Information Sciences 268 (2014) 147–168

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Quantitative information architecture, granular computing and rough set models in the double-quantitative approximation space of precision and grade

Xianyong Zhang [a,b,c,*], Duoqian Miao [a,b]

[a] Department of Computer Science and Technology, Tongji University, Shanghai 201804, PR China
[b] The Key Laboratory of Embedded System and Service Computing, Ministry of Education, Shanghai 201804, PR China
[c] College of Mathematics and Software Science, Sichuan Normal University, Chengdu 610068, PR China

ARTICLE INFO

Article history:
Received 14 May 2012
Received in revised form 3 September 2013
Accepted 7 September 2013
Available online 16 September 2013

Keywords:
Rough set theory
Granular computing
Double quantification
Quantitative information architecture
Variable precision rough set
Graded rough set

ABSTRACT

Because precision and grade act as fundamental quantitative information in approximation space, they are used in relative and absolute quantifications, respectively. At present, the double quantification regarding precision and grade is a novel and valuable subject, but quantitative information fusion has become a key problem. Thus, this paper constructs the double-quantitative approximation space of precision and grade (PG-Approx-Space) and tackles the fusion problem using normal logical operations. It further conducts double-quantification studies on granular computing and rough set models. (1) First, for quantitative information organization and storage, we construct space and plane forms of PG-Approx-Space using the Cartesian product, and for quantitative information extraction and fusion, we establish semantics construction and semantics granules of PG-Approx-Space. (2) Second, by granular computing, we investigate three primary granular issues: quantitative semantics, complete system and optimal calculation. Accordingly, six types of fundamental granules are proposed based on the semantic, microscopic and macroscopic descriptions; their semantics, forms, structures, calculations and relationships are studied, and the granular hierarchical structure is achieved. (3) Finally, we investigate rough set models in PG-Approx-Space. Accordingly, model regions are proposed by developing the classical regions, model expansion is systematically analyzed, some models are constructed as their structures are obtained, and a concrete model is provided. Based on the quantitative information architecture, this paper systematically conducts and investigates double quantification and establishes a fundamental and general exploration framework.

1. Introduction

1.1. Pawlak-Model, VPRS-Model, GRS-Model

Rough set (*RS*) theory, first proposed by Pawlak [23], is a data analysis theory as well as a mathematical tool for dealing with vague, inconsistent and incomplete information. Yao [51] provides this theory with two explanations: the set-oriented and operator-oriented views. As a relatively new methodology on soft computing, RS-Theory has been extensively emphasized in recent years, and its effectiveness has been confirmed by successful applications in many science and engineering

* Corresponding author at: Department of Computer Science and Technology, Tongji University, Shanghai 201804, PR China.
E-mail addresses: xianyongzh@sina.com.cn (X. Zhang), miaoduoqian@163.com.cn (D. Miao).

0020-0255/$ - see front matter
http://dx.doi.org/10.1016/j.ins.2013.09.020

fields. RS-Theory has become one of the focuses of research in artificial intelligence theory and its applications, and some relevant results have been reported in [4,9,13,32,44,59].

U is the finite and non-empty universe, while R denotes an equivalence relation. $x \in U$, equivalence class $[x]_R$ is also called the atom granule, while a granule is generally defined as the union of certain equivalence classes, and $U/R = \{[x]_R : x \in U\}$ represents knowledge. Thus, (U, R) constitutes the approximation space (*Approx-Space*). Suppose $A \subseteq U$, A represents a specific set generally called a concept, and $|A|$ refers to the cardinal number. In the Pawlak-Model [23], the upper and lower approximations are defined by

$$\overline{R}A = \cup\{[x]_R : [x]_R \cap A \neq \phi\}, \quad \underline{R}A = \cup\{[x]_R : [x]_R \subseteq A\}. \tag{1}$$

$$posRA = \underline{R}A, \quad negRA = \sim \overline{R}A, \quad bnRA = \overline{R}A - \underline{R}A \tag{2}$$

denote the positive region, negative region, and boundary region, respectively. In particular, the classified three-way regions have the specific qualitative semantics and reflect certainty and uncertainty, thus underlying knowledge discovery and data mining. Moreover, $\overline{R}$ and $\underline{R}$ are the corresponding approximation operators. As the Pawlak-Model is only a qualitative model, it has some limitations, such as no fault-tolerance mechanisms. Thus, the Pawlak-Model needs improving, while the quantitative RS model has great value.

Uncertainty is a main feature of artificial intelligence, while probability acts as an important tool to describe uncertainty. Thus, probability was introduced into RS-Theory to produce the probabilistic rough set [46,49,68]. The probabilistic rough set has many outstanding merits, such as measurability, generality and flexibility, and it functions well in many concrete models, such as the 0.5-probabilistic rough set [38], variable precision rough set (VPRS) [14,69], decision-theoretic rough set [50,54], game-theoretic rough set [2,7], parameterized rough set [6,26] and Bayesian rough set [36,58]. Meanwhile, the graded rough set (GRS) [52], also a typical quantitative model, has some complementarity with the probabilistic rough set. The VPRS and GRS are actually two fundamental types of quantitative models with both the fault-tolerance capability and extended feature, and as a result, they serve as this paper's main background models.

Ziarko [69] proposed the VPRS by introducing the relative degree of misclassification: $c([x]_R, A) = 1 - |[x]_R \cap A|/|[x]_R|$. With a threshold of $\beta \in [0, 0.5)$, the upper and lower approximations become

$$\overline{R}_\beta A = \cup\{[x]_R : c([x]_R, A) < 1 - \beta\}, \quad \underline{R}_\beta A = \cup\{[x]_R : c([x]_R, A) \leqslant \beta\}; \tag{3}$$

$$posR_\beta A = \underline{R}_\beta A, \quad negR_\beta A = \sim \overline{R}_\beta A, \quad bnR_\beta A = \overline{R}_\beta A - \underline{R}_\beta A \tag{4}$$

denote the positive region, negative region, and boundary region, respectively; $\overline{R}_\beta$ and $\underline{R}_\beta$ are the corresponding approximation operators. For largely solving the data noise problem, the VPRS has great significance for data acquisition and information analyses; moreover, it expands the Pawlak-Model. The VPRS is applied in many studies and applications, such as the attribute reduction and rule extraction in [1,3,8,10,11,19,37,57,66] and the practical results [20,22,39–41,45] in geological and medical fields, etc.

$p([x]_R, A) = 1 - c([x]_R, A) = |[x]_R \cap A|/|[x]_R|$ is the core measure in Approx-Space. It also corresponds to both the conditional probability in the decision-theoretic rough set [12,15,17,31,55,65] and the rough membership proposed by Pawlak and Skowron [24], and it further develops into the inclusion degree. Greco et al. [6] presented a generalized model of the VPRS by using the absolute and relative rough membership, while Singh and Dey [33] used the rough membership for document categorization. Refs. [19,30,43] used the inclusion degree to extensively explore measures, reasoning and applications on uncertainty. In this paper, $p([x]_R, A)$ is specifically called **precision** of $[x]_R$ with respect to A, while the threshold $\beta \in [0, 0.5)$ serves as the precision parameter. Note that precision is extracted from the VPRS but has already been generally promoted. Thus, $\overline{R}_\beta A = \cup\{[x]_R : p([x]_R, A) > \beta\}$ refers to *union of the equivalence classes whose precision with respect to A is greater than* β. $\underline{R}_\beta A = \cup\{[x]_R : p([x]_R, A) \geqslant 1 - \beta\}$ refers to *union of the equivalence classes whose precision with respect to A is not smaller than* $1 - \beta$. Similar to these quantitative descriptions, the precision description of threshold β is called **precision semantics**. Furthermore, the surplus three-way regions have their own precision semantics. In fact, precision semantics reflect the relative fault-description feature and determine the VPRS application function.

Yao and Lin [52] explored the relationships between rough sets and modal logics and proposed the GRS by utilizing the graded modal logics. The GRS basically describes the absolute quantitative information about knowledge and concepts and expands the Pawlak-Model. The discussion of the GRS mainly focuses on the model construction, such as those in [16,42,53], while its background (the graded modal logic) has fruitful studies. $\mathbf{N}$ denotes the natural number set and threshold $k \in \mathbf{N}$. In the GRS, the upper and lower approximations become

$$\overline{R}_k A = \cup\{[x]_R : |[x]_R \cap A| > k\}, \quad \underline{R}_k A = \cup\{[x]_R : |[x]_R| - |[x]_R \cap A| \leqslant k\}; \tag{5}$$

$\overline{R}_k$ and $\underline{R}_k$ are the corresponding approximation operators. Measures $|[x]_R \cap A|$ and $|[x]_R| - |[x]_R \cap A|$ reflect the absolute number of $[x]_R$ elements inside and outside A, respectively. In this paper, $\overline{g}([x]_R, A) = |[x]_R \cap A|$ and $\underline{g}([x]_R, A) = |[x]_R| - |[x]_R \cap A|$ are called **internal grade** and **external grade** of $[x]_R$ with respect to A, respectively, while threshold k serves as the grade parameter. Internal grade and external grade originate from the GRS, but after extraction and promotion, they have already transcended the GRS and, in fact, describe another group of applicable measures in Approx-Space. Thus, $\overline{R}_k A = \cup\{[x]_R : \overline{g}([x]_R, A) > k\}$ refers to *union of the equivalence classes whose internal grade with respect to A is greater than k*; $\underline{R}_k A = \cup\{[x]_R : \underline{g}([x]_R, A) \leqslant k\}$ refers to