Knowledge-Based Systems 91 (2016) 32–45

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

# A novel three-way decision model based on incomplete information system

Dun Liu [a,*], Decui Liang [b,*], Changchun Wang [a]

[a] School of Economics and Management, Southwest Jiaotong University, Chengdu 610031, China
[b] School of Management and Economics, University of Electronic Science and Technology of China, Chengdu 610054, China

ARTICLE INFO

*Article history:*
Received 14 January 2015
Revised 20 July 2015
Accepted 23 July 2015
Available online 30 July 2015

*Keywords:*
Three-way decisions
Decision-theoretic rough sets
Loss function
Incomplete information system
Hybrid information system

ABSTRACT

As a natural extension of three-way decisions with incomplete information, this paper provides a novel three-way decision model based on incomplete information system. First, we define a new relation to describe the similarity degree of incomplete information. Then, in view of the missing values presented in incomplete information system, we utilize interval number to acquire the loss function. A hybrid information table which consist both of the incomplete information and loss function, is used to deal with the new three-way decision model. The key steps and algorithm for constructing the integrated three-way decision model are also carefully investigated. An empirical study of medical diagnosis validates the reasonability and effectiveness of our proposed model.



## 1. Introduction

Three-way decisions (TWD), which were proposed by Yao in 2010 [59,60], have gradually became an important granular computing methodology and attracted many attentions in the nearly five years. The idea of three-way decisions is generated from decision-theoretic rough sets (DTRS) [66,68]. Intuitively, two thresholds $\alpha$ and $\beta$ of DTRS can divide the universe into three pairwise disjoint regions (positive, negative and boundary regions) by considering the minimum expected overall decision risk. The positive decision rules generated by the positive region make decision of acceptance. The negative decision rules generated by the negative region make decision of rejection. With the different from the two-way decisions of acceptance or rejection, the boundary region lead to a third way of decision, namely, noncommitment or deferment [71]. For simplicity, the three types of decision rules generated from the three regions of rough sets, form three-way decisions. As we stated in [35], three-way decisions are the natural extensions of DTRS, and they are common problem solving methodologies and consistent with human's real decision cognition.

In view of the semantics of DTRS, Yao systematically investigated the notion of three-way decisions and its potential applications recently [66,68]. As well, Liu et al. [28] briefly reviewed the two decades' researches on DTRS. Followed by their viewpoints, the existing studies of DTRS in three-way decisions can divide into three main aspects as follows.

- The extended models, modified models and their corresponding approaches on three-way decisions

In order to introduce the general binary relations, Abd El-Monsef and Kilany [1] constructed two new approximations (semilower and semiupper approximations) and proposed a generalization and modification of DTRS model. Herbert and Yao, Azam and Yao [3,12,13] systematically studied three-way decision-making in game-theoretic rough sets. Li and Zhou [18] considered the decision risks in DTRS, and further investigated the three-way decisions with optimistic decision, equable decision, and pessimistic decision. Yao and Zhou [67], Deng and Yao [6,7] utilized the Bayes theorem viewpoint and information-theoretic viewpoint to interpret the thresholds acquisition in probabilistic rough sets, respectively. Ma and Sun [38] extended the probabilistic rough set model to two universes by considering the Bayesian risk in decision making. In consideration of the multiple classifications, the multi-agents and the multi-granulation problems, Liu et al. [31] and Zhou [81] extended the three-way decisions from two-category to multi-category; Yang et al. [56] studied three-way decision-making with DTRS in the context of multi-agent systems; Qian et al. [47] introduced multigranulation

* Corresponding authors.
*E-mail addresses:* newton83@163.com (D. Liu), decuiliang@126.com (D. Liang), changchun2013@gmail.com (C. Wang).

http://dx.doi.org/10.1016/j.knosys.2015.07.036
0950-7051/

method to DTRS and proposed multigranulation decision-theoretic rough sets. In consideration of the uncertainty decision environment in three-way decisions, Liu and Liang proposed a series of new three-way decision models, including three-way decisions with random sets [36], interval sets [24], linguistic assessment [27], fuzzy interval sets [33], triangular fuzzy sets [23], intuitionistic fuzzy sets [26], hesitant fuzzy sets [25], and they further proposed a three-way decision model with logistic regression [35], dynamic three-way decision model [34] and function based three-way decision model [37]. Hu [14] systematically investigated three-way decision spaces in rough sets. Salehi et al. [48] did the systematic mapping studies on granular computing. Ciucci and Dubois [4,5] discussed the dependencies among three-valued logics, and further compared three-valued representations of imperfect information. Yu et al. [72,73] and Lingras et al. [22] investigated the three-way decision approaches with clustering analysis. All the above stated work make soiled contributions on the theoretical researches of three-way decisions.

- Attribute reduction on three-way decisions

Attribute reduction in rough sets is one of the most important issues, it provides an effective way to discover intrinsic knowledge hidden behind the data set by deleting the redundant information from the information system [45]. In DTRS model, the attribute reduction methods mainly focus on two scenarios, positive region based reduction and minimum cost based reduction. As to the first scenario, Yao and Zhao [64] analyzed various criteria for attribute reduction for probabilistic rough sets, such as decision-monotocity, generality and cost. Li et al. [19] further investigated the monotonicity of positive region in DTRS model, and presented a new definition of attribute reduction in DTRS model. For the second scenario, Jia et al. [15] discussed the minimum decision cost attribute reduction in DTRS model and proposed cost-based optimal reducts. Min et al. [41] proposed a test-cost-sensitive attribute reduction for rough set model. Zhao and Zhu [75] proposed an optimal cost-sensitive granularization method to address different sizes of the granule by considering variable test and misclassification costs. In addition, Ma et al. [39] investigated a decision region distribution preservation reduction in DTRS model. Ju et al. [16] considered $\delta$-cut DTRS and discussed the attribute reductions of the new model. Zhang and Miao [79] discussed the region-based quantitative and hierarchical attribute reduction in the two-category decision theoretic rough set model. To sum up, the researches on attribute reduction in DTRS can be easily related to, and interpreted by, more practical notions such as costs, losses and benefits.

- Different areas of application on three-way decisions

The essential ideas of three-way decisions have been widely applied in many fields, such as management decisions (e.g., environmental management [10], government management [32], oil investment management [29,30,40,42,55,74], model selection [9]), information and engineering (e.g., email spam filtering [76,80], E-learning [2], products inspecting process [54]), medical management (e.g., medical clinic [43], medical decision support system [58]), three-way recommender systems [77], etc.

As we stated above, the aforementioned literatures mainly consider the complete information of three-way decisions. The loss functions in their researches are directly given by the experts or dealt with as imprecise values (e.g., random numbers, interval numbers, fuzzy numbers, etc.), but the semantic relation between loss functions and information table are rarely discussed. Furthermore, in real-life applications, since some data could not be obtained for various reasons (e.g., capacity, technology, financing), missing data appears frequently in many information systems. In consideration of the information system with miss values, this paper introduces the incomplete information into DTRS and analyzes three-way decision approach based on the incomplete information system (IIS).

The remainder of this paper is organized as follows. In Section 2, we review some basic concepts of three-way decisions, DTRS, and rough sets under IIS. In Section 3, a novel three-way decision model with incomplete information is proposed. A hybrid information table, which consists both of the incomplete information and loss function, is used to deal with the three-way decision model in IIS. Then, a case study of medical diagnosis is given to illustrate our method in Section 4. Section 5 concludes the paper and elaborates on future studies.

## 2. Preliminaries

The basic concepts, notations and results of rough sets [44–46,49–51,78,82], DTRS [59–65] and three-way decisions [28,30,66,68–71] are briefly reviewed in this section.

### 2.1. Three-way decisions in rough sets

The idea of three-way decisions in rough sets is generated by rough set approximations [59,60]. As we know, Pawlak approximation space $apr = (U, R)$ is defined by a finite and non-empty set $U$ and an equivalence relation $R$. A partition of $U$, which generated from the equivalence relation $R$, can be denoted as $[x]_R$ or $[x]$. For $\forall X \subseteq U$, the lower and upper approximations of $X$ can be defined as:

$$\begin{aligned} \underline{apr}(X) &= \{x \in U | [x] \subseteq X\}; \\ \overline{apr}(X) &= \{x \in U | [x] \cap X \neq \varnothing\}. \end{aligned} \tag{1}$$

In (1), the condition $[x] \subseteq X$ in the lower approximation represents $[x]$ is contained in $X$. As well, the condition $[x] \cap X \neq \varnothing$ in the upper approximation means $[x]$ has an overlap with $X$. The two conditions clearly indicate the qualitative relationships between $[x]$ and $X$ in Pawlak rough sets. However, the definition of Pawlak approximations in (1) does not allow any errors, and the degree of overlap is not considered [69]. Observed by the limitation of Pawlak rough sets, probabilistic rough sets utilize two parameters, $\alpha$ and $\beta$ ($\alpha \geq \beta$), to extend Pawlak rough sets to a more generalized model. The two approximations of probabilistic rough sets can be rewritten as:

$$\begin{aligned} \underline{apr}_{(\alpha,\beta)}(X) &= \{x \in U | Pr(X|[x]) \geq \alpha\}, \\ \overline{apr}_{(\alpha,\beta)}(X) &= \{x \in U | Pr(X|[x]) > \beta\}, \end{aligned} \tag{2}$$

where, the rough membership function $Pr(X|[x]) = \frac{|X \cap [x]|}{|[x]|}$ is the conditional probability of the classification [46]. The two approximations of probabilistic rough sets can lead to three decision regions as $(\alpha, \beta)$-probabilistic positive, boundary and negative regions:

$$\begin{aligned} \mathrm{POS}_{(\alpha,\beta)}(X) &= \{x \in U | Pr(X|[x]) \geq \alpha\}, \\ \mathrm{BND}_{(\alpha,\beta)}(X) &= \{x \in U | \beta < Pr(X|[x]) < \alpha\}, \\ \mathrm{NEG}_{(\alpha,\beta)}(X) &= \{x \in U | Pr(X|[x]) \leq \beta\}. \end{aligned} \tag{3}$$

According to (3), the three regions lead to three-way decisions, namely, decision of acceptance, deferment and rejection, respectively.

Specially, if $\alpha < \beta$, we set "$\gamma = \alpha = \beta$", (3) can be rewritten as:

$$\begin{aligned} \mathrm{POS}_{(\gamma,\gamma)}(X) &= \{x \in U | Pr(X|[x]) \geq \gamma\}, \\ \mathrm{NEG}_{(\gamma,\gamma)}(X) &= \{x \in U | Pr(X|[x]) < \gamma\}. \end{aligned} \tag{4}$$

For simplicity and clarity, we denote (4) as the two-way decision model. Obviously, (4) is a special case of the three-way decision model when $\alpha = \beta$. In addition, if we set "$\alpha = 1, \beta = 0$", (3) converts to Pawlak rough set model; if we set $\alpha = \beta = 0.5$, (3) converts to 0.5 probabilistic rough sets [30].