Measuring Uncertainty in Neighborhood Information System

Si-Yuan Jing[1], Kun She[2]

[1] **Sichuan Province University Key Laboratory of Internet Natural Language Intelligent Processing, Leshan Normal University, Leshan, 614000, China**

[2] **School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, 610000, China**

Abstract

Measuring uncertainty of information system plays an important role in rough set theory. Shannon's information entropy is an effective tool for measuring uncertainty in information system and it has been successfully applied to measure uncertainty of different systems in rough set theory. However, previous studies are only for classical rough set theory which can only deal with nominal attributes. Neighborhood rough set is a more comprehensive model which can handle numerical attributes and nominal attributes simultaneously. Some basic knowledge about neighborhood rough set is firstly studied in this paper. Neighborhood information entropy, neighborhood conditional information entropy and a measure of neighborhood mutual information are introduced respectively. Some of their important properties are also given. These results will be very helpful for understanding the essence of knowledge content and uncertainty measurement in neighborhood information systems.

Keywords: *Neighborhood information system; Uncertainty; Entropy; Mutual information; Rough set theory*.

1. Introduction

Granular computing (GrC), proposed by L. A. Zadeh in 1996 [1], is an emerging soft computing paradigm of information processing and knowledge discovering. It concerns the processing of complex information entities called information granules, which arise in the process of data abstraction and derivation of knowledge from information [2,3,10, 23]. Rough set theory (RST) is one of the most important tools in GrC and it has been proven to be effective to manage uncertainty that arises from inexact, noisy, or incomplete information. The essence behind RST it to classify objects of discourse, which are contained in a finite universe, into equivalence classes with respect to some attributes. The objects in each class are indiscernible, and this indiscernible relation induces a partition of the universe into some blocks, called knowledge granules or elemental concepts. Then, RST use these basic knowledge granules to characterize and approximate arbitrary concepts in complex universe. The focus of RST is on the ambiguity caused by limited discernibility of objects in the domain of discourse. Its key concepts are those of object indiscernibility and set approximation [4,5,6,23].

One of the major issues in RST is to study uncertainty of information (or knowledge) in an information system. Shannon's entropy is the most important tool for information uncertainty measurement [13]. The entropy of a system gives a measure of the uncertainty about its actual structure. It has been a useful mechanism for characterizing the information content in various modes and applications in many diverse fields. In RST, Shannon's entropy has been successfully applied to measure uncertainty of information systems under different situations. M. J. Wierman first used a variant of Shonnon's entropy to measure uncertainty in RST [14]. Liang et al. studied information entropy, rough entropy and knowledge granulation and their application to attribute reduction in RST [15,16]. I. Düntsch and G. Gediga used Shannon's entropy to measure uncertainty of rules in RST [18]. T. Beaubouef et al. used a variant of Shannon's entropy to measure uncertainty in rough sets and rough relational databases [19]. Liang et al. proposed a new information entropy which can be used to measure the fuzziness of rough set and rough classification [21]. Qian et al. proposed the concepts of combination entropy and combination granulation for measuring the uncertainty in complete information systems, which information gains possess intuitionistic knowledge content nature [17]. However, previous works only focus on classical rough sets. In another word, they just study the problem in information systems which only contain nominal attributes (also called categorical attributes or discrete attributes).

In this paper, we study the problem of uncertainty measurement in neighborhood rough set theory (NRST). Neighborhood rough set is a variation of RST which can handle nominal attributes and numerical attributes simultaneously [7,11,12]. Moreover, literature [20] proposed a more universal NRST model to deal with incomplete information. Neighborhood information system is a knowledge representation tool in which a family of knowledge granules is induced by a neighborhood relation.

Neighborhood information entropy, neighborhood conditional information entropy and a measure of neighborhood mutual information are introduced respectively. Some of their important properties are also given. These results will be very helpful for understanding the essence of knowledge content and uncertainty measurement in neighborhood information systems.

The rest of the paper is organized as follow. Section 2 briefly recalls the knowledge about complete information system and incomplete information system. Section 3 introduces some basic knowledge of NRST. In section 4, some concepts and properties of neighborhood information entropy, neighborhood conditional information entropy and neighborhood mutual information are studied. Finally, the conclusions are given in section 5.

2. Complete and Incomplete Information System

Information system is a major form of knowledge representation in RST. It provides a convenient tool for the representation of objects in terms of their attribute values. According to whether or not there are missing data (null values), information systems can be classified into two categories: complete information system and incomplete information system.

2.1 Complete information system

An information system can be formally denoted as a pair $IS=(U,A)$, where,

(1) U is a non-empty finite set of objects;
(2) A is a non-empty finite set of attributes;
(3) For every $a\in A$, there is a mapping $a:U\to V_a$, where V_a is called the value set of a.

If V_a contains a null value for at least one attribute $a\in A$, then IS is called an incomplete information system, otherwise it is a complete information system.

Each subset of attributes $B\subseteq A$ determines a binary indistinguishable relation $IND(B)$ as follow:

$$IND(B)=\{(x,y)\in U\times U \mid \forall a\in B, a(x)=a(y)\}$$

Obviously, $IND(B)=\bigcap_{a\in B} IND(a)$

It can be easily shown that $IND(B)$ is an equivalence relation on the set U. The quotient set induced by this equivalence relation, denoted by $U/IND(B)$ or U/B, constitutes a partition of the universe. The equivalence class including x is denoted by $[x]_B$. Each equivalence class is called a knowledge granule.

2.2 Incomplete information system

It may happen that some of the attribute values for an object are missing. For example, in medical information systems there may exist a group of patients for which it is impossible to perform all the required tests. These missing values can be represented by the set of all possible values for the attribute or equivalence by the domain of the attribute. To indicate such a situation, a distinguished value, a so-called null value is usually assigned to those attributes.

As mentioned above, if V_a contains a null value for at least one attribute $a\in A$, then IS is called an incomplete information system. Furthermore, the null value can be denoted by *.

Literatures define a binary relation on U, also named tolerance relation, just as follow:

$$SIM(B)=\{(x,y)\in U\times U \mid \forall a\in B, a(x)=a(y) \text{ or } x=* \text{ or } y=*\}$$

It can be easily shown that $SIM(B)=\bigcap_{a\in B} SIM(a)$.

Let $S_B(u)$ denote the set $\{v\in U \mid (u,v)\in SIM(B)\}$. $S_B(u)$ is the maximal set of objects which are possibly indistinguishable by B with u. Let $U/SIM(B)$ denote the family sets induced by $SIM(B)$. A member $S_B(u)$ from $U/SIM(B)$ called a tolerance class or a granule of information in an incomplete information system. It should be noticed that the tolerance classes in $U/SIM(B)$ do not constitute a partition of U. They constitute a covering of U.

Let $IS=(U,A)$ be an incomplete information system and $B_1,B_2\subseteq A$. We say that B_2 is coarser than B_1 (or B_1 is finer than B_2), denoted by $B_1\preceq B_2$, if and only if $S_{B_1}(x)\subseteq S_{B_2}(x)$ for $\forall x\in U$. If $B_1\preceq B_2$ and $B_1\neq B_2$, we say that B_2 is strictly coarser than B_1 (or B_1 is strictly finer than B_2) and denoted by $B_1\prec B_2$. In fact, $B_1\prec B_2 \Leftrightarrow$ for $\forall x\in U$, we have that $S_{B_1}(x)\subseteq S_{B_2}(x)$, and $\exists y\in U, y\neq x$, such that $S_{B_1}(y)\subset S_{B_2}(y)$.

3. Neighborhood Rough Set Theory

Classical rough set, also called Pawlak rough set, can only deal with nominal attributes. This limits its application in practice. In this section, we briefly recall some knowledge in NRST [7,8,9,11,12], which is a much more comprehensive than the classical one.

Definition 3.1: A neighborhood information system is a triple $NIS=(U,A,\Delta)$, where U is a non-empty finite set of objects; A is a non-empty finite set of attributes; Δ represents distance function in A; A and Δ form a family of neighborhood relation on U.

If there is null value for at least one object x in each attributes, NIS is called an incomplete neighborhood information system, otherwise it is complete. Similarly, we denote the null value by $*$.

A suitable distance function is the key to a successful application of neighbor theory. In literature [20], author proposed a distance function which can deal with nominal attribute, numerical attribute and set-valued attribute as well as missing data. In addition to the distance function given above, there are a number of distances for heterogeneous features and missing data [22]. Generally, for $\forall x_1,x_2,x_3\in U$, distance function Δ usually satisfies:

(1) $\Delta(x_1,x_2)\ge 0$, $\Delta(x_1,x_2)=0$ if and only if $x_1=x_2$;

(2) $\Delta(x_1,x_2)=\Delta(x_2,x_1)$;

(3) $\Delta(x_1,x_3)\le\Delta(x_1,x_2)+\Delta(x_2,x_3)$.

Definition 3.2: Given a neighborhood information system $NIS=(U,A,\Delta)$, threshold $\delta\ge 0$, we can define a binary relation on U, called neighborhood relation:

$$SIM_N^{\delta}(A)=\left\{(x,y)\middle|\forall B\subseteq A,\Delta_B(x,y)\le\delta\right\}$$

It can be easily found that $SIM_N^{\delta}(A)$ is reflexive and symmetrical but not transitive. It induces a covering on U. The information granules induced by $SIM_N^{\delta}(A)$ can be defined as:

$$SN_A^{\delta}(x)=\left\{y\middle|(x,y)\in SIM_N^{\delta}(A),y\in U\right\}$$

Proposition 3.1: Given a neighborhood relation $SIM_N^{\delta}(A)$ on neighborhood information system $NIS=(U,A,\Delta)$, $B\subseteq A$, it includes some properties as follows:

(1) $SIM_N^{\delta}(B)=\bigcap_{a\in B}SIM_N^{\delta}(a)$;

(2) $SN_B^{\delta}(x)\ne\varnothing$;

(3) $\bigcup_{x\in U}SN_B^{\delta}(x)=U$.

The proof procedures are omitted because it's obvious.

$SN_B^{\delta}(x)$ is the neighborhood information granule centered with sample x and the size of the neighborhood depends on threshold δ. More samples fall into the neighborhood of x if δ takes a great value.

With the above discussion, we can see there are two key factors to impact on the neighborhood. One is the used distance, the other is threshold δ. The first one determines the shape of neighborhoods and the latter controls the size of neighborhood granules. Furthermore, we can also see that a neighborhood granule degrades to an equivalent class if we let $\delta=0$. In this case, the samples in the same neighborhood granule are equivalent to each other and the neighborhood rough set model degenerates to Pawlak's one. Therefore, the neighborhood rough sets are a natural generalization of Pawlak rough sets.

In order to deal with heterogeneous features, some definitions to compute neighborhood of samples with mixed numerical and categorical attributes are given.

Definition 3.3: Given $B_1\subseteq A$ and $B_2\subseteq A$, and they represent numerical attribute set and nominal attribute set respectively. The neighborhood granule of sample x induced by B_1, B_2 and $B_1\cup B_2$ are defined as follow:

(1) $SN_{B_1}^{\delta}(x)=\left\{x_i\middle|\Delta_{B_1}(x,x_i)\le\delta,x_i\in U\right\}$;

(2) $SN_{B_2}^{\delta}(x)=\left\{x_i\middle|\Delta_{B_2}(x,x_i)=0,x_i\in U\right\}$;

(3) $SN_{B_1\cup B_2}^{\delta}(x)=\left\{x_i\middle|\Delta_{B_1}(x,x_i)\le\delta\wedge\Delta_{B_2}(x,x_i)=0,x_i\in U\right\}$.

Where $\wedge$ means "and" operation.

The first item is designed for numerical attributes; the second one is for categorical attributes, and the last one is for mixed numerical and categorical attributes. Therefore Definition 3.3 is applicable to numerical, categorical data and their mixture. According to this definition, the samples in a neighborhood granule have the same values in terms of categorical features and the distance in term of numerical features is less than threshold δ.

Proposition 3.2: Given a neighborhood information system $NIS=(U,A,\Delta)$, threshold δ, $B_1,B_2\subseteq A$. If $B_1\preceq B_2$, then $\forall x\in U$, $SN_{B_1}^{\delta}(x)\subseteq SN_{B_2}^{\delta}(x)$.

Proposition 3.3: Given a neighborhood information system $NIS=(U,A,\Delta)$, threshold δ_1 and δ_2, $B\subseteq A$. If $\delta_1\le\delta_2$, then $\forall x\in U$, $SN_B^{\delta_1}(x)\subseteq SN_B^{\delta_2}(x)$.

The proof procedures are omitted and the detailed content can be found in [7,20].

4. Uncertainty Measuring in Neighborhood Information System

In this section, neighborhood entropy, neighborhood joint entropy, neighborhood conditional entropy and neighborhood mutual information are introduced respectively. Moreover, some of their important properties are also studied.

Definition 4.1: Given a neighborhood information system $NIS=(U,A,\Delta)$ and a threshold $\delta\geq 0$. $B\subseteq A$ is a subset of attributes. The neighborhood of object $x_i\in U$ in B is denoted by $SN_B^{\delta}(x)$. Then the neighborhood uncertainty of the sample is defined as:

$$NH_{\delta}^{x_i}(B)=-\log\frac{\left|SN_B^{\delta}(x_i)\right|}{|U|} \tag{1}$$

Furthermore, the neighborhood entropy of the universe can be computed as:

$$NH_{\delta}(B)=\frac{1}{|U|}\sum_{i=1}^{|U|}NH_{\delta}^{x_i}(B)=-\frac{1}{|U|}\sum_{i=1}^{|U|}\log\frac{\left|SN_B^{\delta}(x_i)\right|}{|U|} \tag{2}$$

Since $\forall x_i\in U$, $SN_B^{\delta}(x_i)\subseteq U$, $\left|SN_B^{\delta}(x_i)\right|/|U|\leq 1$, therefore we can get $0\leq NH_{\delta}^{x_i}(B)\leq\log n$ and $0\leq NH_{\delta}(B)\leq\log n$. Moreover, $NH_{\delta}^{x_i}(B)=\log n$ if and only if $\left|SN_B^{\delta}(x_i)\right|=1$; $NH_{\delta}^{x_i}(B)=0$ if and only if $\left|SN_B^{\delta}(x_i)\right|=|U|$. $NH_{\delta}(B)=\log n$ if and only if for $\forall x_i\in U$, $\left|SN_B^{\delta}(x_i)\right|=1$; $NH_{\delta}(B)=0$ if and only if for $\forall x_i\in U$, $\left|SN_B^{\delta}(x_i)\right|=|U|$.

Proposition 4.1: Given a neighborhood information system $NIS=(U,A,\Delta)$ and two thresholds $\delta_1\geq 0$, $\delta_2\geq 0$. $B\subseteq A$ is a subset of attributes. If $\delta_1\leq\delta_2$, $NH_{\delta_1}(B)\geq NH_{\delta_2}(B)$.

Proof: Since $\delta_1\leq\delta_2$, for $\forall x_i\in U$, we have $SN_B^{\delta_1}(x_i)\subseteq SN_B^{\delta_2}(x_i)$, then $\left|SN_B^{\delta_1}(x_i)\right|\leq\left|SN_B^{\delta_2}(x_i)\right|$. Hence,

$$NH_{\delta_1}(B)=-\frac{1}{|U|}\sum_{i=1}^{|U|}\log\frac{\left|SN_B^{\delta_1}(x_i)\right|}{|U|}\geq -\frac{1}{|U|}\sum_{i=1}^{|U|}\log\frac{\left|SN_B^{\delta_2}(x_i)\right|}{|U|}=NH_{\delta_2}(B)$$

This completes the proof.

Proposition 4.1 states that the neighborhood entropy increases with the decrease of distance threshold δ, in other words, neighborhood information granules become small.

Proposition 4.2: Given a neighborhood information system $NIS=(U,A,\Delta)$ and a threshold $\delta\geq 0$. $B_1,B_2\subseteq A$ are two subsets of attributes. If $B_1\preceq B_2$, $NH_{\delta}(B_1)\geq NH_{\delta}(B_2)$.

Proof: Since $B_1\preceq B_2$, for $\forall x_i\in U$, we have $SN_{B_1}^{\delta}(x_i)\subseteq SN_{B_2}^{\delta}(x_i)$, then $\left|SN_{B_1}^{\delta}(x_i)\right|\leq\left|SN_{B_2}^{\delta}(x_i)\right|$. Hence,

$$NH_{\delta}(B_1)=-\frac{1}{|U|}\sum_{i=1}^{|U|}\log\frac{\left|SN_{B_1}^{\delta}(x_i)\right|}{|U|}\geq -\frac{1}{|U|}\sum_{i=1}^{|U|}\log\frac{\left|SN_{B_2}^{\delta}(x_i)\right|}{|U|}=NH_{\delta}(B_2)$$

This completes the proof.

Definition 4.2: Given a neighborhood information system $NIS=(U,A,\Delta)$ and a threshold $\delta\geq 0$. $B_1,B_2\subseteq A$ are two subsets of attributes. Then the neighborhood joint entropy is defined as:

$$NH_{\delta}(B_1,B_2)=-\frac{1}{|U|}\sum_{i=1}^{|U|}\log\frac{\left|SN_{B_1\cup B_2}^{\delta}(x_i)\right|}{|U|} \tag{3}$$

Where $SN_{B_1\cup B_2}^{\delta}(x_i)$ is the neighborhood of object x_i in attribute space $B_1\cup B_2$.

Proposition 4.3: $NH_{\delta}(B_1,B_2)\geq NH_{\delta}(B_1)$, $NH_{\delta}(B_1,B_2)\geq NH_{\delta}(B_2)$

Proof: Since $B_1\cup B_2\preceq B_1$, $B_1\cup B_2\preceq B_2$, therefore, for $\forall x_i\in U$, we have $SN_{B_1\cup B_2}^{\delta}(x_i)\subseteq SN_{B_1}^{\delta}(x_i)$, $SN_{B_1\cup B_2}^{\delta}(x_i)\subseteq SN_{B_2}^{\delta}(x_i)$. Then, $\left|SN_{B_1\cup B_2}^{\delta}(x_i)\right|\leq\left|SN_{B_1}^{\delta}(x_i)\right|$, $\left|SN_{B_1\cup B_2}^{\delta}(x_i)\right|\leq\left|SN_{B_2}^{\delta}(x_i)\right|$ .Hence, $NH_{\delta}(B_1,B_2)\geq NH_{\delta}(B_1)$, $NH_{\delta}(B_1,B_2)\geq NH_{\delta}(B_2)$.

This completes the proof.

Definition 4.3: Given a neighborhood information system $NIS=(U,A,\Delta)$ and a threshold $\delta \geq 0$. $B_1, B_2 \subseteq A$ are two subsets of attributes. The neighborhood conditional entropy of B_1 to B_2 is defined as:

$$NH_\delta(B_1|B_2) = -\frac{1}{|U|}\sum_{i=1}^{|U|}\log\frac{\left|SN_{B_1\cup B_2}^\delta(x_i)\right|}{\left|SN_{B_2}^\delta(x_i)\right|} \tag{4}$$

Proposition 4.4: $NH_\delta(B_1|B_2) = NH_\delta(B_1,B_2) - NH_\delta(B_2)$

Proof:

$$\begin{aligned} & NH_\delta(B_1,B_2) - NH_\delta(B_2) \\ &= -\frac{1}{|U|}\sum_{i=1}^{|U|}\log\frac{\left|SN_{B_1\cup B_2}^\delta(x_i)\right|}{|U|} + \frac{1}{|U|}\sum_{i=1}^{|U|}\log\frac{\left|SN_{B_2}^\delta(x_i)\right|}{|U|} \\ &= -\frac{1}{|U|}\sum_{i=1}^{|U|}\left(\log\frac{\left|SN_{B_1\cup B_2}^\delta(x_i)\right|}{|U|} - \log\frac{\left|SN_{B_2}^\delta(x_i)\right|}{|U|}\right) \\ &= -\frac{1}{|U|}\sum_{i=1}^{|U|}\log\frac{\left|SN_{B_1\cup B_2}^\delta(x_i)\right|}{\left|SN_{B_2}^\delta(x_i)\right|} = NH_\delta(B_1|B_2) \end{aligned}$$

This completes the proof.

Proposition 4.5: Given a neighborhood information system $NIS=(U,A,\Delta)$ and a threshold $\delta \geq 0$. $B_1, B_2 \subseteq A$ are two subsets of attributes. If $B_1 \preceq B_2$, then $NH_\delta(B_2|B_1) = 0$.

Proof: Since $B_1 \preceq B_2$, hence for $\forall x_i \in U$, we have $SN_{B_1}^\delta(x_i) \subseteq SN_{B_2}^\delta(x_i)$.

Then $SN_{B_1\cup B_2}^\delta(x_i) = SN_{B_1}^\delta(x_i) \cap SN_{B_2}^\delta(x_i) = SN_{B_1}^\delta(x_i)$.

Hence,

$$\begin{aligned} NH_\delta(B_2|B_1) &= -\frac{1}{|U|}\sum_{i=1}^{|U|}\log\frac{\left|SN_{B_1\cup B_2}^\delta(x_i)\right|}{\left|SN_{B_1}^\delta(x_i)\right|} \\ &= -\frac{1}{|U|}\sum_{i=1}^{|U|}\log\frac{\left|SN_{B_1}^\delta(x_i)\right|}{\left|SN_{B_1}^\delta(x_i)\right|} \\ &= 0 \end{aligned}$$

This completes the proof.

Proposition 4.6: Given a neighborhood information system $NIS=(U,A,\Delta)$ and a threshold $\delta \geq 0$. $B_1, B_2, B_3 \subseteq A$ are two subsets of attributes. If $B_1 \preceq B_2$, then $NH_\delta(B_1|B_3) \geq NH_\delta(B_2|B_3)$.

Proof: Since $B_1 \preceq B_2$, hence for $\forall x_i \in U$, we have $SN_{B_1}^\delta(x_i) \subseteq SN_{B_2}^\delta(x_i)$.

Hence, $\left(SN_{B_1}^\delta(x_i) \cap SN_{B_3}^\delta(x_i)\right) \subseteq \left(SN_{B_2}^\delta(x_i) \cap SN_{B_3}^\delta(x_i)\right)$,

and, $\left|SN_{B_1}^\delta(x_i) \cap SN_{B_3}^\delta(x_i)\right| \leq \left|SN_{B_2}^\delta(x_i) \cap SN_{B_3}^\delta(x_i)\right|$.

Hence,

$$\begin{aligned} NH_\delta(B_1|B_3) &= -\frac{1}{|U|}\sum_{i=1}^{|U|}\log\frac{\left|SN_{B_1\cup B_3}^\delta(x_i)\right|}{\left|SN_{B_3}^\delta(x_i)\right|} \\ &= -\frac{1}{|U|}\sum_{i=1}^{|U|}\log\frac{\left|SN_{B_1}^\delta(x_i) \cap SN_{B_3}^\delta(x_i)\right|}{\left|SN_{B_3}^\delta(x_i)\right|} \\ &\geq -\frac{1}{|U|}\sum_{i=1}^{|U|}\log\frac{\left|SN_{B_2}^\delta(x_i) \cap SN_{B_3}^\delta(x_i)\right|}{\left|SN_{B_3}^\delta(x_i)\right|} \\ &= NH_\delta(B_2|B_3) \end{aligned}$$

This completes the proof.

Proposition 4.7: Given a neighborhood information system $NIS=(U,A,\Delta)$ and a threshold $\delta \geq 0$. $B_1, B_2, B_3 \subseteq A$ are two subsets of attributes. If $B_3 \preceq B_1 \preceq B_2$, then $NH_\delta(B_3|B_1) \leq NH_\delta(B_3|B_2)$.

Proof: Since $B_3 \preceq B_1 \preceq B_2$, hence for $\forall x_i \in U$, we have $SN_{B_3}^\delta(x_i) \subseteq SN_{B_1}^\delta(x_i) \subseteq SN_{B_2}^\delta(x_i)$, and $\left|SN_{B_3}^\delta(x_i)\right| \leq \left|SN_{B_1}^\delta(x_i)\right| \leq \left|SN_{B_2}^\delta(x_i)\right|$.

Then, we have $SN_{B_3}^\delta(x_i) \cap SN_{B_1}^\delta(x_i) = SN_{B_3}^\delta(x_i) \cap SN_{B_2}^\delta(x_i) = SN_{B_3}^\delta(x_i)$.

Hence,

$$\begin{aligned} & NH_\delta(B_3|B_1) - NH_\delta(B_3|B_2) \\ &= -\frac{1}{|U|}\sum_{i=1}^{|U|}\log\frac{\left|SN_{B_3\cup B_1}^\delta(x_i)\right|}{\left|SN_{B_1}^\delta(x_i)\right|} - \left(-\frac{1}{|U|}\sum_{i=1}^{|U|}\log\frac{\left|SN_{B_3\cup B_2}^\delta(x_i)\right|}{\left|SN_{B_2}^\delta(x_i)\right|}\right) \\ &= -\frac{1}{|U|}\sum_{i=1}^{|U|}\log\frac{\left|SN_{B_3}^\delta(x_i) \cap SN_{B_1}^\delta(x_i)\right|}{\left|SN_{B_1}^\delta(x_i)\right|} \\ &\quad + \frac{1}{|U|}\sum_{i=1}^{|U|}\log\frac{\left|SN_{B_3}^\delta(x_i) \cap SN_{B_2}^\delta(x_i)\right|}{\left|SN_{B_2}^\delta(x_i)\right|} \\ &= -\frac{1}{|U|}\sum_{i=1}^{|U|}\log\frac{\left|SN_{B_3}^\delta(x_i)\right|}{\left|SN_{B_1}^\delta(x_i)\right|} + \frac{1}{|U|}\sum_{i=1}^{|U|}\log\frac{\left|SN_{B_3}^\delta(x_i)\right|}{\left|SN_{B_2}^\delta(x_i)\right|} \end{aligned}$$

$$= \frac{1}{|U|}\sum_{i=1}^{|U|}\left(-\log\frac{\left|SN_{B_3}^{\delta}(x_i)\right|}{\left|SN_{B_1}^{\delta}(x_i)\right|}+\log\frac{\left|SN_{B_3}^{\delta}(x_i)\right|}{\left|SN_{B_2}^{\delta}(x_i)\right|}\right)$$
$$= \frac{1}{|U|}\sum_{i=1}^{|U|}\left(\log\frac{\left|SN_{B_1}^{\delta}(x_i)\right|}{\left|SN_{B_2}^{\delta}(x_i)\right|}\right)$$
$$\le 0$$

This completes the proof.

Proposition 4.8: Given a neighborhood information system $NIS=(U,A,\Delta)$ and a threshold $\delta\ge 0$. $B_1,B_2,B_3\subseteq A$ are two subsets of attributes. If $B_1\preceq B_2\preceq B_3$, then $NH_\delta(B_3|B_1)=NH_\delta(B_3|B_2)=0$.

Proof: Since $B_1\preceq B_2\preceq B_3$, hence for $\forall x_i\in U$, we have $SN_{B_1}^{\delta}(x_i)\subseteq SN_{B_2}^{\delta}(x_i)\subseteq SN_{B_3}^{\delta}(x_i)$, and
$SN_{B_3}^{\delta}(x_i)\cap SN_{B_1}^{\delta}(x_i)=SN_{B_1}^{\delta}(x_i)$,
$SN_{B_3}^{\delta}(x_i)\cap SN_{B_2}^{\delta}(x_i)=SN_{B_2}^{\delta}(x_i)$.

Hence,

$$NH_\delta(B_3|B_1)=-\frac{1}{|U|}\sum_{i=1}^{|U|}\log\frac{\left|SN_{B_3\cup B_1}^{\delta}(x_i)\right|}{\left|SN_{B_1}^{\delta}(x_i)\right|}$$
$$=-\frac{1}{|U|}\sum_{i=1}^{|U|}\log\frac{\left|SN_{B_3}^{\delta}(x_i)\cap SN_{B_1}^{\delta}(x_i)\right|}{\left|SN_{B_1}^{\delta}(x_i)\right|}$$
$$=-\frac{1}{|U|}\sum_{i=1}^{|U|}\log\frac{\left|SN_{B_1}^{\delta}(x_i)\right|}{\left|SN_{B_1}^{\delta}(x_i)\right|}$$
$$=0$$

$$NH_\delta(B_3|B_2)=-\frac{1}{|U|}\sum_{i=1}^{|U|}\log\frac{\left|SN_{B_3\cup B_2}^{\delta}(x_i)\right|}{\left|SN_{B_2}^{\delta}(x_i)\right|}$$
$$=-\frac{1}{|U|}\sum_{i=1}^{|U|}\log\frac{\left|SN_{B_3}^{\delta}(x_i)\cap SN_{B_2}^{\delta}(x_i)\right|}{\left|SN_{B_2}^{\delta}(x_i)\right|}$$
$$=-\frac{1}{|U|}\sum_{i=1}^{|U|}\log\frac{\left|SN_{B_2}^{\delta}(x_i)\right|}{\left|SN_{B_2}^{\delta}(x_i)\right|}$$
$$=0$$

This completes the proof.

Definition 4.4: Given a neighborhood information system $NIS=(U,A,\Delta)$ and a threshold $\delta\ge 0$. $B_1,B_2\subseteq A$ are two subsets of attributes. The neighborhood mutual information of B_1 and B_2 is defined as:

$$NMI_\delta(B_1;B_2)=-\frac{1}{|U|}\sum_{i=1}^{|U|}\log\frac{\left|SN_{B_1}^{\delta}(x_i)\right|\cdot\left|SN_{B_2}^{\delta}(x_i)\right|}{|U|\cdot\left|SN_{B_1\cup B_2}^{\delta}(x_i)\right|} \quad (5)$$

Proposition 4.9: Given a neighborhood information system $NIS=(U,A,\Delta)$ and a threshold $\delta\ge 0$. $B_1,B_2\subseteq A$ are two subsets of attributes. Then following equations hold:

(1) $NMI_\delta(B_1;B_2)=NMI_\delta(B_2;B_1)$

(2) $NMI_\delta(B_1;B_2)=NH_\delta(B_1)+NH_\delta(B_2)-NH_\delta(B_1,B_2)$

(3) $NMI_\delta(B_1;B_2)=NH_\delta(B_1)-NH_\delta(B_1|B_2)$
$=NH_\delta(B_2)-NH_\delta(B_2|B_1)$

Proof: (1) This conclusion is straightforward. (2)

$$NH_\delta(B_1)+NH_\delta(B_2)-NH_\delta(B_1,B_2)$$
$$=-\frac{1}{|U|}\sum_{i=1}^{|U|}\log\frac{\left|SN_{B_1}^{\delta}(x_i)\right|}{|U|}-\frac{1}{|U|}\sum_{i=1}^{|U|}\log\frac{\left|SN_{B_2}^{\delta}(x_i)\right|}{|U|}$$
$$+\frac{1}{|U|}\sum_{i=1}^{|U|}\log\frac{\left|SN_{B_1\cup B_2}^{\delta}(x_i)\right|}{|U|}$$
$$=-\frac{1}{|U|}\sum_{i=1}^{|U|}\left(\log\frac{\left|SN_{B_1}^{\delta}(x_i)\right|}{|U|}+\log\frac{\left|SN_{B_2}^{\delta}(x_i)\right|}{|U|}\right.$$
$$\left.-\log\frac{\left|SN_{B_1\cup B_2}^{\delta}(x_i)\right|}{|U|}\right)$$
$$=-\frac{1}{|U|}\sum_{i=1}^{|U|}\left(\log\frac{\left|SN_{B_1}^{\delta}(x_i)\right|}{|U|}\frac{\left|SN_{B_2}^{\delta}(x_i)\right|}{|U|}\frac{|U|}{\left|SN_{B_1\cup B_2}^{\delta}(x_i)\right|}\right)$$
$$=-\frac{1}{|U|}\sum_{i=1}^{|U|}\left(\log\frac{\left|SN_{B_1}^{\delta}(x_i)\right|\cdot\left|SN_{B_2}^{\delta}(x_i)\right|}{|U|\cdot\left|SN_{B_1\cup B_2}^{\delta}(x_i)\right|}\right)$$
$$=NMI_\delta(B_1;B_2)$$

This completes the proof.

(3) Follows from above conclusion and proposition 4.4

$$NH_\delta(B_1)-NH_\delta(B_1|B_2)$$
$$=NH_\delta(B_1)-\left(NH_\delta(B_1,B_2)-NH_\delta(B_2)\right)$$
$$=NMI_\delta(B_1;B_2)$$

$$NH_\delta(B_2)-NH_\delta(B_2|B_1)$$
$$=NH_\delta(B_2)-\left(NH_\delta(B_1,B_2)-NH_\delta(B_1)\right)$$
$$=NMI_\delta(B_1;B_2)$$

This completes the proof.

Proposition 4.10: Given a neighborhood information system $NIS=(U,A,\Delta)$ and a threshold $\delta\geq 0$. $B_1,B_2,B_3\subseteq A$ are two subsets of attributes. If $B_3\preceq B_1\preceq B_2$, then $NMI_\delta(B_1;B_3)\geq NMI_\delta(B_2;B_3)$.

Proof: Since $B_3\preceq B_1\preceq B_2$, hence for $\forall x_i\in U$, we have $SN_{B_3}^\delta(x_i)\subseteq SN_{B_1}^\delta(x_i)\subseteq SN_{B_2}^\delta(x_i)$, and

$\left|SN_{B_3}^\delta(x_i)\right|\leq\left|SN_{B_1}^\delta(x_i)\right|\leq\left|SN_{B_2}^\delta(x_i)\right|$.

Then, we have

$SN_{B_3}^\delta(x_i)\cap SN_{B_1}^\delta(x_i)=SN_{B_3}^\delta(x_i)\cap SN_{B_2}^\delta(x_i)=SN_{B_3}^\delta(x_i)$.

Hence,

$$
\begin{aligned}
&NMI_\delta(B_1;B_3)-NMI_\delta(B_2;B_3)\\
&=-\frac{1}{|U|}\sum_{i=1}^{|U|}\log\frac{\left|SN_{B_1}^\delta(x_i)\right|\cdot\left|SN_{B_3}^\delta(x_i)\right|}{|U|\cdot\left|SN_{B_1\cup B_3}^\delta(x_i)\right|}\\
&\quad+\frac{1}{|U|}\sum_{i=1}^{|U|}\log\frac{\left|SN_{B_2}^\delta(x_i)\right|\cdot\left|SN_{B_3}^\delta(x_i)\right|}{|U|\cdot\left|SN_{B_2\cup B_3}^\delta(x_i)\right|}\\
&=-\frac{1}{|U|}\sum_{i=1}^{|U|}\left(\log\frac{\left|SN_{B_1}^\delta(x_i)\right|\cdot\left|SN_{B_3}^\delta(x_i)\right|}{|U|\cdot\left|SN_{B_1\cup B_3}^\delta(x_i)\right|}\right.\\
&\quad\left.-\log\frac{\left|SN_{B_2}^\delta(x_i)\right|\cdot\left|SN_{B_3}^\delta(x_i)\right|}{|U|\cdot\left|SN_{B_2\cup B_3}^\delta(x_i)\right|}\right)\\
&=-\frac{1}{|U|}\sum_{i=1}^{|U|}\log\frac{\left|SN_{B_1}^\delta(x_i)\right|\cdot\left|SN_{B_2\cup B_3}^\delta(x_i)\right|}{\left|SN_{B_1\cup B_3}^\delta(x_i)\right|\cdot\left|SN_{B_2}^\delta(x_i)\right|}\\
&=-\frac{1}{|U|}\sum_{i=1}^{|U|}\log\frac{\left|SN_{B_1}^\delta(x_i)\right|\cdot\left|SN_{B_3}^\delta(x_i)\right|}{\left|SN_{B_3}^\delta(x_i)\right|\cdot\left|SN_{B_2}^\delta(x_i)\right|}\\
&=-\frac{1}{|U|}\sum_{i=1}^{|U|}\log\frac{\left|SN_{B_1}^\delta(x_i)\right|}{\left|SN_{B_2}^\delta(x_i)\right|}\\
&\geq 0
\end{aligned}
$$

This completes the proof.

Proposition 4.11: Given a neighborhood information system $NIS=(U,A,\Delta)$ and a threshold $\delta\geq 0$. $B_1,B_2,B_3\subseteq A$ are two subsets of attributes. If $B_1\preceq B_2\preceq B_3$, then

$NMI_\delta(B_1;B_3)=NMI_\delta(B_2;B_3)=NH_\delta(B_3)$.

Proof: Since $B_1\preceq B_2\preceq B_3$, hence for $\forall x_i\in U$, we have $SN_{B_1}^\delta(x_i)\subseteq SN_{B_2}^\delta(x_i)\subseteq SN_{B_3}^\delta(x_i)$, and

$SN_{B_3}^\delta(x_i)\cap SN_{B_1}^\delta(x_i)=SN_{B_1}^\delta(x_i)$,
$SN_{B_3}^\delta(x_i)\cap SN_{B_2}^\delta(x_i)=SN_{B_2}^\delta(x_i)$.

Hence,

$$
\begin{aligned}
NMI_\delta(B_1;B_3)&=-\frac{1}{|U|}\sum_{i=1}^{|U|}\log\frac{\left|SN_{B_1}^\delta(x_i)\right|\cdot\left|SN_{B_3}^\delta(x_i)\right|}{|U|\cdot\left|SN_{B_1\cup B_3}^\delta(x_i)\right|}\\
&=-\frac{1}{|U|}\sum_{i=1}^{|U|}\log\frac{\left|SN_{B_1}^\delta(x_i)\right|\cdot\left|SN_{B_3}^\delta(x_i)\right|}{|U|\cdot\left|SN_{B_1}^\delta(x_i)\cap SN_{B_3}^\delta(x_i)\right|}\\
&=-\frac{1}{|U|}\sum_{i=1}^{|U|}\log\frac{\left|SN_{B_1}^\delta(x_i)\right|\cdot\left|SN_{B_3}^\delta(x_i)\right|}{|U|\cdot\left|SN_{B_1}^\delta(x_i)\right|}\\
&=-\frac{1}{|U|}\sum_{i=1}^{|U|}\log\frac{\left|SN_{B_3}^\delta(x_i)\right|}{|U|}\\
&=NH_\delta(B_3)
\end{aligned}
$$

$$
\begin{aligned}
NMI_\delta(B_2;B_3)&=-\frac{1}{|U|}\sum_{i=1}^{|U|}\log\frac{\left|SN_{B_2}^\delta(x_i)\right|\cdot\left|SN_{B_3}^\delta(x_i)\right|}{|U|\cdot\left|SN_{B_2\cup B_3}^\delta(x_i)\right|}\\
&=-\frac{1}{|U|}\sum_{i=1}^{|U|}\log\frac{\left|SN_{B_2}^\delta(x_i)\right|\cdot\left|SN_{B_3}^\delta(x_i)\right|}{|U|\cdot\left|SN_{B_2}^\delta(x_i)\cap SN_{B_3}^\delta(x_i)\right|}\\
&=-\frac{1}{|U|}\sum_{i=1}^{|U|}\log\frac{\left|SN_{B_2}^\delta(x_i)\right|\cdot\left|SN_{B_3}^\delta(x_i)\right|}{|U|\cdot\left|SN_{B_2}^\delta(x_i)\right|}\\
&=-\frac{1}{|U|}\sum_{i=1}^{|U|}\log\frac{\left|SN_{B_3}^\delta(x_i)\right|}{|U|}\\
&=NH_\delta(B_3)
\end{aligned}
$$

This completes the proof.

5. Conclusions

In this paper, the concepts of the neighborhood entropy $NH_\delta(B)$, the neighborhood joint entropy $NH_\delta(B_1,B_2)$, the neighborhood conditional entropy $NH_\delta(B_1|B_2)$ and neighborhood mutual information $NMI_\delta(B_1;B_2)$ are introduced to measure uncertainty of neighborhood information system. Some important properties of these concepts are derived. These results have a wide variety of applications, such as measuring the knowledge content, measuring the significance of an attribute set, constructing a decision tree and building a heuristic function in a heuristic reduct algorithm in neighborhood information systems.

Acknowledgments

This work is supported by the Scientific Research Fund of Sichuan Provincial Education Department (Grand No. 14ZB0247), Scientific Research Fund of Leshan Normal University (Grand No. Z1325), and Sichuan Provincial Department of Science and Technology Project (No.2014JY0036).Instead, write “F. A. Author thanks ... .” Sponsor and financial support acknowledgments are also placed here.

References

[1] L. A. Zadeh, "Fuzzy logic=computing with words", IEEE Transactions on Fuzzy System, Vol. 4, No. 1, 1996, pp.103-111.
[2] L. A. Zadeh, "Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic", Fuzzy sets and System, Vol. 19, No. 1, 1997, pp. 111-127.
[3] L. A. Zadeh, "Some reflections on soft computing, granular computing and their roles in the conception, design and utilization of information/intelligent system", Soft Computing, Vol. 2, No. 1, 1998, pp. 23-25.
[4] Z. Pawlak, "Rough sets", International Journal of Computer and Information Science, Vol. 11, No. 5, 1982, pp. 341-356.
[5] Z. Pawlak, "Rough sets - Theoretical Aspects of Reasoning about Data", Kluwer Academic Publishers, Dordrecht, 1991.
[6] Z. Pawlak, A. Skowron, "Rough Sets: Some Extensions", Information Sciences, Vol. 177, 2007, pp. 28–40.
[7] Q. H. Hu, D. R. Yu, J. F. Liu, et al. Neighborhood rough set based heterogeneous feature subset selection. Information Sciences, 2008, vol. 178, pp. 3575-3594.
[8] Q. H. Hu, L. Zhang, D. Zhang, et al., "Measuring relevance between discrete and continuous features based on neighborhood mutual information", Expert Systems with Application, Vol. 38, No. 9, 2011, pp. 10737-10750.
[9] Q. H. Hu, W. Pan, S. An, et al., "An efficient gene selection technique for cancer recognition based on neighborhood mutual information", International Journal of Machine Learning and Cybernetics, Vol. 1, 2010, pp. 63-74.
[10] T. Y. Lin, "Rough Set Approach, Granular Computing: An Emerging Paradigm", Physica-Verlag, Heidelberg, Germany, 2001.
[11] T. Y. Lin, "Neighborhood systems: mathematical models of information granulations", in 2003 IEEE International Conference on Systems, Man & Cybernetics, Washington, DC, USA, 2003.
[12] Y. Y. Yao, "Relational interpretations of neighborhood operators and rough set approximation operators", Information Sciences, Vol. 111, 1998, pp. 239–259.
[13] C. E. Shannon, "The mathematical theory of communication", The Bell System Technology Journal, Vol. 27, 1948, pp. 379-423, 623-656.
[14] M. J. Wierman, "Measuring uncertainty in rough set theory", International Journal of General Systems, Vol. 28, No. 4, 1999, pp. 283-297.
[15] J. Y. Liang, Z. Z. Shi, "The information entropy, rough entropy and knowledge granulation in rough set theory", International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, Vol. 12, No. 1, 2004, pp. 37-46
[16] J. Y. Liang, Z. B. Xu, "The algorithm on knowledge reduction in incomplete information systems", International of Uncertainty, Fuzziness and Knowledge-Based Systems, Vol. 24, No. 1, 2002, pp. 95-103.
[17] Y. H. Qian, J. Y. Liang, "Combination entropy and combination granulation in incomplete information system", Lecture Notes in Artificial Intelligence, Vol. 4062, 2006, pp.184-190.
[18] I. Düntsch, G. Gedig, "Uncertainty measures of rough set prediction", Artificial Intelligence, Vol. 106, 1998, pp. 283-297.
[19] T. Beaubouef, F. E. Petry, G. Arora, "Information-theoretic measures of uncertainty for rough sets and rough relational databases", Information Sciences, Vol. 109, 1998, pp. 185-195.
[20] S. Y. Jing, K. She, S. Ali, "A Universal neighbourhood rough sets model for knowledge discovering from incomplete heterogeneous data", Expert Systems, Vol. 30, No. 1, 2013, pp.89-96.
[21] J. Y. Liang, K. S. Chin, C. Y. Dang, et al., "A new method for measuring uncertainty and fuzziness in rough set theory", International Journal of General Systems, Vol. 31, No. 4, 2002, pp. 331-342.
[22] H. Wang, "Nearest neighbors by neighborhood counting", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 28, No. 6, 2006, pp. 942–953.
[23] S. Ali, S. Y. Jing, K. She, "Profit-Aware DVFS Enabled Resource Management of IaaS Cloud", International Journal of Computer Science Issues, Vol. 10, No. 2, 2003, pp. 237-247.

Si-Yuan Jing received the Ph. D. degree from University of Electronic Science and Technology of China, Chengdu China, in 2013. He is currently an assistant professor with Leshan Normal University. He is the member of ACM and CCF. His research interest is in computational intelligence, big data mining etc.

Kun She received the Ph. D. degree from University of Electronic Science and Technology of China, Chengdu China, in 2006. He is currently a professor with University of Electronic Science and Technology of China. He is the senior member of ACM and CCF. His research interest is in machine learning, information security.