2, \cdots, x_{100}\}$, partitions $P_1 = \{A_1, A_2, A_3, A_4, A_5\}$ and $P_2 = \{B_1, B_2, B_3\}$, in which,

$$\begin{aligned} A_1 &= \{x_1, x_2, \cdots, x_{20}\}, \\ A_2 &= \{x_{21}, x_{22}, \cdots, x_{70}\}, \\ A_3 &= \{x_{71}, x_{72}, \cdots, x_{80}\}, \\ A_4 &= \{x_{81}, x_{82}, \cdots, x_{95}\}, \\ A_5 &= \{x_{96}, x_{97}, x_{98}, x_{99}, x_{100}\}, \\ B_1 &= \{x_1, x_2, \cdots, x_{30}\}, \\ B_2 &= \{x_{31}, x_{32}, \cdots, x_{60}\}, \\ B_3 &= \{x_{61}, x_{62}, \cdots, x_{100}\}. \end{aligned}$$

According to Eqs. (3), (4), (5) and (6), it is not difficult to obtain that

$$\begin{aligned} \bar{x}_{P_1} &= \frac{100}{5} = 20,\ \bar{x}_{P_2} = \frac{100}{3}, \\ {S_{P_1}}^2 &= (\bar{x}_{P_1} - 20)^2 + (\bar{x}_{P_1} - 50)^2 + (\bar{x}_{P_1} - 10)^2 + \\ & \quad (\bar{x}_{P_1} - 15)^2 + (\bar{x}_{P_1} - 5)^2 \\ &= 1250, \\ {S_{P_2}}^2 &= (\bar{x}_{P_2} - 30)^2 + (\bar{x}_{P_2} - 30)^2 + (\bar{x}_{P_2} - 40)^2 \\ &= \frac{200}{3}, \\ SE_{P_1} &= 1 - \frac{{S_{P_1}}^2}{(\bar{x}_{P_1} - 1)^2 + (100 - \bar{x}_{P_1} - 1)^2} \\ &= 1 - \frac{1250}{19^2 + 79^2} \\ &\doteq 0.81, \\ SE_{P_2} &= 1 - \frac{{S_{P_2}}^2}{(\bar{x}_{P_2} - 1)^2 + (100 - \bar{x}_{P_2} - 1)^2} \\ &= 1 - \frac{\frac{200}{3}}{(\frac{97}{3})^2 + (\frac{197}{3})^2} \\ &\doteq 0.99, \\ AR_{p_1} &= 0.05, AR_{p_2} = 0.03. \end{aligned}$$

Let $w = 0.3$, according to Eq. (7), we have

$$\begin{aligned} VE_{p_1} &= 0.3 \times SE_{P_1} + 0.7 \times AR_{P_1} \\ &\doteq 0.3 \times 0.81 + 0.7 \times 0.05 = 0.278, \\ VE_{p_2} &= 0.3 \times SE_{P_2} + 0.7 \times AR_{P_2} \\ &\doteq 0.3 \times 0.99 + 0.7 \times 0.03 = 0.318. \end{aligned}$$

Due to $VE_{p_1} < VE_{p_2}$, intuitively, the partition P_2 is better than the partition P_1, and we can select the partition P_2 to solve the corresponding problem.

5. Illustrative example

In this section, we explain our method in evaluating environment pollution. Department of the Environment often selects many facts to evaluate environment pollution of some areas, *e.g.*, air and water, or soil and crops. Let five areas be $U = \{u_1, u_2, u_3, u_4, u_5\}$. Their environment pollution information (obtained from Department of the Environment) are shown in Table 1 and Table 2, respectively.

Table 1. Evaluating data based on air and water.

	u_1	u_2	u_3	u_4	u_5
air	5	2	5	1	2
water	5	3	5	5	4

Table 2. Evaluating data based on soil and crops.

	u_1	u_2	u_3	u_4	u_5
soil	3	4	2	3	5
crops	2	5	3	1	1

Based on Table 1 and Table 2, firstly, we use the following fuzzy clustering method [4] to classify U:

1. Establishing fuzzy similar matrix on U, *i.e.*, $\widetilde{R}_{|U|\times|U|} = (r_{ij})$ and

$$r_{ij} = \begin{cases} 1, & i = j; \\ 1 - c\sum_{k=1}^{m}(|x_{ik}| - |x_{jk}|), & i \neq j. \end{cases}$$

   In which, $i, j \in \{1,2,3,4,5\}$, m is the number of evaluating factors, *e.g.*, in Table 1, $x_{31} = 5$ is the evaluating value of air of area u_3, c is a parameter decided by experts.

2. Obtaining fuzzy transitive closure $\widetilde{R}^*$ of $\widetilde{R}_{|U|\times|U|}$, *i.e.*, $\widetilde{R}^*$ satisfies that (1) $\widetilde{R}^* \circ \widetilde{R}^* = \widetilde{R}^*$; (2) $\widetilde{R}^* = (\widetilde{R}_{|U|\times|U|})^p = \widetilde{R}_{|U|\times|U|} \circ \cdots \circ \widetilde{R}_{|U|\times|U|}$, $\circ$ is multiplication of fuzzy matrix.

3. Selecting $\lambda-$level value to obtain classifying matrix, *i.e.*, $\widetilde{R}^*_\lambda$ is an equivalent relation.

According to the above mentioned steps, for Table 1 and Table 2, we select $c = 0.2$, and obtain the following two fuzzy similar matrixes on U:

$$\widetilde{R1} = \begin{pmatrix} 1 & 0 & 1 & 0.2 & 0.2 \\ 0 & 1 & 0 & 0.4 & 0.8 \\ 1 & 0 & 1 & 0.2 & 0.2 \\ 0.2 & 0.4 & 0.2 & 1 & 0.6 \\ 0.2 & 0.8 & 0.2 & 0.6 & 1 \end{pmatrix},$$

$$\widetilde{R2} = \begin{pmatrix} 1 & 0.2 & 0.6 & 0.8 & 0.4 \\ 0.2 & 1 & 0.2 & 0 & 0 \\ 0.6 & 0.2 & 1 & 0.4 & 0.2 \\ 0.8 & 0 & 0.4 & 1 & 0.6 \\ 0.4 & 0 & 0.2 & 0.6 & 1 \end{pmatrix}.$$

It is easy to check that the following matrixes are fuzzy transitive closures of $\widetilde{R1}$ and $\widetilde{R2}$, respectively.

$$\widetilde{R1}^* = \begin{pmatrix} 1 & 0.2 & 1 & 0.2 & 0.2 \\ 0.2 & 1 & 0.2 & 0.6 & 0.8 \\ 1 & 0.2 & 1 & 0.2 & 0.2 \\ 0.2 & 0.6 & 0.2 & 1 & 0.6 \\ 0.2 & 0.8 & 0.2 & 0.6 & 1 \end{pmatrix}$$

$$\widetilde{R2}^* = \begin{pmatrix} 1 & 0.2 & 0.6 & 0.8 & 0.6 \\ 0.2 & 1 & 0.2 & 0.2 & 0.2 \\ 0.6 & 0.2 & 1 & 0.6 & 0.6 \\ 0.8 & 0.2 & 0.6 & 1 & 0.6 \\ 0.6 & 0.2 & 0.6 & 0.6 & 1 \end{pmatrix}$$

For $\widetilde{R1}^*$, we select $\alpha = 0.8$ and 0.6, the corresponding classifying matrixes are:

$$\widetilde{R1}^*_{0.8} = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{pmatrix}$$

$$\widetilde{R1}^*_{0.6} = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 \end{pmatrix}$$

The corresponding classes of U are as follows, respectively,

$$\begin{aligned} U^{R_1}_{0.8} &= \{\{u_1,u_3\},\{u_2,u_5\},\{u_4\}\}, \\ U^{R_1}_{0.6} &= \{\{u_1,u_3\},\{u_2,u_4,u_5\}\}. \end{aligned}$$

According to Eqs. (3), (6) and (7), let $w = 0.3$. We obtain

$$VE_{U^{R_1}_{0.8}} = 0.772, VE_{U^{R_1}_{0.6}} = 0.693.$$

Due to $VE_{U^{R_1}_{0.8}} > VE_{U^{R_1}_{0.6}}$, we select $VE_{U^{R_1}_{0.8}}$ as the evaluation result of air and water. Similarly, we can obtain $VE_{U^{R_2}_{0.8}} = 0.861$ as the evaluation result of soil and crops.

Compared $VE_{U^{R_1}_{0.8}} = 0.772$ with $VE_{U^{R_2}_{0.8}} = 0.861$, we select soil and crops to evaluate environment pollution of $U = \{u_1, u_2, u_3, u_4, u_5\}$.

6. Conclusion

In the real world practice, we always face to select a better partition to help us make decisions. In this paper, we analyze the relation of elements in two partitions of the same set U, define basic factors DBF and nBF, independent factors DIF and nIF, absolute independent elements AIE and $nAIE$, present existence conditions of them. Then, we provide a new degree of congruence or similarity of two partitions $Cong_3$ based on DBF and AIE. To select a better partition from all partitions of U, we provide indexes SE_P, AR_P and VE_P, an illustrative example is given to show their application.

Acknowledgments

This work is partly supported by the research fund of sichuan key laboratory of intelligent network information processing (SGXZD1002-10), the key laboratory of the radio signals intelligent processing (Xihua university) (XZD0818-09) and the national natural science foundation of China (60875034).

References

1. Z. Deng, Z. Wang, A new fast vertical method for mining frequent pattern, International Journal of Computational Intelligence Systems, 3(6), 733-744 (2010).
2. H. Doreswamy, M.N. Vanajaskhi, Similarity measuring approach for engineering materials selection, International Journal of Computational Intelligence Systems, 3(1), 115-122 (2010).
3. P.G. Kumar, Fuzzy classifier design using modified genetic algorithm, International Journal of Computational Intelligence Systems, 3(3), 334-342 (2010).
4. A. Kaufmann and M.M. Gupta, Fuzzy Mathematical Models in Engineering and Management Science, Elsevier Science Publishers, New York, 1988.
5. A. Kandel, M. Last and H. Bunke, Data Mining and computational Intelligence, Springer-Verlag: Herdelberg, 2001.
6. O. Kaynak, L.A. Zadeh, B. Tirksen and I.J. Rudas, Computational Intelligence: Soft Computing and Fuzzy-Neuro Integration Applications, Springer-Verlag, Berlin, 1998.
7. J. Wu, L. Liang, M. Song, Performance based clustering for benchmarking of container ports: an application of dea and cluster analysis technique, International Journal of Computational Intelligence Systems, 3(6), 709-722 (2010).
8. J. Wang, S. Wu, G. Li, Clustering with instance and attribute level side information, International Journal of Computational Intelligence Systems, 3(6), 770-785 (2010).
9. A.P. Engelbrecht, Computational Intelligence: An Introduction, John Wiley & Sons, New York, 2003.
10. M.J.A. Berry and G. Linoff, Data Mining Techniques, John Wiley & Sons, New York, 1997.
11. A. Bargiela, W. Pedrycz, A Granular Description of ECG Signals, IEEE Transactions on Biomedical Engineering, 53(10), 1972-1982 (2006).
12. M. Chen, L. Chen, C.C. Hsu, W. Zeng, An information granulation based data mining approach for classifying imbalanced data, Information Sciences, 178, 3214-3227 (2008).
13. S. Mitra and T. Acharya, Data Mining: Multimedia. Soft Computing and Bioinformatics, New York: Wiley, 2003.
14. D. Ruan, G. Chen, E. Kerre and G. Wets, Intelligent Data Mining: Techniques and Applications, Springer-Verlag, Heidelberg, 2005.
15. B. Leary, Analysing partition: Definition, classification and explanation, Political Geography, 26, 886-908 (2007).
16. A. Berkovich, F.G. Garvan, The BG-rank of a partition and its applications, Advances in Applied Mathematics, 40, 377-400 (2008).
17. M. Merino, J.S. Astorqui, Finite element simulation to design constructive elements: An application to light gypsum plaster for partitions, Construction and Building Materials, 23, 14-27 (2009).
18. J. Montero, L. Martinez, Upgrading ideas about the concept of Soft Computing, International Journal of Computational Intelligence Systems, 3(2), 144-147 (2010).
19. S. Palanivel, M.K. Geetha, Video classification and shot detection for video retrieval applications, International Journal of Computational Intelligence Systems, 2(1), 39-50 (2009).
20. W. Pedrycz, Hierarchical architectures of fuzzy models: from type-1 fuzzy sets to information granules of higher type, International Journal of Computational Intelligence Systems, 3(2), 202-214 (2010).
21. R.R. Yager, Some measure relating partitions useful for computational intelligence, International Journal of Computational Intelligence Systems, 1, 1-18 (2008).
22. Y.Y. Yao, J.T. Yao, Granular computing as a basis for consistent classification problems, in: Proceedings of PAKDD02 Workshop on Toward the Foundation of Data Mining, pp. 101-106 (2002).
23. L.A. Zadeh, Toward a theory of fuzzy information granulation and its centrality in human reasoning and

fuzzy logic, Fuzzy Sets and Systems, 90, 111-127 (1997).
24. L.A. Zadeh, Toward a generalized theory of uncertainty (GTU)–an outline, Information Sciences, 172, 1-40 (2005).
25. L.A. Zadeh, Foreword, International Journal of Computational Intelligence Systems, 3(2), 141-143 (2008).

Published by Atlantis Press
Copyright: the authors