Three-Way Weighted Entropies and Three-Way Attribute Reduction

Xianyong Zhang[1,2] and Duoqian Miao[1]

[1] Department of Computer Science and Technology, Tongji University, Shanghai 201804, P.R. China
[2] College of Mathematics and Software Science, Sichuan Normal University, Chengdu 610068, P.R. China
{zhang_xy,dqmiao}@tongji.edu.cn

Abstract. Rough set theory (RS-Theory) is a fundamental model of granular computing (GrC) for uncertainty information processing, and information entropy theory provides an effective approach for its uncertainty representation and attribute reduction. Thus, this paper hierarchically constructs three-way weighted entropies (i.e., the likelihood, prior, and posterior weighted entropies) by adopting a GrC strategy from the concept level to classification level, and it further explores three-way attribute reduction (i.e., the likelihood, prior, and posterior attribute reduction) by resorting to a novel approach of Bayesian inference. From two new perspectives of GrC and Bayesian inference, this study provides some new insights into the uncertainty measurement and attribute reduction of information theory-based RS-Theory.

Keywords: Rough set theory, uncertainty, granular computing, three-way decision, information theory, weighted entropy, Bayesian inference, attribute reduction.

1 Introduction

Rough set theory (RS-Theory) [1] is a fundamental model of granular computing (GrC) for uncertainty information processing. Information theory [2] is an important way to reflect information and measure uncertainty, and it was first introduced into RS-Theory for uncertainty representation and reduction measurement by Prof. Miao in 1997 [3]; a measurement called rough entropy was further put forward by Prof. Beaubouef in 1998 [4]. In the development of more than a decade, many systematic fruits [5-11] based on the information entropy, conditional entropy, and mutual information, have been widely used, especially for attribute reduction.

The Bayesian inference in machine learning [12] provides an effective approach for practical data processing, i.e., introducing the prior information into the likelihood function to produce the posterior probability. Following this approach, this paper mainly evolves the Bayesian probability formula in RS-Theory from a new perspective of weighted entropies, and it also explores relevant Bayesian expressions at different levels. When the weighted entropies are constructed from

D. Miao et al. (Eds.): RSKT 2014, LNAI 8818, pp. 707–719, 2014.
DOI: 10.1007/978-3-319-11740-9_65

the concept level to classification level, the GrC strategy is adopted, because GrC [13,14] is an effective structural methodology for dealing with hierarchical issues. Finally, the hierarchical weighted entropies are utilized to construct attribute reduction. In particular, three-way decision theory, proposed by Prof. Yao [15,16], plays a key role in decision making. Herein, from the three-way decision viewpoint, relevant Bayesian items and systemic reduction are also considered by using a longitudinal strategy, and we will concretely construct three-way weighted entropies and three-way reducts based on the likelihood, prior, and posterior items.

In summary, we mainly use two new perspectives of GrC and Bayesian inference to preliminary explore uncertainty measuring and attribute reduction. Thus, this study can provide some new insights into information theory-based RS-Theory. Moreover, the constructed three-way pattern regarding likelihood, prior, and posterior can partially enrich the three-way decision theory from a new perspective. Next, Section 2 provides preliminaries, Section 3 and 4 study the three-way weighted entropies at the concept and classification levels, respectively, Section 5 further discusses three-way attribute reduction, Section 6 finally provides conclusions.

2 Preliminaries

The decision table (D-Table) $(U, \mathcal{C} \cup \mathcal{D})$ serves as a main framework. Herein, $X \in U/IND(\mathcal{D}) = \{X_j : j = 1, .., m\}$, $\mathcal{A} \subseteq \mathcal{C}$, $[x]_{\mathcal{A}} \in U/IND(\mathcal{A}) = \{[x]^i_{\mathcal{A}} : i = 1, .., n\}$. $\mathcal{B} \subseteq \mathcal{A} \subseteq \mathcal{C}$ refers to the granulation relationship with a partial order - $\mathcal{A} \preceq \mathcal{B}$. If $\mathcal{A} \preceq \mathcal{B}$, then $\forall [x]_{\mathcal{B}} \in U/IND(\mathcal{B})$, $\exists k \in \mathbf{N}$, s.t., $\bigcup_{t=1}^{k} [x]^t_{\mathcal{A}} = [x]_{\mathcal{B}}$; thus, representative granular merging $\bigcup_{t=1}^{k} [x]^t_{\mathcal{A}} = [x]_{\mathcal{B}}$ can be directly utilized for verifying granulation monotonicity [17]. Moreover, $U/IND(\mathcal{B}) = \{U\}$ if $\mathcal{B} = \emptyset$, and let $\forall b \in \mathcal{B}$.

The conditional entropy and mutual information are

$$H(\mathcal{D}/\mathcal{A}) = -\sum_{i=1}^{n} p([x]^i_{\mathcal{A}}) \sum_{j=1}^{m} p(X_j/[x]^i_{\mathcal{A}}) \log\, p(X_j/[x]^i_{\mathcal{A}})$$

and $I(\mathcal{A}; \mathcal{D}) = H(\mathcal{D}) - H(\mathcal{D}/\mathcal{A})$, respectively. Both uncertainty measures have granulation monotonicity, i.e., if $\mathcal{A} \preceq \mathcal{B}$ then $H(\mathcal{D}/\mathcal{A}) \leq H(\mathcal{D}/\mathcal{B})$ and $I(\mathcal{A}; \mathcal{D}) \geq I(\mathcal{B}; \mathcal{D})$, so they are used to construct two types of D-Table reduct, which are equivalent to the classical D-Table reduct based on regions [1,6,11].

(1) $\mathcal{B}$ is a region-based reduct of $\mathcal{C}$, if $\mathrm{POS}_{\mathcal{B}}(\mathcal{D}) = \mathrm{POS}_{\mathcal{C}}(\mathcal{D})$, $\mathrm{POS}_{\mathcal{B}-\{b\}}(\mathcal{D}) \subset \mathrm{POS}_{\mathcal{B}}(\mathcal{D})$.
(2) $\mathcal{B}$ is a conditional entropy-based reduct of $\mathcal{C}$, if $H(\mathcal{D}/\mathcal{B}) = H(\mathcal{D}/\mathcal{C})$, $H(\mathcal{D}/\mathcal{B}-\{b\}) > H(\mathcal{D}/\mathcal{B})$.
(3) $\mathcal{B}$ is a mutual information-based reduct of $\mathcal{C}$, if $I(\mathcal{B}; \mathcal{D}) = I(\mathcal{C}; \mathcal{D})$, $I(\mathcal{B} - \{b\}; \mathcal{D}) < I(\mathcal{B}; \mathcal{D})$.

Moreover, monotonous information entropy $H(\mathcal{A}) = -\sum_{i=1}^{n} p([x]_{\mathcal{A}}^i)\log\, p([x]_{\mathcal{A}}^i)$ is used to equivalently define the information-system reduct [1,3,7].

(1) $\mathcal{B}$ is a knowledge-based reduct of $\mathcal{C}$, if $U/IND(\mathcal{B}) = U/IND(\mathcal{C}), U/IND(\mathcal{B}-\{b\}) \neq U/IND(\mathcal{B})$.
(2) $\mathcal{B}$ is an entropy-based reduct of $\mathcal{C}$, if $H(\mathcal{B}) = H(\mathcal{C})$, $H(\mathcal{B}-\{b\}) < H(\mathcal{B})$.

3 Three-Way Weighted Entropies of a Concept

By evolving the Bayesian probability formula, this section mainly proposes three-way weighted entropies of a concept. For granule $[x]_{\mathcal{A}}$ and concept X, we first analyze the relevant causality mechanism of three-way probabilities, then discuss three-way entropies, and finally construct three-way weighted entropies.

Suppose $p(T) = \frac{|T|}{|U|}$ $(\forall T \in 2^U)$, then $(U, 2^U, p)$ constitutes a probability space. Thus, there are four types of probability to construct the Bayesian formula

$$p([x]_{\mathcal{A}}/X) = \frac{p([x]_{\mathcal{A}}) \cdot p(X/[x]_{\mathcal{A}})}{p(X)}. \tag{1}$$

For given concept X, $p(X) = \frac{|X|}{|U|}$ becomes a constant, so the surplus three-way probabilities are worth analyzing.

From a causality viewpoint, concept X represents a result while divided granule $[x]_{\mathcal{A}}$ means factors. Furthermore, from a Bayesian viewpoint, $\mathcal{A}$ can be viewed as a granulation parameter within a subset range of $\mathcal{C}$.

(1) $p(X/[x]_{\mathcal{A}}) = \frac{|X \cap [x]_{\mathcal{A}}|}{|[x]_{\mathcal{A}}|}$ is the likelihood probability for granulation parameter $\mathcal{A}$ to describe granular decision X.
(2) $p([x]_{\mathcal{A}}/X) = \frac{|X \cap [x]_{\mathcal{A}}|}{|X|}$ is the posterior probability to describe granulation parameters on a premise of result X.
(3) $p([x]_{\mathcal{A}}) = \frac{|[x]_{\mathcal{A}}|}{|U|}$ is the prior probability to describe cause parameter $\mathcal{A}$.

The three-way probabilities, which correspond to relative and absolute measures [18], respectively, exhibit different probability semantics and decision actions. In particular, likelihood $p(X/[x]_{\mathcal{A}})$ and posterior $p([x]_{\mathcal{A}}/X)$ directly reflect causality from the cause-to-effect and effect-to-cause viewpoints, respectively, so their relevant measures can thoroughly describe correlative relationships between the decision concept and its condition structures. Clearly, $p([x]_{\mathcal{A}}/X)$ is more perfect for reduction because reduction is a concrete effect-to-cause pattern, and its calculation is also more optimal. Moreover, prior $p([x]_{\mathcal{A}})$ mainly measures cause uncertainty by reflecting structural information of $\mathcal{A}$.

Our original intention is to describe the causality system regarding $\mathcal{A}$ and X and to further study attribute reduction by constructing benign measures based

on the three-way probabilities. In view of entropy's importance for measuring uncertainty, we next exhibit three-way entropies.

Definition 1. For concept X,

$$H^X(\mathcal{A}) = -\sum_{i=1}^{n} p([x]_{\mathcal{A}}^i)\log\, p([x]_{\mathcal{A}}^i), \tag{2}$$

$$H(\mathcal{A}/X) = -\sum_{i=1}^{n} p([x]_{\mathcal{A}}^i/X)\log\, p([x]_{\mathcal{A}}^i/X), \tag{3}$$

$$H(X/\mathcal{A}) = -\sum_{i=1}^{n} p(X/[x]_{\mathcal{A}}^i)\log\, p(X/[x]_{\mathcal{A}}^i) \tag{4}$$

are called prior, posterior, and likelihood entropies, respectively.

Definition 1 proposes three-way entropies of a concept. In fact, $H^X(\mathcal{A})$ and $H(\mathcal{A}/X)$ naturally measure uncertainty of granulation $\mathcal{A}$ without limitations and on promise X, respectively, because both $p([x]_{\mathcal{A}})$ and $p([x]_{\mathcal{A}}/X)$ form a probability distribution. Moreover, $H(X/\mathcal{A})$ is also formally proposed to measure the likelihood structure, though $\sum_{i=1}^{n} p(X/[x]_{\mathcal{A}}^i) \neq 1$. For X, $H^X(\mathcal{A})$ conducts absolute pre-evaluation, while $H(X/\mathcal{A})$ and $H(\mathcal{A}/X)$ make relative descriptions from two different causality directions. Thus, the three-way entropies, especially $H(X/\mathcal{A})$ and $H(\mathcal{A}/X)$, can measure causality between granulation parameter $\mathcal{A}$ and decision set X.

Proposition 1. If $\mathcal{A} \preceq \mathcal{B}$, then $H^X(\mathcal{A}) \geq H^X(\mathcal{B})$, $H(\mathcal{B}/X) \geq H(\mathcal{A}/X)$, but neither $H(X/\mathcal{A}) \geq H(X/\mathcal{B})$ nor $H(X/\mathcal{A}) \leq H(X/\mathcal{B})$ necessarily holds.

Proof. The results can be proved by entropy properties, because $p([x]_{\mathcal{B}}) = \sum_{t=1}^{k} p([x]_{\mathcal{A}}^t)$ and $p([x]_{\mathcal{B}}/X) = \sum_{t=1}^{k} p([x]_{\mathcal{A}}^t/X)$ but $p(X/[x]_{\mathcal{B}}) \neq \sum_{t=1}^{k} p(X/[x]_{\mathcal{A}}^t)$. □

Granulation monotonicity is an important feature for evaluating an entropy. Based on Proposition 1, the prior/posterior and likelihood entropies have granulation monotonicity and non-monotonicity, respectively. In particular, the following Example 1 illustrates the non-monotonicity of the likelihood entropy.

Example 1. Given $[x]_{\mathcal{A}}^1$, $[x]_{\mathcal{A}}^2$ and complementary X_1, X_2. Let $|[x]_{\mathcal{A}}^1| = 40 = |[x]_{\mathcal{A}}^2|$, $|X_1| = 29$, $|X_2| = 51$; moreover, $|[x]_{\mathcal{A}}^1 \cap X_1| = 1$, $|[x]_{\mathcal{A}}^2 \cap X_1| = 28$, so $|[x]_{\mathcal{A}}^1 \cap X_2| = 39$, $|[x]_{\mathcal{A}}^2 \cap X_2| = 12$. For $[x]_{\mathcal{A}}^1 \cup [x]_{\mathcal{A}}^2 = [x]_{\mathcal{B}}$ regarding X_1, $p(X_1/[x]_{\mathcal{A}}^1) = 0.025$, $p(X_1/[x]_{\mathcal{A}}^2) = 0.7$, $p(X_1/[x]_{\mathcal{B}}) = 0.3625$, so $-0.025\log 0.025 - 0.7\log 0.7 = 0.4932 < 0.5307 = -0.3625\log 0.3625$; regarding X_2, $p(X_2/[x]_{\mathcal{A}}^1) = 0.975$, $p(X_2/[x]_{\mathcal{A}}^2) = 0.3$, $p(X_2/[x]_{\mathcal{B}}) = 0.6375$, so $-0.975\log 0.975 - 0.3\log 0.3 = 0.5567 > 0.4141 = -0.6375\log 0.6375$. If $U/IND(\mathcal{A}) = \{[x]_{\mathcal{A}}^1, [x]_{\mathcal{A}}^2\}$ and $U/IND(\mathcal{B}) = \{[x]_{\mathcal{B}}\}$, then $\mathcal{A} \preceq \mathcal{B}$ but $H(X_1/\mathcal{A}) \leq H(X_1/\mathcal{B})$, $H(X_2/\mathcal{A}) \geq H(X_2/\mathcal{B})$. □

For the three-way probabilities, $p([x]_{\mathcal{A}}/X)$ and $p(X/[x]_{\mathcal{A}})$ reflect causality between $\mathcal{A}$ and X; for the three-way entropies, only $H^X(\mathcal{A})$ and $H(\mathcal{A}/X)$ exhibit

necessary monotonicity. Thus, $p([x]_{\mathcal{A}}/X)$ and further $H(\mathcal{A}/X)$ hold important significance for describing $\mathcal{A}$ structure based on X. In fact, posterior entropy $H(\mathcal{A}/X)$ reflects the average information content of granulation $U/IND(\mathcal{A})$ for given concept X, and at the entropy level, only it has perfect value for measuring uncertainty of $\mathcal{A}$ for X. This posterior entropy's function underlies latter importance of the posterior weighed entropy and posterior attribute reduction.

Though the posterior entropy is valuable, however, there are no relationships for the three-way entropies. Thus, better three-way measures with granulation monotonicity are worth deeply mining to establish an essential connection. For this purpose, we first creatively evolve the Bayesian probability formula to naturally mine three-way weighted entropies, and we then explore their monotonicity and relationship.

Theorem 1. $-\sum_{i=1}^{n} p(X)p([x]_{\mathcal{A}}^{i}/X)\log\, p([x]_{\mathcal{A}}^{i}/X)$
$= \quad -\sum_{i=1}^{n} p(X/[x]_{\mathcal{A}}^{i})p([x]_{\mathcal{A}}^{i})\log\quad p([x]_{\mathcal{A}}^{i}) \;-\; \sum_{i=1}^{n} p([x]_{\mathcal{A}}^{i})p(X/[x]_{\mathcal{A}}^{i})\log\quad p(X/[x]_{\mathcal{A}}^{i})$
$+\, p(X)\log\, p(X)$.

Proof. First, $p([x]_{\mathcal{A}}^{i}/X) = \frac{p([x]_{\mathcal{A}}^{i})\cdot p(X/[x]_{\mathcal{A}}^{i})}{p(X)}$, $\forall i \in \{1, ..., n\}$. Thus, $-p([x]_{\mathcal{A}}^{i}/X)\log\, p([x]_{\mathcal{A}}^{i}/X) = -\frac{p([x]_{\mathcal{A}}^{i})\cdot p(X/[x]_{\mathcal{A}}^{i})}{p(X)}[\log\, p([x]_{\mathcal{A}}^{i}) + \log\, p(X/[x]_{\mathcal{A}}^{i}) - \log\, p(X)]$. Hence, $-p(X)p([x]_{\mathcal{A}}^{i}/X)\log\, p([x]_{\mathcal{A}}^{i}/X) = -p(X/[x]_{\mathcal{A}}^{i})p([x]_{\mathcal{A}}^{i})\log\, p([x]_{\mathcal{A}}^{i}) - p([x]_{\mathcal{A}}^{i})p(X/[x]_{\mathcal{A}}^{i})\log\, p(X/[x]_{\mathcal{A}}^{i}) + p([x]_{\mathcal{A}}^{i})p(X/[x]_{\mathcal{A}}^{i})\log\, p(X)$. Furthermore, the result is obtained by summation, where $\sum_{i=1}^{n} p([x]_{\mathcal{A}}^{i})p(X/[x]_{\mathcal{A}}^{i})\log\, p(X) = \sum_{i=1}^{n} p([x]_{\mathcal{A}}^{i} \cap X)\log\, p(X) = [\sum_{i=1}^{n} p([x]_{\mathcal{A}}^{i} \cap X)]\log\, p(X) = p(X)\log\, p(X)$. □

Theorem 1 develops the Bayesian theorem in an entropy direction, and there is actually a core form containing both an entropy and weights, i.e., a weighted entropy. Thus, a weighted entropy plays an core role and can establish an equation. This entropy evolution inherits the Bayesian probability formula and inspires our following further works.

Definition 2. For probability distribution (ξ, p_i) and weight $w_i \geq 0$, $H_W(\xi) = -\sum_{i=1}^{n} w_i p_i \log\, p_i$ is called the weighted entropy. In particular, the generalized weighted entropy has not constraint condition $\sum_{i=1}^{n} p_i = 1$.

The weighted entropy mainly introduces weights into the entropy, and weights usually reflect importance degrees for information receivers. In particular, it develops the entropy and degenerates into the latter by setting up $w_i = 1$. Herein, the generalized weighted entropy is mainly used in view of $\sum_{i=1}^{n} p(X/[x]_{\mathcal{A}}^{i}) \neq 1$.

Definition 3. For concept X,

$$H_W^X(\mathcal{A}) = -\sum_{i=1}^{n} p(X/[x]_{\mathcal{A}}^i)p([x]_{\mathcal{A}}^i)\log\, p([x]_{\mathcal{A}}^i), \tag{5}$$

$$H_W(\mathcal{A}/X) = -\sum_{i=1}^{n} p(X)p([x]_{\mathcal{A}}^i/X)\log\, p([x]_{\mathcal{A}}^i/X) = p(X)H(\mathcal{A}/X), \tag{6}$$

$$H_W(X/\mathcal{A}) = -\sum_{i=1}^{n} p([x]_{\mathcal{A}}^i)p(X/[x]_{\mathcal{A}}^i)\log\, p(X/[x]_{\mathcal{A}}^i) \tag{7}$$

are called prior, posterior, and likelihood weighted entropies, respectively.

The three-way weighted entropies originate from three-way entropies by adding probability-based weight coefficients. In fact, $H_W^X(\mathcal{A})$ improves upon absolute $H^X(\mathcal{A})$ by introducing relative $p(X/[x]_{\mathcal{A}}^i)$, while $H_W(\mathcal{A}/X)$ and $H_W(X/\mathcal{A})$ improve upon relative $H(\mathcal{A}/X)$ and $H(X/\mathcal{A})$ by introducing absolute $p(X)$ and $p([x]_{\mathcal{A}}^i)$, respectively. Thus, $H_W^X(\mathcal{A})$, $H_W(\mathcal{A}/X)$, $H_W(X/\mathcal{A})$ inherit uncertainty semantics by different probability weights, and they exhibit systematic completeness and superior stability from the double-quantitative perspective [18], so they can better describe the system regarding cause $\mathcal{A}$ and result X. Moreover, posterior weighted entropy $H_W(\mathcal{A}/X)$ has a simple and perfect structure, because it can be directly decomposed into a product of posterior entropy $H(\mathcal{A}/X)$ and constant $p(X)$. Note that weighted entropy symbol $H_W(.)$ is distinguished from information entropy symbol $H(.)$.

Proposition 2. If $\mathcal{A} \preceq \mathcal{B}$, then $H_W^X(\mathcal{A}) \geq H_W^X(\mathcal{B})$, $H_W(\mathcal{A}/X) \geq H_W(\mathcal{B}/X)$, $H_W(X/\mathcal{A}) \leq H_W(X/\mathcal{B})$.

Proof. Herein, we only provide the proof for the likelihood weighted entropy by utilizing granular merging $\bigcup_{t=1}^{k} [x]_{\mathcal{A}}^t = [x]_{\mathcal{B}}$. $f(u) = -u\log u$ $(u \in [0.1])$ is a concave function; thus, if $\sum_{t=1}^{k} \lambda_t = 1$, then $-\sum_{t=1}^{k} \lambda_t p_t \log\, p_t \leq -[\sum_{t=1}^{k} \lambda_t p_t]\log[\sum_{t=1}^{k} \lambda_t p_t]$.

$$-\sum_{t=1}^{k} p([x]_{\mathcal{A}}^t)p(X/[x]_{\mathcal{A}}^t)\log\, p(X/[x]_{\mathcal{A}}^t) = -\sum_{t=1}^{k} p([x]_{\mathcal{B}})\frac{|[x]_{\mathcal{A}}^t|}{|[x]_{\mathcal{B}}|}p(X/[x]_{\mathcal{A}}^t)\log\, p(X/[x]_{\mathcal{A}}^t)$$

$$= p([x]_{\mathcal{B}})[-\sum_{t=1}^{k} \frac{|[x]_{\mathcal{A}}^t|}{|[x]_{\mathcal{B}}|}p(X/[x]_{\mathcal{A}}^t)\log\, p(X/[x]_{\mathcal{A}}^t)]$$

$$\leq -p([x]_{\mathcal{B}})[\sum_{t=1}^{k} \frac{|[x]_{\mathcal{A}}^t|}{|[x]_{\mathcal{B}}|}p(X/[x]_{\mathcal{A}}^t)]\log\, \frac{\sum_{t=1}^{k} |[x]_{\mathcal{A}}^t \cap X|}{|[x]_{\mathcal{B}}|}$$

$$= -p([x]_{\mathcal{B}})\frac{|[x]_{\mathcal{B}} \cap X|}{|[x]_{\mathcal{B}}|}\log\, \frac{|[x]_{\mathcal{B}} \cap X|}{|[x]_{\mathcal{B}}|} = -p([x]_{\mathcal{B}})p(X/[x]_{\mathcal{B}})\log\, p(X/[x]_{\mathcal{B}}). \quad \square$$

Based on Proposition 2, three weighted entropies exhibit perfect granulation monotonicity and thus hold significance. In particular, $H_W(X/\mathcal{A})$ becomes

monotonicity though $H(X/\mathcal{A})$ is non-monotonicity, and this monotonicity difficulty is proved by utilizing a concave feature of function $-u\log u$.

Theorem 2. $H_W(\mathcal{A}/X) = H_W^X(\mathcal{A}) + H_W(X/\mathcal{A}) + p(X)\log\, p(X)$
$= H_W^X(\mathcal{A}) - [-p(X)\log\, p(X) - H_W(X/\mathcal{A})]$, and $-p(X)\log\, p(X) - H_W(X/\mathcal{A}) \geq 0$.

Theorem 2 provides an important relationship for the three-way weighted entropies (where $-p(X)\log\, p(X)$ is a constant), i.e., the posterior weighted entropy becomes a linear translation of the sum of the prior and likelihood weighted entropies. Thus, the Bayesian probability formula can deduce essential relationships regarding not three-way entropies but three-way weighted entropies. Furthermore, $-p(X)\log\, p(X) - H_W(X/\mathcal{A})$ can be chosen as a new measure to simplify the fundamental equation by eliminating the translation distance.

Definition 4. $H_W^*(X/\mathcal{A}) = -p(X)\log\, p(X) - H_W(X/\mathcal{A})$.

Corollary 1. (1) If $\mathcal{A} \preceq \mathcal{B}$, then $H_W^*(X/\mathcal{A}) \geq H_W^*(X/\mathcal{B})$.
(2) $H_W(\mathcal{A}/X) = H_W^X(\mathcal{A}) - H_W^*(X/\mathcal{A})$.

Herein, $H_W^*(X/\mathcal{A})$ corresponds to $H_W(X/\mathcal{A})$ by a negative linear transformation, so it exhibits opposite granulation monotonicity. Furthermore, the posterior weighted entropy becomes the difference between prior weighted entropy $H_W^X(\mathcal{A})$ and $H_W^*(X/\mathcal{A})$, and the latter corresponds to the likelihood weighted entropy.

4 Three-Way Weighted Entropies of a Classification

Three-way weighted entropies are proposed for a concept in Section 3, and they will be further constructed for a classification in this section by a natural integration strategy of GrC. Moreover, they will be linked to the existing RS-Theory system with the information entropy, conditional entropy and mutual information. Next, classification $U/IND(\mathcal{D}) = \{X_1, .., X_m\}$ with m concepts is given.

Definition 5. For classification $U/IND(\mathcal{D})$,

$$H_W^{\mathcal{D}}(\mathcal{A}) = \sum_{j=1}^{m} H_W^{X_j}(\mathcal{A}),\ H_W(\mathcal{A}/\mathcal{D}) = \sum_{j=1}^{m} H_W(A/X_j),\ H_W(\mathcal{D}/\mathcal{A}) = \sum_{j=1}^{m} H_W(X_j/A)$$

are called prior, posterior, and likelihood weighted entropies, respectively. Moreover, let $H_W^*(\mathcal{D}/\mathcal{A}) = \sum_{j=1}^{m} H_W^*(X_j/A)$.

For decision classification $U/IND(\mathcal{D})$, the three-way weighted entropies are corresponding sum of concepts' weighted entropies regarding classification's internal concepts, because we naturally adopt a GrC strategy from an internal concept to its integrated classification. Thus, they inherit relevant causality mechanisms and hold corresponding functions for measuring uncertainty; moreover, they also inherit the essential monotonicity and mutual relationship.

Proposition 3. If $\mathcal{A} \preceq \mathcal{B}$, then $H_W^{\mathcal{D}}(\mathcal{A}) \geq H_W^{\mathcal{D}}(\mathcal{B})$, $H_W(\mathcal{A}/\mathcal{D}) \geq H_W(\mathcal{B}/\mathcal{D})$, $H_W(\mathcal{D}/\mathcal{A}) \leq H_W(\mathcal{D}/\mathcal{B})$, $H_W^*(\mathcal{D}/\mathcal{A}) \geq H_W^*(\mathcal{D}/\mathcal{B})$.

Theorem 3 (Weighted Entropies' Bayesian Formula).

$$H_W(\mathcal{A}/\mathcal{D}) = H_W^{\mathcal{D}}(\mathcal{A}) - [H(\mathcal{D}) - H_W(\mathcal{D}/\mathcal{A})] = H_W^{\mathcal{D}}(\mathcal{A}) - H_W^*(\mathcal{D}/\mathcal{A}).$$

Theorem 3 describes an important relationship of the three-way weighted entropies by introducing $H(\mathcal{D})$. Thus, the posterior weighted entropy is difference between the prior weighted entropy and $H^*_W(\mathcal{D}/\mathcal{A})$, and the latter is a linear transformation of the likelihood weighted entropy. In particular, Theorem 3 essentially evolves the Bayesian probability formula, so it is called by *Weighted Entropies' Bayesian Formula* to highlight its important values.

Next, we summarize the above GrC works via Table 1. There are three GrC levels which are located at the micro bottom, meso layer, and macro top, respectively.

(1) At Level (1), the three-way probabilities describe granule $[x]_{\mathcal{A}}$ and concept X, and the Bayesian probability formula holds by using premise $p(X)$.
(2) At Level (2), the three-way weighted entropies describe granulation $\mathcal{A}$ and concept X and exhibit granulation monotonicity. In particular, $H_W(\mathcal{A}/X) = H^X_W(\mathcal{A}) - H^*_W(X/\mathcal{A})$ acts as an evolutive Bayesian formula, where $H^*_W(X/\mathcal{A})$ is a linear adjustment of $H_W(X/\mathcal{A})$ by using premise $-p(X)\log p(X)$.
(3) At Level (3), the three-way weighted entropies describe granulation $\mathcal{A}$ and classification $\mathcal{D}$ and inherit granulation monotonicity. In particular, $H_W(\mathcal{A}/\mathcal{D}) = H^{\mathcal{D}}_W(\mathcal{A}) - H^*_W(\mathcal{D}/\mathcal{A})$ acts as an evolutive Bayesian result, where $H^*_W(\mathcal{D}/\mathcal{A})$ is a linear adjustment of $H_W(\mathcal{D}/\mathcal{A})$ by using premise $H(\mathcal{D})$.

Thus, our GrC works establish an integrated description for $\mathcal{A}$ and $\mathcal{D}$ by using a bottom-top strategy, so they underlie the further discussion, especially for attribute reduction. In fact, attribute reduction is mainly located at Level (3), where $H(\mathcal{D})$ is a constant from the causality perspective.

Table 1. GrC-Based Weighted Entropies and Relevant Bayesian Formulas

Level	Objects	Three-Way Measures	Relevant Bayesian Formulas
(1)	$[x]_{\mathcal{A}}$, X	$p([x]_{\mathcal{A}})$, $p([x]_{\mathcal{A}}/X)$, $p(X/[x]_{\mathcal{A}})$	$p([x]_{\mathcal{A}}/X) = \frac{p([x]_{\mathcal{A}}) \cdot p(X/[x]_{\mathcal{A}})}{p(X)}$
(2)	$\mathcal{A}$, X	$H^X_W(\mathcal{A})$, $H_W(\mathcal{A}/X)$, $H_W(X/\mathcal{A})$ (or $H^*_W(X/\mathcal{A})$)	$H_W(\mathcal{A}/X) = H^X_W(\mathcal{A}) - H^*_W(X/\mathcal{A})$
(3)	$\mathcal{A}$, $\mathcal{D}$	$H^{\mathcal{D}}_W(\mathcal{A})$, $H_W(\mathcal{A}/\mathcal{D})$, $H_W(\mathcal{D}/\mathcal{A})$ (or $H^*_W(\mathcal{D}/\mathcal{A})$)	$H_W(\mathcal{A}/\mathcal{D}) = H^{\mathcal{D}}_W(\mathcal{A}) - H^*_W(\mathcal{D}/\mathcal{A})$

Finally, we explain the novel system of the three-way weighted entropies by the previous system based on information theory, and we also analyze both systems' relationships.

Theorem 4. $H^{\mathcal{D}}_W(\mathcal{A}) = H(\mathcal{A})$, $H_W(\mathcal{A}/\mathcal{D}) = H(\mathcal{A}/\mathcal{D})$, $H_W(\mathcal{D}/\mathcal{A}) = H(\mathcal{D}/\mathcal{A})$, $H^*_W(\mathcal{D}/\mathcal{A}) = I(\mathcal{A};\mathcal{D})$.

Proof. (1) $H^{\mathcal{D}}_W(\mathcal{A}) = \sum_{j=1}^{m} H^{X_j}_W(\mathcal{A}) = -\sum_{j=1}^{m}[\sum_{i=1}^{n} p(X_j/[x]^i_{\mathcal{A}})p([x]^i_{\mathcal{A}})\log p([x]^i_{\mathcal{A}})]$

$= -\sum_{j=1}^{m}[p(X_j/[x]^1_{\mathcal{A}})p([x]^1_{\mathcal{A}})\log p([x]^1_{\mathcal{A}}) - ... - p(X_j/[x]^n_{\mathcal{A}})p([x]^n_{\mathcal{A}})\log p([x]^n_{\mathcal{A}})]$

$= -[\sum_{j=1}^{m} p(X_j/[x]^1_{\mathcal{A}})]p([x]^1_{\mathcal{A}})\log p([x]^1_{\mathcal{A}}) - ... - [\sum_{j=1}^{m} p(X_j/[x]^n_{\mathcal{A}})]p([x]^n_{\mathcal{A}})\log p([x]^n_{\mathcal{A}})$

$= -p([x]^1_{\mathcal{A}})\log p([x]^1_{\mathcal{A}}) - ... - p([x]^n_{\mathcal{A}})\log p([x]^n_{\mathcal{A}}) = H(\mathcal{A})$.

(2) $H_W(\mathcal{A}/\mathcal{D}) = \sum_{j=1}^{m} H_W(A/X_j) = \sum_{j=1}^{m} p(X_j)H(A/X_j)$

$= -\sum_{j=1}^{m} p(X_j) \sum_{i=1}^{n} p([x]_{\mathcal{A}}^i/X_j)\log\, p([x]_{\mathcal{A}}^i/X_j) = H(\mathcal{A}/\mathcal{D})$.

(3) $H_W(\mathcal{D}/\mathcal{A}) = H_W(X_1/A) + ... + H_W(X_m/A)$

$= [-p([x]_{\mathcal{A}}^1)p(X_1/[x]_{\mathcal{A}}^1)\log\, p(X_1/[x]_{\mathcal{A}}^1) - ... - p([x]_{\mathcal{A}}^n)p(X_1/[x]_{\mathcal{A}}^1)\log\, p(X_1/[x]_{\mathcal{A}}^n)] +$
...
$+[-p([x]_{\mathcal{A}}^1)p(X_m/[x]_{\mathcal{A}}^1)\log\, p(X_m/[x]_{\mathcal{A}}^1) - ... - p([x]_{\mathcal{A}}^n)p(X_m/[x]_{\mathcal{A}}^1)\log\, p(X_m/[x]_{\mathcal{A}}^n)]$
$= -p([x]_{\mathcal{A}}^1)[p(X_1/[x]_{\mathcal{A}}^1)\log\, p(X_1/[x]_{\mathcal{A}}^1) + ... + p(X_m/[x]_{\mathcal{A}}^1)\log\, p(X_m/[x]_{\mathcal{A}}^1)] - ...$
$- p([x]_{\mathcal{A}}^n)[p(X_1/[x]_{\mathcal{A}}^1)\log\, p(X_1/[x]_{\mathcal{A}}^n) + ... + p(X_m/[x]_{\mathcal{A}}^n)\log\, p(X_m/[x]_{\mathcal{A}}^1)]$
$= -\sum_{i=1}^{n} p([x]_{\mathcal{A}}^i) \sum_{j=1}^{m} p(X_j/[x]_{\mathcal{A}}^i)\log\, p(X_j/[x]_{\mathcal{A}}^i) = H(\mathcal{D}/\mathcal{A})$.

Thus, $H_W^*(\mathcal{D}/\mathcal{A}) = H(\mathcal{D}) - H_W(\mathcal{D}/\mathcal{A}) = H(\mathcal{D}) - H(\mathcal{D}/\mathcal{A}) = I(\mathcal{A};\mathcal{D})$ □.

Theorem 5. $H_W(\mathcal{A}/\mathcal{D}) = H_W^{\mathcal{D}}(\mathcal{A}) - H_W^*(\mathcal{D}/\mathcal{A})$ is equivalent to $H(\mathcal{A}/\mathcal{D}) = H(\mathcal{A}) - I(\mathcal{A};\mathcal{D})$.

Based on Theorem 4, three-way weighted entropies $H_W^{\mathcal{D}}(\mathcal{A})$, $H_W(\mathcal{A}/\mathcal{D})$, $H_W(\mathcal{D}/\mathcal{A})$ are equivalent to prior entropy $H(\mathcal{A})$, conditional entropy $H(\mathcal{A}/\mathcal{D})$, conditional entropy $H(\mathcal{D}/\mathcal{A})$, respectively; moreover, $H_W^*(\mathcal{D}/\mathcal{A})$ corresponds to mutual information $I(\mathcal{A};\mathcal{D})$. Furthermore, Theorem 5 reflects equivalence between $H_W(\mathcal{A}/\mathcal{D}) = H_W^{\mathcal{D}}(\mathcal{A}) - H_W^*(\mathcal{D}/\mathcal{A})$ and $H(\mathcal{A}/\mathcal{D}) = H(\mathcal{A}) - I(\mathcal{A};\mathcal{D})$, which are from two different systems. Thus, the weighted entropy system (including its Bayesian formula) has been explained/verified by the previous information theory system. In contrast, the former can thoroughly explain the latter as well. Therefore, both systems exhibit theoretical equivalence. However, the weighted entropy approach conducts a GrC construction, and it also emphasizes the causality semantics and application direction based on the Bayesian mechanism. Thus, the three-way weighted entropies hold at least two fundamental values. First, they construct, explain, and deepen the existing information system of RS-Theory by the GrC construction and Bayesian formula; moreover, they underlie systemic attribute reduction by the essential uncertainty measure and effective Bayesian inference.

5 Three-Way Attribute Reduction

The three-way weighted entropies (of a classification) and their monotonicity and relationship have been provided in Section 4. This section mainly uses them to systemically construct three-way attribute reduction, and the poster reduction will be emphasized via the Bayesian inference and causality theory.

Definition 6. $\mathcal{B}$ is called likelihood, prior, and posterior reducts of $\mathcal{C}$, if it satisfies the following three conditions, respectively.

(1) $H_W^*(\mathcal{D}/\mathcal{B}) = H_W^*(\mathcal{D}/\mathcal{C})$, $H_W^*(\mathcal{D}/\mathcal{B} - \{b\}) < H_W^*(\mathcal{D}/\mathcal{B})$
(or $H_W(\mathcal{D}/\mathcal{B}) = H_W(\mathcal{D}/\mathcal{C})$, $H_W(\mathcal{D}/\mathcal{B} - \{b\}) > H_W(\mathcal{D}/\mathcal{B})$).
(2) $H_W^D(\mathcal{B}) = H_W^D(\mathcal{C})$, $H_W^D(B - \{b\}) < H_W^D(\mathcal{B})$.
(3) $H_W(\mathcal{B}/\mathcal{D}) = H_W(\mathcal{C}/\mathcal{D})$, $H_W(\mathcal{B} - \{b\}/\mathcal{D}) < H_W(\mathcal{B}/\mathcal{D})$.

Theorem 6.

(1) A likelihood reduct is equivalent to a D-Table reduct. Furthermore, a likelihood reduct based on $H^*_W(\mathcal{D}/\mathcal{A})$ or $H_W(\mathcal{D}/\mathcal{A})$ is equivalent to a D-Table reduct based on the mutual information or conditional entropy, respectively.
(2) A prior reduct is equivalent to an information-system reduct.
(3) A posterior reduct is different from both D-Table and information-system reducts.

Proof. (1) For the D-Table reduct, the region-based method is equivalent to the mutual information-based and conditional entropy-based ways [1,6,11]; furthermore, $I(\mathcal{A};\mathcal{D})$ and $H(\mathcal{D}/\mathcal{A})$ correspond to $H^*_W(\mathcal{D}/\mathcal{A})$ and $H_W(\mathcal{D}/\mathcal{A})$, respectively, so the likelihood reduct is equivalent to the two information-based reducts and the classical region-based reduct. (2) For the information-system reduct, the prior reduct is equivalent to the entropy-based reduct and further knowledge-based reduct, because $H^{\mathcal{D}}_W(\mathcal{A}) = H(\mathcal{A})$. (3) The difference of the posterior reduct is verified by the following D-Table example. □.

Example 2. In D-Table $S = (U, \mathcal{C} \cup \mathcal{D})$ provided by Table 2,
$U = \{x_1, ..., x_{12}\}$, $C = \{a, b, c\}$, $D = \{d\}$, $U/\{d\} = \{X_1, X_2, X_3\}$,
$X_1 = \{x_1, ..., x_4\}$, $X_2 = \{x_5, ..., x_8\}$, $X_3 = \{x_9, ..., x_{12}\}$. Thus, $U/\{a, b, c\} =$
$U/\{a, b\} = \{\{x_2, x_3, x_6, x_{11}\}, \{x_4, x_8, x_{12}\}, \{x_1\}, \{x_5\}, \{x_7, x_{10}\}, \{x_9\}\}$,
$U/\{a\} = U/\{a, c\} = \{\{x_2, x_3, x_6, x_7, x_{10}, x_{11}\}, \{x_4, x_8, x_{12}\}, \{x_1\}, \{x_5\}, \{x_9\}\}$,
$U/\{b\} = \{\{x_2, x_3, x_6, x_{11}\}, \{x_4, x_8, x_{12}\}, \{x_1, x_7, x_{10}\}, \{x_5, x_9\}\}$,
$U/\{c\} = \{\{x_2, x_3, x_6, x_7, x_{10}, x_{11}\}, \{x_4, x_8, x_{12}\}, \{x_1, x_5, x_9\}\}$,
$U/\{b, c\} = \{\{x_2, x_3, x_6, x_{11}\}, \{x_4, x_8, x_{12}\}, \{x_1\}, \{x_5, x_9\}, \{x_7, x_{10}\}\}$.

Table 2. D-Table in Example 2

U	a	b	c	d	U	a	b	c	d	U	a	b	c	d
x_1	3	3	3	1	x_5	4	4	3	2	x_9	5	4	3	3
x_2	1	1	1	1	x_6	1	1	1	2	x_{10}	1	3	1	3
x_3	1	1	1	1	x_7	1	3	1	2	x_{11}	1	1	1	3
x_4	2	2	2	1	x_8	2	2	2	2	x_{12}	2	2	2	3

First, there are only two D-Table reducts $\{a\}$, $\{c\}$ and one information-system reduct $\{a, b\}$. Herein, $H_W(\mathcal{C}/\mathcal{D}) = P(X_1)H(\mathcal{C}/X_1) + P(X_2)H(\mathcal{C}/X_2) + P(X_3)$ $H(\mathcal{C}/X_3) = \frac{4}{12}[(-0.5\log0.5 - 2\times0.25\log0.25) - 4\times0.25\log0.25 - 4\times0.25\log0.25] =$ $1.8333 = H_W(\{b\}/\mathcal{D})$, $H_W(\{a\}/\mathcal{D}) = H_W(\{a, c\}/\mathcal{D}) = H_W(\{c\}/\mathcal{D}) = \frac{4}{12}$ $(-0.5\log0.5 - 2 \times 0.25\log0.25) \times 3 = 1.500 < 1.8333$. Thus, $\{b\}$ becomes the sole posterior reduct and is neither the $(U, \mathcal{C} \cup \mathcal{D})$ reduct nor $(U, \mathcal{C})$ reduct; in contrast, neither $\{a\}$, $\{c\}$ nor $\{a, b\}$ is a posterior reduct. Note that the key granular merging regarding $\{x_1, x_7, x_{10}\}$ is allowed not for the other reducts but for the posterior reduct. □

The three weighted entropies can measure uncertainty, and they are used to naturally define the three-way reducts. In spite of measuring of all weighted entropies, the three-way reducts exhibit different reduction essence. The likelihood reduct and prior reduct, which are also related to the mutual information, conditional entropy and information entropy, mainly correspond to the qualitative reducts regarding D-Table and information-system, respectively. In contrast, the posterior reduct completely corresponds to a quantitative reduct. Thus, the posterior weighted entropy exhibit more essential metrizability, and the posterior reduct exhibits novelty and transcendence, so both are worth emphasizing.

Next, based on the posterior weighted entropy, we analyze important significance of the posterior reduct for D-Table reduct.

(1) The D-Table reduct usually uses likelihood information in the cause-to-effect (or condition-to-decision) direction. According to the Bayesian inference, the posterior weighted entropy adjusts the likelihood weighted entropy by strengthening the prior knowledge, so the posterior reduct improves upon the likelihood reduct by pursuing quantitative uncertainty rather than qualitative absoluteness. In fact, by considering the granulation distribution, the posterior reduct achieves the highest posterior uncertainty and lowest risk according to the granulation monotonicity and maximum entropy principle, respectively, so it can avoid the over-fitting problem due to its measurability, generality, and robustness.
(2) In D-Table $(U, \mathcal{C} \cup \mathcal{D})$, granulation $U/IND(\mathcal{A})$ and classification $\mathcal{D}$ correspond to the condition cause and decision effect, respectively. D-Table reduction aims to choose appropriate granulation parameters $\mathcal{A}$ to preserve specific decision information regarding $\mathcal{D}$, i.e., it mainly seeks condition parameters $\mathcal{A}$ on a stable premise of $\mathcal{D}$. Thus, from the causality viewpoint, posterior weighted entropy $H_W(\mathcal{A}/\mathcal{D})$ not only reflects the causality relationship between C and $\mathcal{D}$ but also more adheres to the operational pattern of D-Table reduction, so the posterior reduct holds practical significance by adopting the cause-to-effect (or decision-to-condition) strategy.

In summary, within a new framework of Bayesian inference, the posterior weighted entropy bears important information of uncertainty distribution, and it also positively improves upon the likelihood weighted entropy by considering the prior information. Moreover, $H_W(\mathcal{A}/\mathcal{D})$ (i.e., condition entropy $H(\mathcal{A}/\mathcal{D})$) is simpler than $H^*_W(\mathcal{D}/\mathcal{A})$ (i.e., mutual information $I(\mathcal{A};\mathcal{D})$) and $H_W(\mathcal{D}/\mathcal{A})$ (i.e., condition entropy $H(\mathcal{D}/\mathcal{A})$). Thus, the posterior reduct holds advantages regarding uncertainty semantics, causality directness and calculation optimization.

6 Conclusions

Based on the GrC technology and Bayesian inference approach, we construct three-way weighted entropies and three-way attribute reduction, and the relevant results deepen information theory-based RS-Theory, especially the GrC uncertainty measurement and attribute reduction. The three-way weighted entropies and three-way attribute reduction actually correspond to the likelihood,

prior, posterior decisions, so they also enrich the three-way decision theory from a new viewpoint. In particular, hierarchies of three-way attribute reduction are worth deeply exploring, and the posterior weighted entropy and posterior attribute reduction need in-depth theoretical exploration and further practical verification.

Acknowledgments. This work was supported by the National Science Foundation of China (61273304 and 61203285), Specialized Research Fund for Doctoral Program of Higher Education of China (20130072130004), China Postdoctoral Science Foundation Funded Project (2013T60464 and 2012M520930), and Shanghai Postdoctoral Scientific Program (13R21416300).

References

1. Pawlak, Z.: Rough sets. International Journal of Information and Computer Science 11(5), 341–356 (1982)
2. Shannon, C.E.: The mathematical theory of communication. The Bell System Technical Journal 27(3–4), 373–423 (1948)
3. Miao, D.Q.: Rough set theory and its application in machine learning (Ph. D. Thesis). Beijing, Institute of Automation, The Chinese Academy of Sciences (1997)
4. Beaubouef, T., Petry, F.E., Arora, G.: Information-theoretic measures of uncertainty for rough sets and rough relational databases. Information Sciences 109(1-4), 185–195 (1998)
5. Slezak, D.: Approximate entropy reducts. Fundamenta Informaticae 53, 365–390 (2002)
6. Miao, D.Q., Hu, G.R.: A heuristic algorithm for reduction of knowledge. Chinese Journal of Computer Research & Development 36, 681–684 (1999)
7. Miao, D.Q., Wang, J.: An information representation of the concepts and operations in rough set theory. Journal of Software 10(2), 113–116 (1999)
8. Liang, J.Y., Qian, Y.H., Chu, D.Y., Li, D.Y., Wang, J.H.: The algorithm on knowledge reduction in incomplete information systems. Fuzziness and Knowledge-Based Systems 10, 95–103 (2002)
9. Liang, J., Shi, Z., Li, D., Wierman, M.: Information entropy, rough entropy and knowledge granularity in incomplete information systems. International Journal of General Systems 35(6), 641–654 (2006)
10. Wang, G.Y., Yu, H., Yang, D.C.: Decision table reduction based on conditional information entropy. Chinese Journal of Computers 25, 759–766 (2002)
11. Wang, G.Y., Zhao, J., An, J.J., Wu, Y.: A comparative study of algebra viewpoint and information viewpoint in attribute reduction. Fundamenta Informaticae 68(3), 289–301 (2005)
12. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer (2006)
13. Zadeh, L.A.: Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. Fuzzy Sets and Systems 90, 111–127 (1997)
14. Lin, T.Y.: Granular computing: from rough sets and neighborhood systems to information granulation and computing with words. In: European Congress on Intelligent Techniques and Soft Computing, pp. 1602–1606 (1997)
15. Yao, Y.: An outline of a theory of three-way decisions. In: Yao, J., Yang, Y., Słowiński, R., Greco, S., Li, H., Mitra, S., Polkowski, L. (eds.) RSCTC 2012. LNCS, vol. 7413, pp. 1–17. Springer, Heidelberg (2012)

16. Yao, Y.Y.: Three-way decisions with probabilistic rough sets. Information Sciences 180, 341–353 (2010)
17. Miao, D.Q., Zhang, X.Y.: Change uncertainty of three-way regions in knowledge-granulation. In: Liu, D., Li, T.R., Miao, D.Q., Wang, G.Y., Liang, J.Y. (eds.) Three-Way Decisions and Granular Computing, pp. 116–144. Science Press, Beijing (2013)
18. Zhang, X.Y., Miao, D.Q.: Quantitative information architecture, granular computing and rough set models in the double-quantitative approximation space on precision and grade. Information Sciences 268, 147–168 (2014)