International Journal of Approximate Reasoning 88 (2017) 1–22

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

A three-way decisions model with probabilistic rough sets for stream computing ☆

Jianfeng Xu [a,b,c], Duoqian Miao [a,c,*], Yuanjian Zhang [a,c], Zhifei Zhang [a,d]

[a] *Department of Computer Science and Technology, Tongji University, Shanghai 201804, P.R. China*
[b] *Software College, Nanchang University, Jiangxi 330077, P.R. China*
[c] *Key Laboratory of Embedded System and Service Computing, Ministry of Education, Tongji University, Shanghai 201804, P.R. China*
[d] *Big Data and Network Security Research Center, Tongji University, Shanghai 200092, P.R. China*

ARTICLE INFO

Article history:
Received 29 June 2016
Received in revised form 30 December 2016
Accepted 8 May 2017
Available online 11 May 2017

Keywords:
Three-way decisions
Probabilistic rough sets
Stream computing learning method
Knowledge updating

ABSTRACT

Stream computing paradigm, with the characteristics of real-time arrival and departure, has been admitted as a major computing paradigm in big data. Relevant theories are flourishing recently with the surge development of stream computing platforms such as Storm, Kafka and Spark. Rough set theory is an effective tool to extract knowledge with imperfect information, however, related discussions on synchronous immigration and emigration of objects have not been investigated. In this paper, stream computing learning method is proposed on the basis of existing incremental learning studies. This method aims at solving challenges resulted from simultaneous addition and deletion of objects. Based on novel learning method, a stream computing algorithm called single-object stream-computing-based three-way decisions (SS3WD) is developed. In this algorithm, the probabilistic rough set model is applied to approximate the dynamic variation of concepts. Three-way regions can be determined without multiple scans of existing information granular. Extensive experiments not only demonstrate better efficiency and robustness of SS3WD in the presence of objects streaming variation, but also illustrate that stream computing learning method is an effective computing strategy for big data.

1. Introduction

The unprecedented popularity of novel information technology and application schema, such as cloud computing, Internet of things (IOT), and mobile interconnection, accumulate a large scale of data and promote the development of big data [2,7,39]. The essential characteristics of big data have been summarized by many scientists, and currently 5V model and 5R model [8] are widely accepted. Generally speaking, data with any properties or requirements mentioned in 5V and 5R model can be considered as big data. Fast arrival, for example, is admitted as one of the most remarkable challenges. On one hand, desirable result of up-to-date data cannot be achieved in a limited time because of high velocity (defined in 5V), on the other hand the value of complicated applications [13,15,33] will be diminished if hidden knowledge is not extracted real-time (defined in 5R). Specifically, there are two major reasons:

☆ This paper is part of the virtual special issue on tri-partition, edited by Davide Ciucci and Yiyu Yao.

* Corresponding author at: Department of Computer Science and Technology, Tongji University, Shanghai 201804, P.R. China.
E-mail address: dqmiao@tongji.edu.cn (D. Miao).

http://dx.doi.org/10.1016/j.ijar.2017.05.001

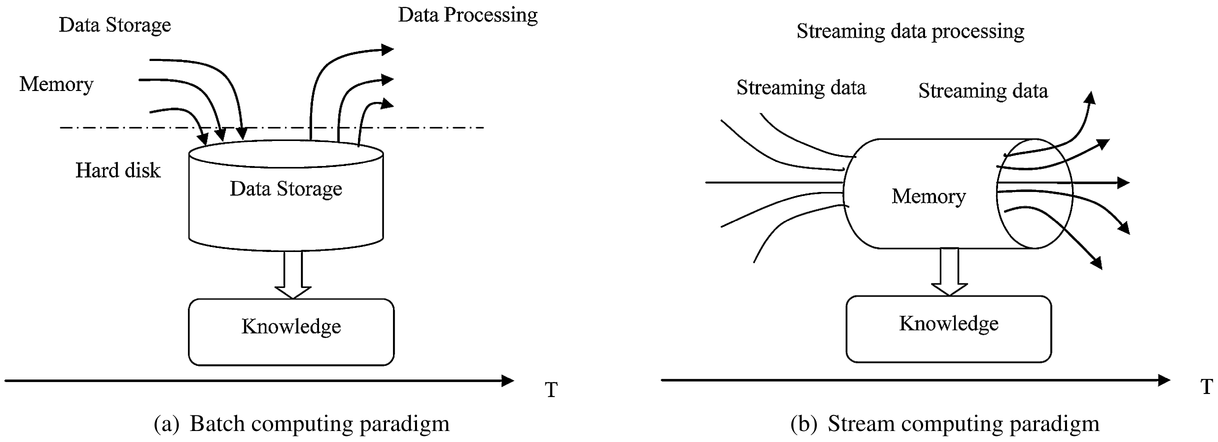

Fig. 1. Typical computing paradigm for big data [37].

The explosive generation of data in a limited time is ubiquitous. For example, CERN's large hadron collider produce petabytes of data per second in the working status. Repeating such experiment is quite costly and thus one-pass scan of such application is very important.

In some other applications, however, data is poured from a tremendous number of interacting instances, despite the seemingly negligible contribution of each participant. Typical examples of this style include click stream, RFID data and GPS location information.

Obviously, these kind of data need to be processed in the manner of *stream computing*, i.e. computed the whole data segmentally and sequentially. To solve the dilemma of real-time and accurate, scholars have suggested a wealth of ideas and they can be categorized into two computing paradigms: *batch computing paradigm* and *stream computing paradigm*.

(1) Batch computing paradigm [46] stores and computes data in batches, whereas relations between batches are neglected. In most cases, both stages are handled in a highly centralized way. As depicted in Fig. 1(a), batch computing launches only when accumulated data is abundant.

(2) Stream computing paradigm [31], however, performs data storage and computing in memory and consider the relation of batches in the way of sliding window. As described in Fig. 1(b), no data exchange occurs between hard disk and memory.

Generally, any machine learning method can be customized into batch learning paradigm and stream computing paradigm. Incremental learning method can accelerate the speed of stream data. However, most of existing incremental learning method compute information addition and deletion separately, although both operations may be considered [14]. To implement stream learning paradigm more effectively, we term a new glossary called *stream computing learning method*.

Stream computing learning method is defined as a novel strategy whose operations towards objects immigrations and emigrations are conducted at the same time.

Obviously, stream computing learning method is more consistent to the connotation of stream computing paradigm as compared to classical incremental learning method. Currently, researches on the variation mechanism of data, fast real-time computing and approximate real-time computing are rather preliminary. It is undoubtedly that approximation instead of accurate answer is more likely to be achieved, therefore it is imperative to introduce new theory to facilitate the research of stream computing learning method.

Three-way decisions theory [43] is an important extension of rough set theory [34]. Decisions are determined if it is informative, otherwise will be deferred. It is a rather inclusive paradigm since the hidden structure used to support decision-making can be generated by any kinds of learning mechanism. Gradually, it has been recognized that the theory has incomparable advantages in solving complicated problem because of analogous cognitive mechanism shadowed in human [22,32,45]. Currently, the research direction of three-way decisions are mainly concentrated on the following aspects: 1) the basic theory of three-way decisions [9,11,16,17]; 2) three-way decision and rough sets theory [18,41,44,53,54]; and 3) clustering/classification based on three-way decisions [47–49].

The contributions of this paper are as follows. Firstly, it is the first time to systematically clarify the hierarchical structure of stream computing. From the coarsest to refinement, we have stream computing paradigm, stream computing learning method, and stream computing learning algorithms. Major differences against incremental learning lies in the level of stream computing learning method. While incremental learning method performs computations in the unit of information variation direction of dataset, i.e. either addition or deletion, stream computing learning method combines both immigrations and emigrations into an atomic operation unit. It is straightforward to see that the stream computing learning method is tightly coupled as compared to incremental learning method. Consequently, detailed stream computing learning algorithm is also different from incremental learning algorithm, making the knowledge updating more purposeful. Secondly, the use of three-way decisions in stream computing algorithms is also initially explored. Probabilistic rough set is used to approximate the concept given variations of addition and deletion are computed simultaneously. Thirdly, a novel stream computing algorithm called single-object stream-computing-based three-way decisions (SS3WD) is proposed. This algorithm utilizes conditional probability to approximate the transition of three-way regions, and finally by theoretical and practical analysis, we claim that at any time SS3WD not only run faster than incremental learning algorithm, but also scalable in different datasets. It demonstrates that stream computing learning method is effective.

The rest of the paper is organized as follows. In Section 2, the necessity and status on stream computing paradigm and probabilistic rough sets three-way decisions are discussed. In Section 3, problem formulation and related theorems/corollaries are presented as foundations of proposed algorithm. In Section 4, the effectiveness is further demonstrated from case study and benchmark datasets. Finally, the whole paper is concluded in Section 5.

2. Related work

Rough set is widely used in incremental learning. This section will discuss the status of dynamic learning on the basis of incremental learning so that similar ideas can be transferred to flourish stream computing learning method.

Currently, studies of incremental learning with rough sets are based on different extensions of rough sets [3,23,24,27,52] including both complete information system [19,20] and incomplete information system [27,35,36]. Inspired by the analysis in rough set, corresponding researches are mainly focused in two aspects, known as element variation and learning task. While variation of objects [6,26,29] and attributes [10,50,51] are critically analyzed in element variation, variations of lower and upper approximations [5,25,35,50,52], attribute reduction [12,19,21,38,40] and decision rules [1,26,28] are frequently investigated in learning tasks. Representative work are shown as follows:

(1) Variation of objects: Luo et al. [29] presented an efficient incremental learning algorithm using probabilistic rough sets. Chen et al. [6] formalized a matrix-based method which consider the variation of objects and attributes simultaneously.

(2) Variation of attributes: Chen et al. [3] discussed a dynamic maintenance approach for approximations in coarsening and refining attribute values based on rough sets theory in an incomplete system. Furthermore, Chen et al. [4] interpreted the maintenance of approximations in incomplete ordered decision systems while attribute values coarsening or refining and proposed algorithms for incremental updating approximations of upward and downward union of classes. Liu et al. [28] investigated two incremental algorithms based on adding attributes and deleting attributes under probability rough sets.

(3) Lower and upper approximations: Zhang et al. [51] proposed the attribute set in the set-valued information system may evolve over time when new information arrives, and the incremental approaches for updating the relation matrix are proposed to update rough sets approximations. To compute dynamic approximations of the multiple type of data, Zhang et al. [52] defines a composite rough set. Luo et al. [30] extended the relation matrix on decision-theoretic rough set. Chen et al. [6] further investigated a method called UAGOAAS to estimate three-way region alteration when attributes and objects are changed simultaneously.

(4) Attribute reduction: Das et al. [10] claimed that attribute reduction launches only when error upper bound is significantly improved. Given varying attribute values, Wang et al. [38] developed an attribute reduction algorithm based on complementary entropy, conditional entropy, and conditional complementary entropy.

(5) Decision rules: Liu et al. [28] introduced a new concept of interesting knowledge based on both accuracy and coverage for dynamic information system, and an incremental paradigm and approach for inducing decision rules are proposed when the object set varies over time.

Furthermore, Although few studies investigated object immigration and emigration algorithms [26,29], they simply put them together and failed to consider the realization of stream computing learning method given addition and deletion of objects occuring simultaneously. Table 1 is a comparison of research status in incremental learning method and stream computing learning method.

From Table 1, we can see that studies of stream computing learning method is quite limited as compared to incremental learning method. Updating of hidden structure and uncertainty measure is important in incremental learning, and it should be well considered to further reduce the computational cost. Therefore, much work can be realized in designing stream computing algorithm.

3. Preliminary

In this section, we will review the basic notions and concepts for three-way decisions from the perspective of probability [43,44]. Information system is the foundation of intelligent information processing. Typically, an information system is defined as follows.

Definition 1. An information system is defined by a quadruple tuple: $IS = (U, A, V, f)$ where U is a finite non-empty set of data objects. $A = C \bigcup D$ is a finite non-empty set of attributes, where C is a set of condition attribute, D is a set of decision attribute. V is a non-empty set of values of $a \in A$, and f is an information function from U to V.

Table 1
Research status of dynamic computing using rough sets.

Research topic	Incremental learning method		Stream computing learning method	
	Researches	Status	Researches	Status
Variation of objects	Chen et al. [6] Luo et al. [29]	To be perfected	Not given	To be analyzed
Variation of attributes	Chen et al. [4] Li et al. [23] Li et al. [25] Zeng et al. [50]	To be perfected	Not given	To be analyzed
Lower and upper approximations	Chen et al. [6] Qian et al. [35] Zhang et al. [51] Luo et al. [29]	To be perfected	Not given	To be analyzed
Attribute reduction	Dey et al. [12] Wang et al. [38]	To be perfected	Eskandari et al. [14]	To be perfected
Decision rules	Błaszczyński et al. [1] Liu et al. [26]	To be perfected	Not given	To be analyzed

The hidden structure of information, or information granular, represent the similarity/dissimilarity relations among objects. Equivalence relation is regarded as the fundamental criterion to discern objects. Definition of equivalence relation are shown as follows.

Definition 2. Given a subset of attributes $B \subseteq A$ in IS, the $IND(B)$ denotes an equivalence relation, which can be defined as follows:

$$IND(B) = \{(x, y) \in U \times U | \forall a \in B, f(x,a) = f(y,a)\}, \tag{1}$$

where $R_i \in U/C$, and U/C is equivalent partition of condition attribute, with its basis $|U/C| = m$;
$D_j \in U/D$, and U/D is equivalent partition of decision attribute, with its basis $|U/D| = n$.

The affiliation of objects to class can be determined by adopting maximum inclusion degree of information granular among all classes, and the decisions should suffice the requirement of thresholds meanwhile. In real practice, useful rules can be extracted given the thresholds located in the interval $[0, 1]$. Regarding inclusion degree as conditional probability, the equation probability is defined as follows.

Definition 3. Given a subset $D_j \subseteq U$ in IS, the conditional probability of an object belonging to D_j given that the object belongs to [x]. This probability may be simply estimated as follows:

$$Pr(D_j|[\mathrm{x}]) = \frac{\left|D_j \cap [x]\right|}{|[x]|} \tag{2}$$

where $|\bullet|$ denotes the cardinality of a set.

The result of conditional probability divides the whole universe into three regions named as positive region (POS), boundary region (BND) and negative region (NEG) respectively. Details of three regions are described as follows.

Definition 4. Given a pair of thresholds α and β with $0 \le \beta < \alpha \le 1$, the positive, boundary and negative regions are defined as follows:

$$\begin{aligned} &POS_{(\alpha,\bullet)}(D_j) = \{x \in U \| \Pr(D_j|[x]) \ge \alpha\};\\ &BND_{(\alpha,\beta)}(D_j) = \{x \in U | \beta < \Pr(D_j|[x]) < \alpha\};\\ &NEG_{(\bullet,\beta)}(D_j) = \{x \in U | \Pr(D_j|[x]) \le \beta\}. \end{aligned} \tag{3}$$

For objects allocated in different regions, corresponding decisions can be determined. The acceptance decision confirms affiliations of object *w.r.t* class, whereas the rejection decision contributes to the class boundary. However, there is a third possibility signifies deferment, and it can be explained as insufficiency of given information. Detailed definitions are given as follows.

Definition 5. According to the three probabilistic regions, one can make three-way decisions of acceptance, deferment and rejection, respectively.

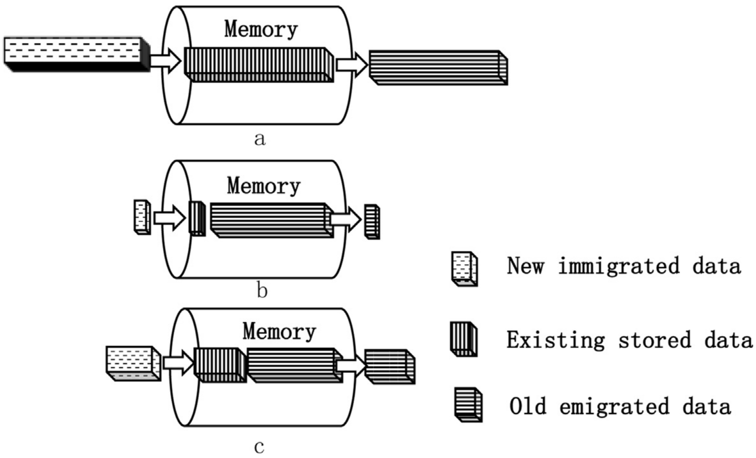

Fig. 2. Cases of stream computing learning method.

$$\begin{aligned}&DES_{Accept}(R_i \to D_j), \quad \text{for} \quad R_i \subseteq POS_{(\alpha,\bullet)}(D_j);\\&DES_{Defer}(R_i \to D_j), \quad \text{for} \quad R_i \subseteq BND_{(\alpha,\beta)}(D_j);\\&DES_{Reject}(R_i \to D_j), \quad \text{for} \quad R_i \subseteq NEG_{(\bullet,\beta)}(D_j).\end{aligned} \tag{4}$$

4. Stream computing algorithm for three-way knowledge updating

Real-time memory calculation is the main component of stream computing paradigm. According to dynamic characteristics of object, real-time stream computing can be divided into three cases, as shown in Fig. 2:

(1) Case 1 (Simple stream computing learning method): As shown in case (a) of Fig. 2, the calculation interval within two computing period is not continued, and in each calculation there are only new data in memory whereas the previous used data is flushed.

(2) Case 2 (Single-object dynamic stream computing learning method): As described in case (b) of Fig. 2, the calculation interval reduces significantly. The adding of one more data triggers the deleting of data with earliest timestamp.

(3) Case 3 (Batched-object dynamic stream calculating learning method): As illustrated in case (c) of Fig. 2, the calculation interval is very similar to case 2. However, the amount of added data and deleted data are expanded from one to n.

Since the special requirements for high-speed computing in stream computing learning method, developing algorithms which can perform faster and more effective is not only an important research topic, but also one of main research directions in the big data computing [42] research. In what follows, we will analyze the knowledge updating in stream computing learning method with single-object (case 2 in Fig. 2).

The formalization of single-object variation in the context of stream computing learning method is given as follows:

Let $IS^{(t)} = \left\{U^{(t)}, C^{(t)} \cup D^{(t)}\right\}$ be an information system at time t, with $U^{(t)}/C^{(t)} = \left\{R_1^{(t)}, R_2^{(t)}, \cdots, R_m^{(t)}\right\}$ and $U^{(t)}/D^{(t)} = \left\{D_1^{(t)}, D_2^{(t)}, \cdots, D_n^{(t)}\right\}$ be equivalent partition of condition and decision respectively. Assuming there is an object $\overline{x}$ immigrates to IS whereas $\underline{x}$ emigrates from IS, then the equivalent partitions of condition are correspondingly renewed as $U^{(t+1)}/C^{(t+1)} = \left\{R_1^{(t+1)}, R_2^{(t+1)}, \cdots, R_m^{(t+1)}\right\}$ and equivalent partitions of decision are correspondingly denoted as $U^{(t+1)}/D^{(t+1)} = \left\{D_1^{(t+1)}, D_2^{(t+1)}, \cdots, D_n^{(t+1)}\right\}$.

$$R_i^{(t+1)} = \begin{cases} R_i^{(t)} - \{\underline{x}\}, & \underline{x} \in R_i^{(t)} \wedge \overline{x} \notin R_i^{(t+1)}, \quad 1 \le i \le m \\ R_i^{(t)} \cup \{\overline{x}\}, & \underline{x} \notin R_i^{(t)} \wedge \overline{x} \in R_i^{(t+1)}, \quad 1 \le i \le m \\ R_i^{(t)} \cup \{\overline{x}\} - \{\underline{x}\}, & \underline{x} \in R_i^{(t)} \wedge \overline{x} \in R_i^{(t+1)}, \quad 1 \le i \le m \\ R_i^{(t)}, & \underline{x} \notin R_i^{(t)} \wedge \overline{x} \notin R_i^{(t+1)}, \quad 1 \le i \le m \\ \{\overline{x}\}, & \overline{x} \in R_i^{(t+1)}, \quad i = m+1 \end{cases} \tag{5}$$

$$D_j^{(t+1)} = \begin{cases} D_j^{(t)} - \{\underline{x}\}, & \underline{x} \in D_j^{(t)} \wedge \overline{x} \notin D_j^{(t+1)}, \quad 1 \le j \le n \\ D_j^{(t)} \cup \{\overline{x}\}, & \underline{x} \notin D_j^{(t)} \wedge \overline{x} \notin D_j^{(t+1)}, \quad 1 \le j \le n \\ D_j^{(t)} \cup \{\overline{x}\} - \{\underline{x}\}, & \underline{x} \in D_j^{(t)} \wedge \overline{x} \in D_j^{(t+1)}, \quad 1 \le j \le n \\ D_j^{(t)}, & \underline{x} \notin D_j^{(t)} \wedge \overline{x} \notin D_j^{(t+1)}, \quad 1 \le j \le n \\ \{\overline{x}\}, & \overline{x} \in D_j^{(t+1)}, \quad j = n+1 \end{cases} \tag{6}$$

Equations (5) and (6) enumerate the variation of condition and decision in terms of each equivalence class. The variations can be classified into trivial and non-trivial groups. The trivial variation is that the immigration of new object does not belong to any existing equivalence class of condition ($R_i^{(t+1)} = \{\overline{x}\}$, where $i = m+1$) or any existing equivalence class of decision ($D_j^{(t+1)} = \{\overline{x}\}$, where $j = n+1$). In this case, $\underline{x}$ can be directly determined into the positive region of its corresponding decision label. The non-trivial case is the remaining combination of equation (4) and equation (5), and the variation of conditional probability will be critically investigated later, as shown in Theorems 1–4.

4.1. Stream computing learning algorithm towards conditional probability

Theorem 1. *Let $IS^{(t)} = \left\{U^{(t)}, C^{(t)} \cup D^{(t)}\right\}$ and $IS^{(t+1)} = \left\{U^{(t+1)}, C^{(t+1)} \cup D^{(t+1)}\right\}$ be an information system at time t and $t+1$ respectively. If variation of immigrated object $\overline{x}$ satisfies $\overline{x} \notin R_i^{(t+1)} \wedge \overline{x} \notin D_j^{(t+1)}$, variation trend of conditional probability w.r.t. variation of emigrated object are given as follows:*

$$(1)\ \left(\underline{x} \notin R_i^{(t)} \wedge \underline{x} \notin D_j^{(t)}\right) \Rightarrow \Pr\left(D_j^{(t+1)} \middle| R_i^{(t+1)}\right) = \Pr\left(D_j^{(t)} \middle| R_i^{(t)}\right);$$

$$(2)\ \left(\underline{x} \in R_i^{(t)} \wedge \underline{x} \notin D_j^{(t)}\right) \Rightarrow \Pr\left(D_j^{(t+1)} \middle| R_i^{(t+1)}\right) > \Pr\left(D_j^{(t)} \middle| R_i^{(t)}\right);$$

$$(3)\ \left(\underline{x} \notin R_i^{(t)} \wedge \underline{x} \in D_j^{(t)}\right) \Rightarrow \Pr\left(D_j^{(t+1)} \middle| R_i^{(t+1)}\right) = \Pr\left(D_j^{(t)} \middle| R_i^{(t)}\right);$$

$$(4)\ \left(\underline{x} \in R_i^{(t)} \wedge \underline{x} \in D_j^{(t)}\right) \Rightarrow \Pr\left(D_j^{(t+1)} \middle| R_i^{(t+1)}\right) < \Pr\left(D_j^{(t)} \middle| R_i^{(t)}\right).$$

Proof. (1) Since $\left(\overline{x} \notin R_i^{(t+1)} \wedge \overline{x} \notin D_j^{(t+1)}\right) \wedge \left(\underline{x} \notin R_i^{(t)} \wedge \underline{x} \notin D_j^{(t)}\right) \Rightarrow \left(R_i^{(t+1)} = R_i^{(t)}\right) \wedge \left(D_j^{(t+1)} = D_j^{(t)}\right)$, according to Definition 3, $\Pr(D_j^{(t+1)} \left| R_i^{(t+1)}\right.) = \frac{\left|D_j^{(t+1)} \cap R_i^{(t+1)}\right|}{\left|R_i^{(t+1)}\right|} = \frac{\left|D_j^{(t)} \cap R_i^{(t)}\right|}{\left|R_i^{(t)}\right|} = \Pr(D_j^{(t)} \left| R_i^{(t)}\right.)$, namely, $\Pr(D_j^{(t+1)} \left| R_i^{(t+1)}\right.) = \Pr(D_j^{(t)} \left| R_i^{(t)}\right.)$. Hence

$$\Pr\left(D_j^{(t+1)} \middle| R_i^{(t+1)}\right) = \Pr\left(D_j^{(t)} \middle| R_i^{(t)}\right).$$

(2) Since $\left(\overline{x} \notin R_i^{(t+1)} \wedge \overline{x} \notin D_j^{(t+1)}\right) \wedge \left(\underline{x} \in R_i^{(t)} \wedge \underline{x} \notin D_j^{(t)}\right) \Rightarrow \left(R_i^{(t+1)} = R_i^{(t)} - \{\underline{x}\}\right) \wedge \left(D_j^{(t+1)} = D_j^{(t)}\right)$, according to Definition 3, $\Pr\left(D_j^{(t+1)} \middle| R_i^{(t+1)}\right) = \frac{\left|D_j^{(t+1)} \cap R_i^{(t+1)}\right|}{\left|R_i^{(t+1)}\right|} = \frac{\left|D_j^{(t)} \cap R_i^{(t)} - \{\underline{x}\}\right|}{\left|R_i^{(t)} - \{\underline{x}\}\right|} = \frac{\left|D_j^{(t)} \cap R_i^{(t)}\right|}{\left|R_i^{(t)}\right| - 1} > \frac{\left|D_j^{(t)} \cap R_i^{(t)}\right|}{\left|R_i^{(t)}\right|} = \Pr\left(D_j^{(t)} \middle| R_i^{(t)}\right)$. Hence

$$\Pr\left(D_j^{(t+1)} \middle| R_i^{(t+1)}\right) > \Pr\left(D_j^{(t)} \middle| R_i^{(t)}\right).$$

(3) Since $\left(\overline{x} \notin R_i^{(t+1)} \wedge \overline{x} \notin D_j^{(t+1)}\right) \wedge \left(\underline{x} \notin R_i^{(t)} \wedge \underline{x} \in D_j^{(t)}\right) \Rightarrow \left(R_i^{(t+1)} = R_i^{(t)}\right) \wedge \left(D_j^{(t+1)} = D_j^{(t)} - \{\underline{x}\}\right)$, according to Definition 3, $\Pr\left(D_j^{(t+1)} \middle| R_i^{(t+1)}\right) = \frac{\left|D_j^{(t+1)} \cap R_i^{(t+1)}\right|}{\left|R_i^{(t+1)}\right|} = \frac{\left|(D_j^{(t)} - \{\underline{x}\}) \cap R_i^{(t)}\right|}{\left|R_i^{(t)}\right|} = \frac{\left|D_j^{(t)} \cap R_i^{(t)}\right|}{\left|R_i^{(t)}\right|} = \Pr\left(D_j^{(t)} \middle| R_i^{(t)}\right)$. Hence

$$\Pr\left(D_j^{(t+1)} \middle| R_i^{(t+1)}\right) = \Pr\left(D_j^{(t)} \middle| R_i^{(t)}\right).$$

(4) Since $\left(\overline{x} \notin R_i^{(t+1)} \wedge \overline{x} \notin D_j^{(t+1)}\right) \wedge \left(\underline{x} \in R_i^{(t)} \wedge \underline{x} \in D_j^{(t)}\right) \Rightarrow \left(R_i^{(t+1)} = R_i^{(t)} - \{\underline{x}\}\right) \wedge \left(D_j^{(t+1)} = D_j^{(t)} - \{\underline{x}\}\right)$, according to Definition 3, $\Pr\left(D_j^{(t+1)} \middle| R_i^{(t+1)}\right) = \frac{\left|D_j^{(t+1)} \cap R_i^{(t+1)}\right|}{\left|R_i^{(t+1)}\right|} = \frac{\left|(D_j^{(t)} - \{\underline{x}\}) \cap (R_i^{(t)} - \{\underline{x}\})\right|}{\left|R_i^{(t)} - \{\underline{x}\}\right|} = \frac{\left|D_j^{(t)} \cap R_i^{(t)}\right| - 1}{\left|R_i^{(t)}\right| - 1} < \frac{\left|D_j^{(t)} \cap R_i^{(t)}\right|}{\left|R_i^{(t)}\right|} = \Pr\left(D_j^{(t)} \middle| R_i^{(t)}\right)$. Hence $\Pr\left(D_j^{(t+1)} \middle| R_i^{(t+1)}\right) < \Pr\left(D_j^{(t)} \middle| R_i^{(t)}\right)$. □

Theorem 1 summarizes the change of conditional probability for single-object variation of all $\underline{x}$ given $\overline{x} \notin R_i^{(t+1)} \wedge \overline{x} \notin D_j^{(t+1)}$. Therefore, it lays a solid foundation for estimating changes of conditional probability if immigrated object neither belongs to equivalence class of considered conditions nor belongs to equivalence class of considered decisions.

Theorem 2. *Let $IS^{(t)} = \left\{U^{(t)}, C^{(t)} \cup D^{(t)}\right\}$ and $IS^{(t+1)} = \left\{U^{(t+1)}, C^{(t+1)} \cup D^{(t+1)}\right\}$ be an information system at time t and $t+1$ respectively. If variation of immigrated object $\overline{x}$ satisfies $\overline{x} \notin R_i^{(t+1)} \wedge \overline{x} \in D_j^{(t+1)}$, variation trend of conditional probability w.r.t. variation of emigrated object are given as follows:*

(1) $\left(\underline{x} \notin R_i^{(t)} \wedge \underline{x} \notin D_j^{(t)}\right) \Rightarrow \Pr\left(D_j^{(t+1)} \middle| R_i^{(t+1)}\right) = \Pr\left(D_j^{(t)} \middle| R_i^{(t)}\right)$;

(2) $\left(\underline{x} \in R_i^{(t)} \wedge \underline{x} \notin D_j^{(t)}\right) \Rightarrow \Pr\left(D_j^{(t+1)} \middle| R_i^{(t+1)}\right) > \Pr\left(D_j^{(t)} \middle| R_i^{(t)}\right)$;

(3) $\left(\underline{x} \notin R_i^{(t)} \wedge \underline{x} \in D_j^{(t)}\right) \Rightarrow \Pr\left(D_j^{(t+1)} \middle| R_i^{(t+1)}\right) = \Pr\left(D_j^{(t)} \middle| R_i^{(t)}\right)$;

(4) $\left(\underline{x} \in R_i^{(t)} \wedge \underline{x} \in D_j^{(t)}\right) \Rightarrow \Pr\left(D_j^{(t+1)} \middle| R_i^{(t+1)}\right) < \Pr\left(D_j^{(t)} \middle| R_i^{(t)}\right)$.

Proof. (1) Since $\left(\overline{x} \notin R_i^{(t+1)} \wedge \overline{x} \in D_j^{(t+1)}\right) \wedge \left(\underline{x} \notin R_i^{(t)} \wedge \underline{x} \notin D_j^{(t)}\right) \Rightarrow \left(R_i^{(t+1)} = R_i^{(t)}\right) \wedge \left(D_j^{(t+1)} = D_j^{(t)} \cup \overline{x}\right)$, according to Definition 3 $\Pr\left(D_j^{(t+1)} \middle| R_i^{(t+1)}\right) = \frac{\left|D_j^{(t+1)} \cap R_i^{(t+1)}\right|}{\left|R_i^{(t+1)}\right|} = \frac{\left|(D_j^{(t)} \cup \{\overline{x}\}) \cap (R_i^{(t)} - \{\underline{x}\})\right|}{\left|R_i^{(t)} - \{\underline{x}\}\right|} = \frac{\left|D_j^{(t)} \cap R_i^{(t)}\right|}{\left|R_i^{(t)}\right|} = \Pr\left(D_j^{(t)} \middle| R_i^{(t)}\right)$. Hence

$$\Pr\left(D_j^{(t+1)} \middle| R_i^{(t+1)}\right) = \Pr\left(D_j^{(t)} \middle| R_i^{(t)}\right).$$

The proofs of (2), (3), and (4) are similar to that of (1). □

Theorem 2 illustrates the change of conditional probability for single-object variation of all $\underline{x}$ given $\overline{x} \notin R_i^{(t+1)} \wedge \overline{x} \in D_j^{(t+1)}$. Therefore, it lays a solid foundation for estimating changes of conditional probability if immigrated object does not belong to equivalence class of considered conditions whereas belongs to equivalence class of considered decisions.

Theorem 3. *Let $IS^{(t)} = \left\{U^{(t)}, C^{(t)} \cup D^{(t)}\right\}$ and $IS^{(t+1)} = \left\{U^{(t+1)}, C^{(t+1)} \cup D^{(t+1)}\right\}$ be an information system at time t and $t+1$ respectively. If variation of immigrated object $\overline{x}$ satisfies $\overline{x} \in R_i^{(t+1)} \wedge \overline{x} \notin D_j^{(t+1)}$, variation trend of conditional probability w.r.t. variation of emigrated object are given as follows:*

(1) $\left(\underline{x} \notin R_i^{(t)} \wedge \underline{x} \notin D_j^{(t)}\right) \Rightarrow \Pr\left(D_j^{(t+1)} \middle| R_i^{(t+1)}\right) < \Pr\left(D_j^{(t)} \middle| R_i^{(t)}\right)$;

(2) $\left(\underline{x} \in R_i^{(t)} \wedge \underline{x} \notin D_j^{(t)}\right) \Rightarrow \Pr\left(D_j^{(t+1)} \middle| R_i^{(t+1)}\right) = \Pr\left(D_j^{(t)} \middle| R_i^{(t)}\right)$;

(3) $\left(\underline{x} \notin R_i^{(t)} \wedge \underline{x} \in D_j^{(t)}\right) \Rightarrow \Pr\left(D_j^{(t+1)} \middle| R_i^{(t+1)}\right) < \Pr\left(D_j^{(t)} \middle| R_i^{(t)}\right)$;

(4) $\left(\underline{x} \in R_i^{(t)} \wedge \underline{x} \in D_j^{(t)}\right) \Rightarrow \Pr\left(D_j^{(t+1)} \middle| R_i^{(t+1)}\right) < \Pr\left(D_j^{(t)} \middle| R_i^{(t)}\right)$.

Proof. (1) Since $\left(\overline{x} \in R_i^{(t+1)} \wedge \overline{x} \notin D_j^{(t+1)}\right) \wedge \left(\underline{x} \notin R_i^{(t)} \wedge \underline{x} \notin D_j^{(t)}\right) \Rightarrow \left(R_i^{(t+1)} = R_i^{(t)} \cup \{\overline{x}\}\right) \wedge \left(D_j^{(t+1)} = D_j^{(t)}\right)$, according to Definition 3, $\Pr\left(D_j^{(t+1)} \middle| R_i^{(t+1)}\right) = \frac{\left|D_j^{(t+1)} \cap R_i^{(t+1)}\right|}{\left|R_i^{(t+1)}\right|} = \frac{\left|D_j^{(t)} \cap \left\{R_i^{(t)} \cup \{\overline{x}\}\right\}\right|}{\left|R_i^{(t)} \cup \{\overline{x}\}\right|} = \frac{\left|D_j^{(t)} \cap R_i^{(t)}\right|}{\left|R_i^{(t)}\right| + 1} < \Pr\left(D_j^{(t)} \middle| R_i^{(t)}\right)$. Hence

$$\Pr\left(D_j^{(t+1)} \middle| R_i^{(t+1)}\right) < \Pr\left(D_j^{(t)} \middle| R_i^{(t)}\right).$$

The proofs of (2), (3), and (4) are similar to that of (1). □

Theorem 3 shows the change of conditional probability for single-object variation of all $\underline{x}$ given $\overline{x} \in R_i^{(t+1)} \wedge \overline{x} \notin D_j^{(t+1)}$. Therefore, it lays a solid foundation for estimating changes of conditional probability if immigrated object belongs to equivalence class of considered conditions whereas does not belong to equivalence class of considered decisions.

Theorem 4. *Let $IS^{(t)} = \left\{U^{(t)}, C^{(t)} \cup D^{(t)}\right\}$ and $IS^{(t+1)} = \left\{U^{(t+1)}, C^{(t+1)} \cup D^{(t+1)}\right\}$ be an information system at time t and $t+1$ respectively. If variation of immigrated object $\overline{x}$ satisfies $\overline{x} \in R_i^{(t+1)} \wedge \overline{x} \in D_j^{(t+1)}$, variation trend of conditional probability w.r.t. variation of emigrated object are given as follows:*

(1) $\left(\underline{x} \notin R_i^{(t)} \wedge \underline{x} \notin D_j^{(t)}\right) \Rightarrow \Pr\left(D_j^{(t+1)} \middle| R_i^{(t+1)}\right) > \Pr\left(D_j^{(t)} \middle| R_i^{(t)}\right)$;

(2) $\left(\underline{x} \in R_i^{(t)} \wedge \underline{x} \notin D_j^{(t)}\right) \Rightarrow \Pr\left(D_j^{(t+1)} \middle| R_i^{(t+1)}\right) > \Pr\left(D_j^{(t)} \middle| R_i^{(t)}\right)$;

(3) $\left(\underline{x} \notin R_i^{(t)} \wedge \underline{x} \in D_j^{(t)}\right) \Rightarrow \Pr\left(D_j^{(t+1)} \middle| R_i^{(t+1)}\right) > \Pr\left(D_j^{(t)} \middle| R_i^{(t)}\right)$;

(4) $\left(\underline{x} \in R_i^{(t)} \wedge \underline{x} \in D_j^{(t)}\right) \Rightarrow \Pr\left(D_j^{(t+1)} \middle| R_i^{(t+1)}\right) = \Pr\left(D_j^{(t)} \middle| R_i^{(t)}\right)$.

Proof. (1) Since $\left(\overline{x} \in R_i^{(t+1)} \wedge \overline{x} \in D_j^{(t+1)}\right) \wedge \left(\underline{x} \notin R_i^{(t)} \wedge \underline{x} \notin D_j^{(t)}\right) \Rightarrow \left(R_i^{(t+1)} = \left\{R_i^{(t)} \cup \{\overline{x}\}\right\}\right) \wedge \left(D_j^{(t+1)} = \left\{D_j^{(t)} \cup \{\overline{x}\}\right\}\right)$, according to Definition 3, $\Pr\left(D_j{}^{(t+1)} \middle| R_i^{(t+1)}\right) = \frac{\left|D_j{}^{(t+1)} \cap R_i^{(t+1)}\right|}{\left|R_i^{(t+1)}\right|} = \frac{\left|\left\{D_j^{(t)} \cup \{\overline{x}\}\right\} \cap \left\{R_i^{(t)} \cup \{\overline{x}\}\right\}\right|}{\left|R_i^{(t)} \cup \{\overline{x}\}\right|} = \frac{\left|D_j^{(t)} \cap R_i^{(t)}\right| + 1}{\left|R_i^{(t)}\right| + 1} > \Pr\left(D_j{}^{(t)} \middle| R_i^{(t)}\right)$. Hence

$$\Pr\left(D_j{}^{(t+1)} \middle| R_i^{(t+1)}\right) > \Pr\left(D_j{}^{(t)} \middle| R_i^{(t)}\right).$$

The proofs of (2), (3), and (4) are similar to that of (1). □

Theorem 4 reveals the change of conditional probability for single-object variation of all $\underline{x}$ given $\overline{x} \in R_i^{(t+1)} \wedge \overline{x} \in D_j^{(t+1)}$. Therefore, it lays a solid foundation for estimating changes of conditional probability if immigrated object belong to both equivalence class of considered conditions and equivalent class of considered decisions.

By analyzing changes of conditional probability, Theorems 1–4 cover all possible variations of single-object that leads to the probable variation of conditional probability. Therefore, there is no need to discern the sequence of immigrated/emigrated object, which can simplify the algorithm design of real-time knowledge updating.

4.2. Stream computing learning algorithm towards three-way regions

Under the background of stream computing, three-way decisions theory can take full advantages of acquired knowledge so that new decisions can be immediately determined. This section attempts to deduce the transition of three-way region by exploring relation between conditional probability change and three-way region change. Based on Theorems 1–4, we summarize trend variation of conditional probability from time t to time $t+1$ and show it in Table 2.

It can be seen from Table 2 that the variation trend of conditional probability at time t and time $t+1$ is ascending, remains or descending. Grouped by the three possible results, we present strategies of region changes in Corollaries 1–3 correspondingly.

Corollary 1. *Let $IS^{(t)} = \left\{U^{(t)}, C^{(t)} \cup D^{(t)}\right\}$ and $IS^{(t+1)} = \left\{U^{(t+1)}, C^{(t+1)} \cup D^{(t+1)}\right\}$ represent an information system at time t and time $t+1$ respectively. If variation of single object is one of the following situations:*

(1) $\left(\overline{x} \notin R_i^{(t+1)} \wedge \overline{x} \notin D_j^{(t+1)}\right) \wedge \left(\underline{x} \notin R_i^{(t)} \wedge \underline{x} \notin D_j^{(t)}\right)$;

(2) $\left(\overline{x} \notin R_i^{(t+1)} \wedge \overline{x} \notin D_j^{(t+1)}\right) \wedge \left(\underline{x} \notin R_i^{(t)} \wedge \underline{x} \in D_j^{(t)}\right)$;

(3) $\left(\overline{x} \notin R_i^{(t+1)} \wedge \overline{x} \in D_j^{(t+1)}\right) \wedge \left(\underline{x} \notin R_i^{(t)} \wedge \underline{x} \notin D_j^{(t)}\right)$;

(4) $\left(\overline{x} \notin R_i^{(t+1)} \wedge \overline{x} \in D_j^{(t+1)}\right) \wedge \left(\underline{x} \notin R_i^{(t)} \wedge \underline{x} \in D_j^{(t)}\right)$;

(5) $\left(\overline{x} \in R_i^{(t+1)} \wedge \overline{x} \notin D_j^{(t+1)}\right) \wedge \left(\underline{x} \in R_i^{(t)} \wedge \underline{x} \notin D_j^{(t)}\right)$;

(6) $\left(\overline{x} \in R_i^{(t+1)} \wedge \overline{x} \in D_j^{(t+1)}\right) \wedge \left(\underline{x} \in R_i^{(t)} \wedge \underline{x} \in D_j^{(t)}\right)$.

then the positive region, boundary region and negative region of $IS^{(t+1)}$ is renewed as follows:

$$P: R_i^{(t)} \subseteq POS_{(\alpha,\beta)}\left(D_j^{(t)}\right) \Rightarrow POS_{(\alpha,\beta)}\left(D_j^{(t+1)}\right) = POS_{(\alpha,\beta)}\left(D_j^{(t)}\right);$$

$$B: R_i^{(t)} \subseteq BND_{(\alpha,\beta)}\left(D_j^{(t)}\right) \Rightarrow BND_{(\alpha,\beta)}\left(D_j^{(t+1)}\right) = BND_{(\alpha,\beta)}\left(D_j^{(t)}\right);$$

$$N: R_i^{(t)} \subseteq NEG_{(\alpha,\beta)}\left(D_j^{(t)}\right) \Rightarrow NEG_{(\alpha,\beta)}\left(D_j^{(t+1)}\right) = NEG_{(\alpha,\beta)}\left(D_j^{(t)}\right).$$

Proof. It can be directly deduced from Definition 4 that three-way regions at time $t+1$ is identical to that at time t. □

Corollary 2. *Let $IS^{(t)} = \left\{U^{(t)}, C^{(t)} \cup D^{(t)}\right\}$ and $IS^{(t+1)} = \left\{U^{(t+1)}, C^{(t+1)} \cup D^{(t+1)}\right\}$ represent an information system at time t and time $t+1$ respectively. If variation of single object is one of the following situations:*

(1) $\left(\overline{x} \notin R_i^{(t+1)} \wedge \overline{x} \notin D_j^{(t+1)}\right) \wedge \left(\underline{x} \in R_i^{(t)} \wedge \underline{x} \notin D_j^{(t)}\right)$;

(2) $\left(\overline{x} \notin R_i^{(t+1)} \wedge \overline{x} \in D_j^{(t+1)}\right) \wedge \left(\underline{x} \in R_i^{(t)} \wedge \underline{x} \notin D_j^{(t)}\right)$;

(3) $\left(\overline{x} \in R_i^{(t+1)} \wedge \overline{x} \in D_j^{(t+1)}\right) \wedge \left(\underline{x} \notin R_i^{(t)} \wedge \underline{x} \notin D_j^{(t)}\right)$;

Table 2
Updating patterns of the conditional probability for stream computing.

Theorem	*Patterns*		*Changes of equivalence classes*		$\Pr\left(D_j^{(t+1)} \mid R_i^{(t+1)}\right)$	Trend
	Immigrated object	*Emigrated object*	$R_i^{(t+1)}$	$D_j^{(t+1)}$		
1	$\overline{x} \notin R_i^{(t+1)} \wedge \overline{x} \notin D_j^{(t+1)}$	$\underline{x} \notin R_i^{(t)} \wedge \underline{x} \notin D_j^{(t)}$	$R_i^{(t)}$	$D_j^{(t)}$	$\frac{\left\|D_j^{(t)} \cap R_i^{(t)}\right\|}{\left\|R_i^{(t)}\right\|}$	Constant
		$\underline{x} \in R_i^{(t)} \wedge \underline{x} \notin D_j^{(t)}$	$R_i^{(t)} - \{\underline{x}\}$	$D_j^{(t)}$	$\frac{\left\|D_j^{(t)} \cap R_i^{(t)}\right\|}{\left\|R_i^{(t)}\right\|-1}$	Increase
		$\underline{x} \notin R_i^{(t)} \wedge \underline{x} \in D_j^{(t)}$	$R_i^{(t)}$	$D_j^{(t)} - \{\underline{x}\}$	$\frac{\left\|D_j^{(t)} \cap R_i^{(t)}\right\|}{\left\|R_i^{(t)}\right\|}$	Constant
		$\underline{x} \in R_i^{(t)} \wedge \underline{x} \in D_j^{(t)}$	$R_i^{(t)} - \{\underline{x}\}$	$D_j^{(t)} - \{\underline{x}\}$	$\frac{\left\|D_j^{(t)} \cap R_i^{(t)}\right\|-1}{\left\|R_i^{(t)}\right\|-1}$	Decrease
2	$\overline{x} \notin R_i^{(t+1)} \wedge \overline{x} \in D_j^{(t+1)}$	$\underline{x} \notin R_i^{(t)} \wedge \underline{x} \notin D_j^{(t)}$	$R_i^{(t)}$	$D_j^{(t)} \bigcup \{\overline{x}\}$	$\frac{\left\|D_j^{(t)} \cap R_i^{(t)}\right\|}{\left\|R_i^{(t)}\right\|}$	Constant
		$\underline{x} \in R_i^{(t)} \wedge \underline{x} \notin D_j^{(t)}$	$R_i^{(t)} - \{\underline{x}\}$	$D_j^{(t)} \bigcup \{\overline{x}\}$	$\frac{\left\|D_j^{(t)} \cap R_i^{(t)}\right\|}{\left\|R_i^{(t)}\right\|-1}$	Increase
		$\underline{x} \notin R_i^{(t)} \wedge \underline{x} \in D_j^{(t)}$	$R_i^{(t)}$	$D_j^{(t)} \bigcup \{\overline{x}\} - \{\underline{x}\}$	$\frac{\left\|D_j^{(t)} \cap R_i^{(t)}-1\right\|}{\left\|R_i^{(t)}\right\|-1}$	Constant
		$\underline{x} \in R_i^{(t)} \wedge \underline{x} \in D_j^{(t)}$	$R_i^{(t)} - \{\underline{x}\}$	$D_j^{(t)} \bigcup \{\overline{x}\} - \{\underline{x}\}$	$\frac{\left\|D_j^{(t)} \cap R_i^{(t)}\right\|-1}{\left\|R_i^{(t)}\right\|-1}$	Decrease
3	$\overline{x} \in R_i^{(t+1)} \wedge \overline{x} \notin D_j^{(t+1)}$	$\underline{x} \notin R_i^{(t)} \wedge \underline{x} \notin D_j^{(t)}$	$R_i^{(t)} \bigcup \{\overline{x}\}$	$D_j^{(t)}$	$\frac{\left\|D_j^{(t)} \cap R_i^{(t)}\right\|}{\left\|R_i^{(t)}\right\|+1}$	Decrease
		$\underline{x} \in R_i^{(t)} \wedge \underline{x} \notin D_j^{(t)}$	$R_i^{(t)} \bigcup \{\overline{x}\} - \{\underline{x}\}$	$D_j^{(t)}$	$\frac{\left\|D_j^{(t)} \cap R_i^{(t)}\right\|}{\left\|R_i^{(t)}-1\right\|}$	Constant
		$\underline{x} \notin R_i^{(t)} \wedge \underline{x} \in D_j^{(t)}$	$R_i^{(t)} \bigcup \{\overline{x}\}$	$D_j^{(t)} - \{\underline{x}\}$	$\frac{\left\|D_j^{(t)} \cap R_i^{(t)}\right\|}{\left\|R_i^{(t)}\right\|+1}$	Decrease
		$\underline{x} \in R_i^{(t)} \wedge \underline{x} \in D_j^{(t)}$	$R_i^{(t)} \bigcup \{\overline{x}\} - \{\underline{x}\}$	$D_j^{(t)} - \{\underline{x}\}$	$\frac{\left\|D_j^{(t)} \cap R_i^{(t)}\right\|-1}{\left\|R_i^{(t)}\right\|}$	Decrease
4	$\overline{x} \in R_i^{(t+1)} \wedge \overline{x} \in D_j^{(t+1)}$	$\underline{x} \notin R_i^{(t)} \wedge \underline{x} \notin D_j^{(t)}$	$R_i^{(t)} \bigcup \{\overline{x}\}$	$D_j^{(t)} \bigcup \{\overline{x}\}$	$\frac{\left\|D_j^{(t)} \cap R_i^{(t)}\right\|+1}{\left\|R_i^{(t)}\right\|+1}$	Increase
		$\underline{x} \in R_i^{(t)} \wedge \underline{x} \notin D_j^{(t)}$	$R_i^{(t)} \bigcup \{\overline{x}\} - \{\underline{x}\}$	$D_j^{(t)} \bigcup \{\overline{x}\}$	$\frac{\left\|D_j^{(t)} \cap R_i^{(t)}\right\|+1}{\left\|R_i^{(t)}\right\|}$	Increase
		$\underline{x} \notin R_i^{(t)} \wedge \underline{x} \in D_j^{(t)}$	$R_i^{(t)} \bigcup \{\overline{x}\}$	$D_j^{(t)} \bigcup \{\overline{x}\} - \{\underline{x}\}$	$\frac{\left\|D_j^{(t)} \cap R_i^{(t)}\right\|+1}{\left\|R_i^{(t)}\right\|+1}$	Increase
		$\underline{x} \in R_i^{(t)} \wedge \underline{x} \in D_j^{(t)}$	$R_i^{(t)} \bigcup \{\overline{x}\} - \{\underline{x}\}$	$D_j^{(t)} \bigcup \{\overline{x}\} - \{\underline{x}\}$	$\frac{\left\|D_j^{(t)} \cap R_i^{(t)}\right\|}{\left\|R_i^{(t)}\right\|}$	Constant

(4) $\left(\overline{x} \in R_i^{(t+1)} \wedge \overline{x} \in D_j^{(t+1)}\right) \wedge \left(\underline{x} \in R_i^{(t)} \wedge \underline{x} \notin D_j^{(t)}\right)$;

(5) $\left(\overline{x} \in R_i^{(t+1)} \wedge \overline{x} \in D_j^{(t+1)}\right) \wedge \left(\underline{x} \notin R_i^{(t)} \wedge \underline{x} \in D_j^{(t)}\right)$,

then the positive region, boundary region and negative region of $IS^{(t+1)}$ at time $t+1$ is renewed as follows:

$$POS_{(\alpha,\beta)}\left(D_j^{(t+1)}\right) = \begin{cases} POS_{(\alpha,\beta)}\left(D_j^{(t)}\right) - R_i^{(t)} \cup R_i^{(t+1)}, & p_1 \\ POS_{(\alpha,\beta)}\left(D_j^{(t)}\right) \cup R_i^{(t+1)}, & p_2 \\ POS_{(\alpha,\beta)}\left(D_j^{(t)}\right) \cup R_i^{(t+1)}, & p_3 \end{cases}$$

$$\begin{aligned} \textit{where} \quad & p_1 : R_i^{(t)} \subseteq POS_{(\alpha,\beta)}\left(D_j^{(t)}\right); \\ & p_2 : R_i^{(t)} \subseteq BND_{(\alpha,\beta)}\left(D_j^{(t)}\right) \wedge \Pr\left(D_j^{(t+1)} \mid R_i^{(t+1)}\right) \geq \alpha; \\ & p_3 : R_i^{(t)} \subseteq NEG_{(\alpha,\beta)}\left(D_j^{(t)}\right) \wedge \Pr\left(D_j^{(t+1)} \mid R_i^{(t+1)}\right) \geq \alpha. \end{aligned}$$

$$BND_{(\alpha,\beta)}\left(D_j^{(t+1)}\right)=\begin{cases}BND_{(\alpha,\beta)}\left(D_j^{(t)}\right)-R_i^{(t)}, & b_1\\ BND_{(\alpha,\beta)}\left(D_j^{(t)}\right)-R_i^{(t)}\cup R_i^{(t+1)}, & b_2\\ BND_{(\alpha,\beta)}\left(D_j^{(t)}\right)\cup R_i^{(t+1)}, & b_3\end{cases}$$

$$\begin{aligned}\textit{where}\quad & b_1: R_i^{(t)}\subseteq BND_{(\alpha,\beta)}\left(D_j^{(t)}\right)\wedge \Pr\left(D_j^{(t+1)}|R_i^{(t+1)}\right)\geq\alpha;\\ & b_2: R_i^{(t)}\subseteq BND_{(\alpha,\beta)}\left(D_j^{(t)}\right)\wedge \beta<\Pr\left(D_j^{(t+1)}|R_i^{(t+1)}\right)<\alpha;\\ & b_3: R_i^{(t)}\subseteq NEG_{(\alpha,\beta)}\left(D_j^{(t)}\right)\wedge \beta<\Pr\left(D_j^{(t+1)}|R_i^{(t+1)}\right)<\alpha.\end{aligned}$$

$$NEG_{(\alpha,\beta)}(D_j^{(t+1)})=\begin{cases}NEG_{(\alpha,\beta)}\left(D_j^{(t)}\right)-R_i^{(t)}, & n_1\\ NEG_{(\alpha,\beta)}\left(D_j^{(t)}\right)-R_i^{(t)}, & n_2\\ NEG_{(\alpha,\beta)}\left(D_j^{(t)}\right)-R_i^{(t)}\cup R_i^{(t+1)}, & n_3\end{cases}$$

$$\begin{aligned}\textit{where}\quad & n_1: R_i^{(t)}\subseteq NEG_{(\alpha,\beta)}\left(D_j^{(t)}\right)\wedge \Pr\left(D_j^{(t+1)}|R_i^{(t+1)}\right)\geq\alpha;\\ & n_2: R_i^{(t)}\subseteq NEG_{(\alpha,\beta)}\left(D_j^{(t)}\right)\wedge \alpha>\Pr\left(D_j^{(t+1)}|R_i^{(t+1)}\right)>\beta;\\ & n_3: R_i^{(t)}\subseteq NEG_{(\alpha,\beta)}\left(D_j^{(t)}\right)\wedge \Pr\left(D_j^{(t+1)}|R_i^{(t+1)}\right)\leq\beta.\end{aligned}$$

Proof. According to Theorems 1, 2 and 4, conditional probability $\Pr\left(D_j^{(t+1)}|R_i^{(t+1)}\right)$ increases, therefore, we have:

1) Updating of $POS_{(\alpha,\beta)}\left(D_j^{(t+1)}\right)$ at time $t+1$:

1.1) Given additional condition p_1 holds, based on Definition 4, we have $\Pr\left(D_j^{(t)}\middle|R_i^{(t)}\right)\geq\alpha$. Since

$$\Pr\left(D_j^{(t+1)}\middle|R_i^{(t+1)}\right)>\Pr\left(D_j^{(t)}\middle|R_i^{(t)}\right)>\alpha,\ R_i^{(t+1)}\subseteq POS_{(\alpha,\beta)}\left(D_j^{(t+1)}\right),$$

which means $R_i^{(t)}$ is an existing equivalence class in positive region at both time t and $t+1$ but the contained elements are changed. Hence

$$POS_{(\alpha,\beta)}\left(D_j^{(t+1)}\right)=POS_{(\alpha,\beta)}\left(D_j^{(t)}\right)-R_i^{(t)}\cup R_i^{(t+1)}.$$

1.2) Given additional condition p_2 holds, based on Definition 4, we have $\alpha>\Pr\left(D_j^{(t)}\middle|R_i^{(t)}\right)>\beta$. Since

$$\Pr\left(D_j^{(t+1)}\middle|R_i^{(t+1)}\right)\geq\alpha,\ R_i^{(t+1)}\subseteq POS_{(\alpha,\beta)}\left(D_j^{(t+1)}\right),$$

which means $R_i^{(t)}$ is an equivalence class transformed from boundary region at time t to positive region at time $t+1$. Hence

$$POS_{(\alpha,\beta)}\left(D_j^{(t+1)}\right)=POS_{(\alpha,\beta)}\left(D_j^{(t)}\right)\cup R_i^{(t+1)}.$$

1.3) Given additional condition p_3 holds, based on Definition 4, we have $\Pr\left(D_j^{(t)}\middle|R_i^{(t)}\right)\leq\beta$. Since

$$\Pr\left(D_j^{(t+1)}\middle|R_i^{(t+1)}\right)\geq\alpha,\ R_i^{(t+1)}\subseteq POS_{(\alpha,\beta)}\left(D_j^{(t+1)}\right),$$

which means $R_i^{(t)}$ is an equivalence class transformed from negative region at time t to positive region at time $t+1$. Hence

$$POS_{(\alpha,\beta)}\left(D_j^{(t+1)}\right)=POS_{(\alpha,\beta)}\left(D_j^{(t)}\right)\cup R_i^{(t+1)}.$$

2) Updating of $BND_{(\alpha,\beta)}\left(D_j^{(t+1)}\right)$ at time $t+1$:

2.1) Given additional condition b_1, based on Definition 4, we have $\alpha>\Pr\left(D_j^{(t)}\middle|R_i^{(t)}\right)>\beta$. Since

$$\Pr\left(D_j^{(t+1)} \middle| R_i^{(t+1)}\right) > \alpha,\ R_i^{(t+1)} \subseteq POS_{(\alpha,\beta)}\left(D_j^{(t+1)}\right),$$

which means $R_i^{(t)}$ is an equivalence class transformed from boundary region at time t to positive region at time $t+1$. Hence

$$BND_{(\alpha,\beta)}\left(D_j^{(t+1)}\right) = BND_{(\alpha,\beta)}\left(D_j^{(t)}\right) - R_i^{(t)}.$$

2.2) Given additional condition b_2, based on Definition 4, we have $\alpha > \Pr\left(D_j^{(t)} \middle| R_i^{(t)}\right) > \beta$. Since

$$\alpha > \Pr\left(D_j^{(t+1)} \middle| R_i^{(t+1)}\right) > \beta,\ R_i^{(t+1)} \subseteq POS_{(\alpha,\beta)}\left(D_j^{(t+1)}\right),$$

which means $R_i^{(t)}$ is an existing equivalence class in boundary region at both time t and $t+1$ but the contained elements are changed. Hence

$$BND_{(\alpha,\beta)}\left(D_j^{(t+1)}\right) = BND_{(\alpha,\beta)}\left(D_j^{(t)}\right) - R_i^{(t)} \cup R_i^{(t+1)}.$$

2.3) Given additional condition b_3, based on Definition 4, we have $\Pr\left(D_j^{(t)} \middle| R_i^{(t)}\right) < \beta$. Since

$$\alpha > \Pr\left(D_j^{(t+1)} \middle| R_i^{(t+1)}\right) > \beta,\ R_i^{(t+1)} \subseteq POS_{(\alpha,\beta)}\left(D_j^{(t+1)}\right),$$

which means $R_i^{(t)}$ is an equivalence class transformed from negative region at time t to boundary region at time $t+1$. Hence

$$BND_{(\alpha,\beta)}\left(D_j^{(t+1)}\right) = BND_{(\alpha,\beta)}\left(D_j^{(t)}\right) \cup R_i^{(t+1)}.$$

3) Updating of $NEG_{(\alpha,\beta)}\left(D_j^{(t+1)}\right)$ at time $t+1$:

3.1) Given additional condition n_1, based on Definition 4, we have $\Pr\left(D_j^{(t)} \middle| R_i^{(t)}\right) \geq \alpha$. Since

$$\Pr\left(D_j^{(t+1)} \middle| R_i^{(t+1)}\right) > \alpha,\ R_i^{(t+1)} \subseteq POS_{(\alpha,\beta)}\left(D_j^{(t+1)}\right),$$

which means $R_i^{(t)}$ is an equivalence class transformed from negative region at time t to positive region at time $t+1$. Hence

$$NEG_{(\alpha,\beta)}\left(D_j^{(t+1)}\right) = NEG_{(\alpha,\beta)}\left(D_j^{(t)}\right) - R_i^{(t)}.$$

3.2) Given additional condition n_2, based on Definition 4, we have $\alpha > \Pr\left(D_j^{(t)} \middle| R_i^{(t)}\right) > \beta$. Since

$$\Pr\left(D_j^{(t+1)} \middle| R_i^{(t+1)}\right) > \alpha,\ R_i^{(t+1)} \subseteq POS_{(\alpha,\beta)}\left(D_j^{(t+1)}\right),$$

which means $R_i^{(t)}$ is an equivalence class transformed from negative region at time t to boundary region at time $t+1$. Hence

$$NEG_{(\alpha,\beta)}\left(D_j^{(t+1)}\right) = NEG_{(\alpha,\beta)}\left(D_j^{(t)}\right) - R_i^{(t)}.$$

3.3) Given additional condition n_3, based on Definition 4, we have $\Pr\left(D_j^{(t)} \middle| R_i^{(t)}\right) < \beta$. Since

$$\Pr\left(D_j^{(t+1)} \middle| R_i^{(t+1)}\right) > \alpha,\ R_i^{(t+1)} \subseteq POS_{(\alpha,\beta)}\left(D_j^{(t+1)}\right),$$

which means $R_i^{(t)}$ is an existing equivalence class in negative region at both time t and $t+1$ but contained elements are changed. Hence

$$NEG_{(\alpha,\beta)}\left(D_j^{(t+1)}\right) = NEG_{(\alpha,\beta)}\left(D_j^{(t)}\right) - R_i^{(t)} \cup R_i^{(t+1)}. \qquad \square$$

Corollary 3. *Let $IS^{(t)} = \left\{U^{(t)}, C^{(t)} \cup D^{(t)}\right\}$ and $IS^{(t+1)} = \left\{U^{(t+1)}, C^{(t+1)} \cup D^{(t+1)}\right\}$ represent an information system at time t and time $t+1$ respectively. If variation of single object is one of the following situations:*

$$(1)\left(\overline{x} \notin R_i^{(t+1)} \wedge \overline{x} \notin D_j^{(t+1)}\right) \wedge \left(\underline{x} \in R_i^{(t)} \wedge \underline{x} \in D_j^{(t)}\right);$$
$$(2)\left(\overline{x} \notin R_i^{(t+1)} \wedge \overline{x} \in D_j^{(t+1)}\right) \wedge \left(\underline{x} \in R_i^{(t)} \wedge \underline{x} \in D_j^{(t)}\right);$$
$$(3)\left(\overline{x} \in R_i^{(t+1)} \wedge \overline{x} \notin D_j^{(t+1)}\right) \wedge \left(\underline{x} \notin R_i^{(t)} \wedge \underline{x} \notin D_j^{(t)}\right);$$
$$(4)\left(\overline{x} \in R_i^{(t+1)} \wedge \overline{x} \notin D_j^{(t+1)}\right) \wedge \left(\underline{x} \notin R_i^{(t)} \wedge \underline{x} \in D_j^{(t)}\right);$$
$$(5)\left(\overline{x} \in R_i^{(t+1)} \wedge \overline{x} \notin D_j^{(t+1)}\right) \wedge \left(\underline{x} \in R_i^{(t)} \wedge \underline{x} \in D_j^{(t)}\right).$$

then the positive region, boundary region and negative region of $IS^{(t+1)}$ *is renewed as follows:*

$$POS_{(\alpha,\beta)}\left(D_j^{(t+1)}\right) = \begin{cases} POS_{(\alpha,\beta)}\left(D_j^{(t)}\right) - R_i^{(t)} \cup R_i^{(t+1)}, & p_1 \\ POS_{(\alpha,\beta)}\left(D_j^{(t)}\right) - R_i^{(t)}, & p_2 \\ POS_{(\alpha,\beta)}\left(D_j^{(t)}\right) - R_i^{(t)}, & p_3 \end{cases}$$

$$\begin{aligned} \textit{where}\quad & p_1 : R_i^{(t)} \subseteq POS_{(\alpha,\beta)}\left(D_j^{(t)}\right) \wedge \Pr\left(D_j^{(t+1)}|R_i^{(t+1)}\right) \geq \alpha; \\ & p_2 : R_i^{(t)} \subseteq POS_{(\alpha,\beta)}\left(D_j^{(t)}\right) \wedge \beta < \Pr\left(D_j^{(t+1)}|R_i^{(t+1)}\right) < \alpha; \\ & p_3 : R_i^{(t)} \subseteq POS_{(\alpha,\beta)}\left(D_j^{(t)}\right) \wedge \Pr\left(D_j^{(t+1)}|R_i^{(t+1)}\right) \leq \beta. \end{aligned}$$

$$BND_{(\alpha,\beta)}\left(D_j^{(t+1)}\right) = \begin{cases} BND_{(\alpha,\beta)}\left(D_j^{(t)}\right) \cup R_i^{(t+1)}, & b_1 \\ BND_{(\alpha,\beta)}\left(D_j^{(t)}\right) - R_i^{(t)} \cup R_i^{(t+1)}, & b_2 \\ BND_{(\alpha,\beta)}\left(D_j^{(t)}\right) - R_i^{(t)}, & b_3 \end{cases}$$

$$\begin{aligned} \textit{where}\quad & b_1 : R_i^{(t)} \subseteq POS_{(\alpha,\beta)}\left(D_j^{(t)}\right) \wedge \beta < \Pr\left(D_j^{(t+1)}|R_i^{(t+1)}\right) < \alpha; \\ & b_2 : R_i^{(t)} \subseteq BND_{(\alpha,\beta)}\left(D_j^{(t)}\right) \wedge \beta < \Pr\left(D_j^{(t+1)}|R_i^{(t+1)}\right) < \alpha; \\ & b_3 : R_i^{(t)} \subseteq BND_{(\alpha,\beta)}\left(D_j^{(t)}\right) \wedge \Pr\left(D_j^{(t+1)}|R_i^{(t+1)}\right) \leq \beta. \end{aligned}$$

$$NEG_{(\alpha,\beta)}\left(D_j^{(t+1)}\right) = \begin{cases} NEG_{(\alpha,\beta)}\left(D_j^{(t)}\right) \cup R_i^{(t+1)}, & n_1 \\ NEG_{(\alpha,\beta)}\left(D_j^{(t)}\right) \cup R_i^{(t+1)}, & n_2 \\ NEG_{(\alpha,\beta)}\left(D_j^{(t)}\right) - R_i^{(t)} \cup R_i^{(t+1)}, & n_3 \end{cases}$$

$$\begin{aligned} \textit{where}\quad & n_1 : R_i^{(t)} \subseteq POS_{(\alpha,\beta)}\left(D_j^{(t)}\right) \wedge \Pr\left(D_j^{(t+1)}|R_i^{(t+1)}\right) \leq \beta; \\ & n_2 : R_i^{(t)} \subseteq BND_{(\alpha,\beta)}\left(D_j^{(t)}\right) \wedge \Pr\left(D_j^{(t+1)}|R_i^{(t+1)}\right) \leq \beta; \\ & n_3 : R_i^{(t)} \subseteq NEG_{(\alpha,\beta)}\left(D_j^{(t)}\right). \end{aligned}$$

Proof. This proof is similar to that of Corollary 2. □

Remark 1. From Corollaries 1–3, we can see that the result of three-way region variation may occur within region (for example, positive to positive) or without regions (for example, negative to boundary) depending on the relation of conditional probability with regard to predefined thresholds (α, β). While the intra-region variations do not alter the remaining regions, the inter-region variation lead to the expanding and diminishing of two regions simultaneously. By this design, we can make minimum computation in stream computing.

4.3. SS3WD: a new stream computing learning algorithm

On the basis of theorems and corollaries presented in Sections 4.1 and 4.2, this section will present an algorithm called Single-object Stream computing using Three-Way Decisions (SS3WD) to realize the stream computing learning method. Details of SS3WD are given as follows.

Algorithm 1. *SS3WD*

Input:

$IS^{(t)} = \{U^{(t)}, C^{(t)} \cup D^{(t)}\}$; $R_i^{(t)}$, $D_j^{(t)}$, $\Pr\left(D_j^{(t)}|R_i^{(t)}\right)$; $POS_{(\alpha,\beta)}\left(D_j^{(t)}\right)$, $BND_{(\alpha,\beta)}\left(D_j^{(t)}\right)$, $NEG_{(\alpha,\beta)}\left(D_j^{(t)}\right)$; (α, β); $|C_i|$, $|D_j|$ $\overline{x}$, *and* $\underline{x}$.

Output:

$POS_{(\alpha,\beta)}\left(D_j^{(t+1)}\right)$, $BND_{(\alpha,\beta)}\left(D_j^{(t+1)}\right)$, $NEG_{(\alpha,\beta)}\left(D_j^{(t+1)}\right)$.

Step 1: Find the equivalence class affiliation of immigrated and emigrated of objects, then update the equivalence classes $R_i^{(t+1)}$ *and* $D_j^{(t+1)}$ *at time* $t+1$.

Step 2: Estimate the trend of conditional probability $\Pr\left(D_j^{(t+1)}|R_i^{(t+1)}\right)$ *for each* i *and* j *according to Theorems 1–4 at time* $t+1$.

Step 3: Update the three-way decisions regions $POS_{(\alpha,\beta)}\left(D_j^{(t+1)}\right)$, $BND_{(\alpha,\beta)}\left(D_j^{(t+1)}\right)$, $NEG_{(\alpha,\beta)}\left(D_j^{(t+1)}\right)$ *respectively according to Corollaries 1–3.*

Step 4: Compute new conditional probability for every equivalence class w.r.t. every concept $(P(D_j|C_i))$ *at time* $t+1$ *and update* $|D_j|$, $|C_i|$. *Go back to Step 1.*

Remark 2. SS3WD is an implementation of stream mining learning method. Firstly, it is related to stream data mining. Secondly, the computations w.r.t. object variations are performed meanwhile.

Now we will analyze the computational complexity of SS3WD. Assuming that an information system IS contains m conditional equivalence classes $m = |U^{(t)}/A|$ $(U^{(t)} \subseteq U)$ and n decision equivalence classes $n = |D|$, in SS3WD requires $(m|C| + n)$ times at any time. Step 2 is the estimation of probabilistic trend, which only spends $O(1)$. Step 3 is the region re-allocation. Updating regions implies the transition of objects within regions. No renovation for regions is the best situation, and in this case complexity of both algorithm can be regarded as $O(1)$. In worst cases, it requires to be transferred two times. Let $|R_i|$ and $|R_j|$ stands for the cardinal of equivalence class i *w.r.t.* immigrated object $(\overline{x})$ and class j *w.r.t.* emigrated object $(\underline{x})$, the time cost for region adjustment is denoted as $O(|R_i| + |R_j|)$. Step 4 requires $O(1)$ since new conditional probability can be computed quickly given that $|R_j|$ and $|D_i|$ is known. Time cost in step 2 and step 4 is negligible. Finally, the overall complexity of SS3WD is at most $O(|U^{(t)}/C||C| + |D|) + O(|R_i| + |R_j|)$.

Meanwhile, an incremental learning algorithm called Incremental Learning with Three-Way Decisions (IL3WD) is also designed based on [26,29]. It is a hybrid incremental learning algorithm which maintain decision knowledge w.r.t. object addition and deletion incrementally. Details of IL3WD are given as follows.

Algorithm 2. *IL3WD*

Input:

$IS^{(t)} = \{U^{(t)}, C^{(t)} \cup D^{(t)}\}$; $R_i^{(t)}$, $D_j^{(t)}$, $\Pr\left(D_j^{(t)}|R_i^{(t)}\right)$; $POS_{(\alpha,\beta)}\left(D_j^{(t)}\right)$, $BND_{(\alpha,\beta)}\left(D_j^{(t)}\right)$, $NEG_{(\alpha,\beta)}\left(D_j^{(t)}\right)$; (α, β); $|C_i|$, $|D_j|$ $\overline{x}$, *and* $\underline{x}$.

Output:

$POS_{(\alpha,\beta)}\left(D_j^{(t+1)}\right)$, $BND_{(\alpha,\beta)}\left(D_j^{(t+1)}\right)$, $NEG_{(\alpha,\beta)}\left(D_j^{(t+1)}\right)$.

Step 1: Update three-way region at time $t+1$ *given immigrated object* $\overline{x}$.

Step 1.1: Find the equivalence class affiliation of immigrated object, then update the two equivalence classes $R_i^{(t+1)}$ *and* $D_j^{(t+1)}$ *at time* $t+1$.

Step 1.2: Estimate the trend of conditional probability $\Pr\left(D_j^{(t+1)}|R_i^{(t+1)}\right)$ *for each* i *and* j *according to Refs. [26,29] at time* $t+1$.

Step 1.3: Update the three-way decisions regions $POS_{(\alpha,\beta)}\left(D_j^{(t+1)}\right)$, $BND_{(\alpha,\beta)}\left(D_j^{(t+1)}\right)$, $NEG_{(\alpha,\beta)}\left(D_j^{(t+1)}\right)$ *respectively according to Refs. [26,29].*

Step 1.4: Compute new conditional probability for every equivalence class w.r.t. every concept $(\Pr(D_j|C_i))$ *at time t+1 and update* $|D_j|$, $|C_i|$.

Step 2: Update three-way region at time $t+1$ *given emigrated object* $\underline{x}$.

Step 2.1: Find the equivalence class affiliation of emigrated object, then update the two equivalence classes $R_i^{(t+1)}$ *and* $D_j^{(t+1)}$ *at time* $t+1$.

Step 2.2: Estimate the trend of conditional probability $\Pr\left(D_j^{(t+1)}|R_i^{(t+1)}\right)$ *for each* i *and* j *in accordance with Refs. [26,29] at time* $t+1$.

Step 2.3: Update the three-way decisions regions $POS_{(\alpha,\beta)}\left(D_j^{(t+1)}\right)$, $BND_{(\alpha,\beta)}\left(D_j^{(t+1)}\right)$, $NEG_{(\alpha,\beta)}\left(D_j^{(t+1)}\right)$ *respectively according to Refs. [26,29].*

Step 2.4: Compute new conditional probability for every equivalence class w.r.t. every concept ($P(D_j|C_i)$) *at time* $t+1$ *and update* $|D_j|$, $|C_i|$.

Step 3: Go back to Step 1.

Now we will analyze the computational complexity of IL3WD. Assuming that an information system IS contains m conditional equivalence classes $m=|U^{(t)}/C|$ $(U^{(t)}\subseteq U)$ and n decision equivalence classes $n=|D|$, in steps 1.1 and 2.1 of IL3WD, the three-way decision-making needs to calculate the conditional probability for $(m|C|+n)$ times each, denoted as $O(2m|C|+2n)$ altogether. This lies in the fact that every variation, whether immigration or emigration, is separately processed. Steps 1.2 and 2.2 are the estimation of probabilistic trend, which only spends $O(1)$. Steps 1.3 and 2.3 are the region re-allocation. Updating regions implies the transition of objects within regions. No renovation for regions is the best situation, and in this case complexity of both algorithm can be regarded as $O(1)$. In worst cases, it requires to be transferred two times (no matter immigration or emigration first). Let $|R_i|$ and $|R_j|$ stands for the cardinal of equivalence class i *w.r.t.* immigrated object ($\overline{x}$) and class j *w.r.t.* emigrated object ($\underline{x}$), the time cost for region adjustment is denoted as $O(|R_i|+|R_j|)$. Finally, the overall complexity of IL3WD is at most $O(2|U^{(t)}/C||C|+2|D|)+O(|R_i|+|R_j|)$;

SS3WD and IL3WD is analogous in many aspects. However, time cost of step 3 in SS3WD and IL3WD is not invariantly identical. A special case is that time cost in step 3 of SS3WD can be significantly economical than IL3WD if the following conditions are sufficed simultaneously: i) the conditional probability of certain equivalence class is approximate to one threshold; ii) at least condition attribute in $\underline{x}$ and $\overline{x}$ are identical of that equivalence class. In this case, IL3WD will have to adjust twice whereas SS3WD require at most one adjustment. Step 4 requires $O(1)$ since new conditional probability can be computed quickly given that $|R_j|$ and $|D_i|$ is known. Time cost in step 2 and step 4 is negligible. Whether step 1 or step 3 spends more time is closely pertinent to the dataset. However, the uncertainty in time cost does not hamper us to conclude that theoretically, SS3WD is expected to be faster than IL3WD. This is straightforward since the time cost of SS3WD is no worse than IL3WD. Finally, we have following assertions:

i) If step 1 dominates the main time cost of SS3WD (which means step 1.1 and 2.1 dominates the time cost of IL3WD) and no special case of step 3 occurs (means time cost in step 3 of SS3WD is equivalent to that in step 1.3 and step 2.3 of IL3WD), speedup will be close to 2;

ii) If special case of step 3 in SS3WD occurs, the speedup can be almost infinite;

iii) If step 3 dominates the main time cost of SS3WD, speedup will be close to 1.

We can trace the change of decision regions at different stages according to the variation of objects according to this algorithm, as shown in Fig. 3.

The SS3WD algorithm reflects the three-way decision-based stream computing process of arbitrary two adjacent time $(t,t+1)$. As shown in the Fig. 3, thick solid arrows represent the structure process of 3WD, which are divided into three steps: firstly, streaming updating the conditional equivalence classes and decision equivalence class; secondly, streaming updating the conditional probability; thirdly, streaming updating the three-way decisions based on the decision equivalence classes and the value of conditional probability. In order to reduce the redundant calculation as much as possible, the structure of 3WD at time $t+1$ utilize information stored at time t (indicated by virtual arrow) and changes of equivalence class at $t+1$ (indicated by solid arrows) in real time. For situations either immigration count or emigration count exceeds one, SS3WD is supposed to perform more times to depict knowledge variation between time t and time $t+1$.

5. Examples and experiments

This section attempts to apply SS3WD for stream computing. Firstly, an example are presented to validate its correctness and feasibility step by step. We then compare SS3WD with IL3WD on nine UCI datasets with different sizes of objects and attributes. Finally, we present the results of experiments from three different perspectives.

5.1. Example

Given an information system $IS^{(t)}=\{U^{(t)},C^{(t)}\cup D^{(t)}\}$, as shown in Table 3. where $U^{(t)}=\{x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8\}$ represents a set of data objects at time t, $C^{(t)}=\{c_1,c_2,c_3,c_4\}$ represents a set of conditional attribute, $D^{(t)}=\{d\}$ presents a set of decision attribute.

The corresponding structured memory information of Table 3 in timestamp t are stored as follows:

- equivalence class of condition: $U^{(t)}/C^{(t)}=\{R_1^{(t)},R_2^{(t)},R_3^{(t)}\}$, where $R_1^{(t)}=\{x_1,x_3,x_4\}$, $R_2^{(t)}=\{x_2,x_5\}$, $R_3^{(t)}=\{x_6,x_7,x_8\}$.
- equivalence class of decision: $U^{(t)}/D^{(t)}=\{D_1^{(t)},D_2^{(t)}\}$, where $D_1^{(t)}=\{x_1,x_3,x_6,x_7\}$, $D_2^{(t)}=\{x_2,x_4,x_5,x_8\}$.
- Conditional probabilities:

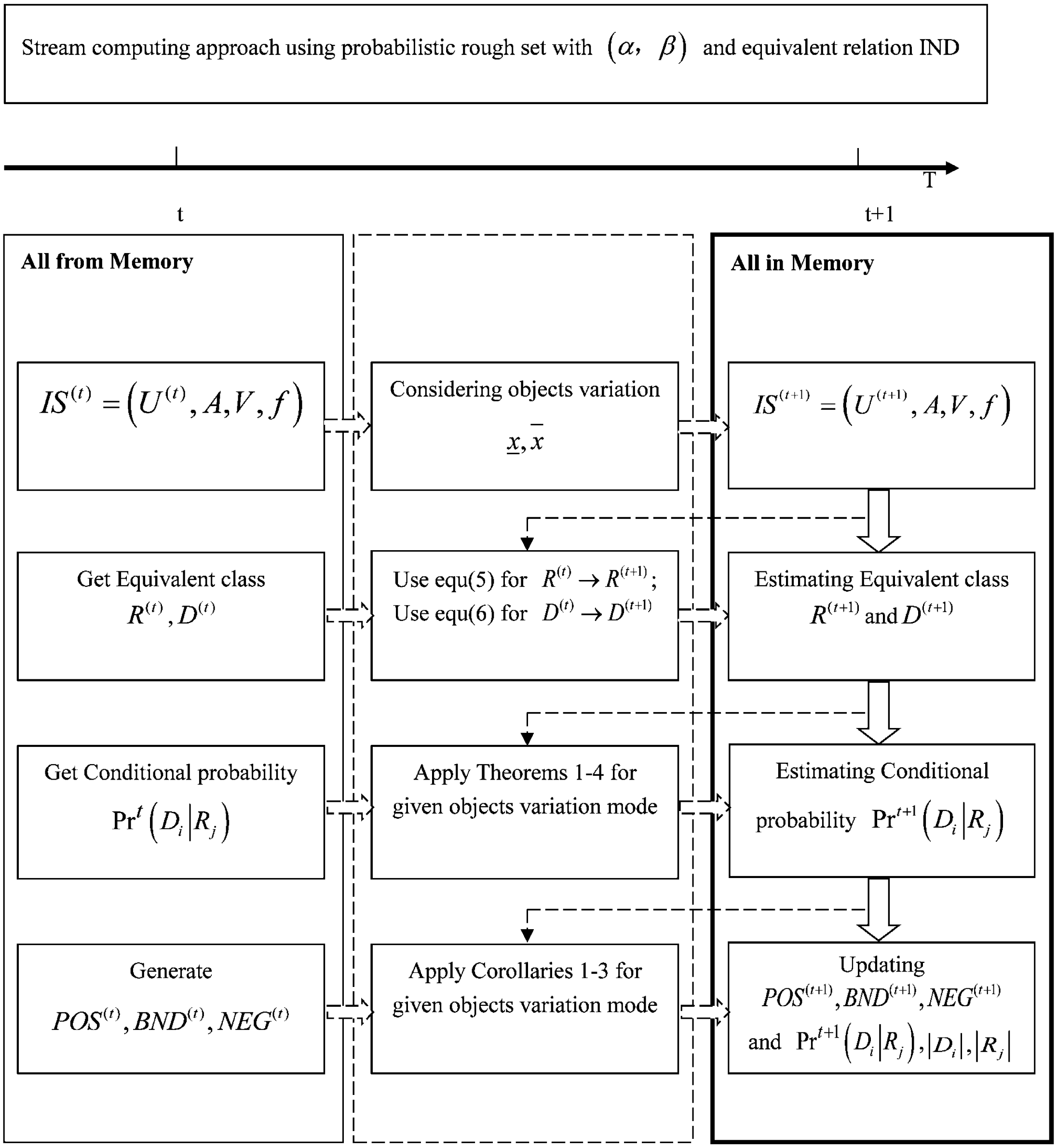

Fig. 3. Description of SS3WD Algorithm.

Table 3
The original information system at time t.

U	c_1	c_2	c_3	c_4	d	U	c_1	c_2	c_3	c_4	d
x_1	1	0	0	1	1	x_5	1	1	0	0	0
x_2	1	1	0	0	0	x_6	1	1	1	1	1
x_3	1	0	0	1	1	x_7	1	1	1	1	1
x_4	1	0	0	1	0	x_8	1	1	1	1	0

$$D_1^{(t)} : \Pr\left(D_1^{(t)}|R_1^{(t)}\right) = 2/3, \Pr\left(D_1^{(t)}|R_2^{(t)}\right) = 0, \Pr\left(D_1^{(t)}|R_3^{(t)}\right) = 2/3;$$
$$D_2^{(t)} : \Pr\left(D_2^{(t)}|R_1^{(t)}\right) = 1/3, \Pr\left(D_2^{(t)}|R_2^{(t)}\right) = 1, \Pr\left(D_2^{(t)}|R_3^{(t)}\right) = 1/3;$$

- Thresholds:$(\alpha, \beta) = (0.75, 0.35)$.
- Three-way decision regions of D_1 D_2:

Table 4
The information system $U^{t+1} : U^{t+1} = U^t \cup \{x_9\} - \{x_1\}$.

U	c_1	c_2	c_3	c_4	d	U	c_1	c_2	c_3	c_4	d
x_2	1	1	0	0	0	x_6	1	1	1	1	1
x_3	1	0	0	1	1	x_7	1	1	1	1	1
x_4	1	0	0	1	0	x_8	1	1	1	1	0
x_5	1	1	0	0	0	x_9	1	1	1	1	1

$$D_1^{(t)} : \begin{cases} POS_{(\alpha,\beta)}\left(D_1^{(t)}\right) = \emptyset; \\ BND_{(\alpha,\beta)}\left(D_1^{(t)}\right) = R_1^{(t)} \cup R_3^{(t)}; \\ NEG_{(\alpha,\beta)}\left(D_1^{(t)}\right) = R_2^{(t)}. \end{cases} \quad \text{and} \quad D_2^{(t)} : \begin{cases} POS_{(\alpha,\beta)}\left(D_2^{(t)}\right) = R_2^{(t)}; \\ BND_{(\alpha,\beta)}\left(D_2^{(t)}\right) = \emptyset; \\ NEG_{(\alpha,\beta)}\left(D_2^{(t)}\right) = R_1^{(t)} \cup R_3^{(t)}. \end{cases}.$$

Consequently, we can make decisions of acceptance, deferment and rejection at time t, respectively according to Definition 5.

In what follows, we propose two different examples of the data stream objects to illustrate the real time stream computing process of three-way decisions.

Example. We assume at time $t+1$, the object x_9 is immigrated, whereas the object x_1 is emigrated, as shown in Table 4.

Step 1 of SS3WD

For the equivalence classes induced by $C^{(t)}$ and $D^{(t)}$, we have the knowledge of equivalence classes as following:

$R_1^{(t+1)} = R_1^{(t)} - \{\underline{x}\} = \{x_3, x_4\}$;
$R_2^{(t+1)} = R_2^{(t)} = \{x_2, x_5\}$;
$R_3^{(t+1)} = R_3^{(t)} \cup \{\overline{x}\} = \{x_6, x_7, x_8, x_9\}$;
$D_1^{(t+1)} = D_1^{(t)} \cup \{\overline{x}\} - \{\underline{x}\} = \{x_3, x_6, x_7, x_9\}$;
$D_2^{(t+1)} = D_2^{(t)} = \{x_2, x_4, x_5, x_8\}$.

As a result, $\left(\overline{x} \in R_3^{(t+1)} \wedge \overline{x} \in D_1^{(t+1)}\right) \wedge \left(\underline{x} \in R_1^{(t)} \wedge \underline{x} \in D_1^{(t)}\right)$.

Step 2 of SS3WD

- Update the three-way decisions of $D_1^{(t+1)}$:
  We can estimate the trend of conditional probability as follows:

$$\Pr\left(D_1^{(t+1)} \middle| R_1^{(t+1)}\right) < \Pr\left(D_1^{(t)} \middle| R_1^{(t)}\right); \quad (Theorem \quad 2(4))$$
$$\Pr\left(D_1^{(t+1)} \middle| R_2^{(t+1)}\right) = \Pr\left(D_1^{(t)} \middle| R_2^{(t)}\right); \quad (Theorem \quad 2(3))$$
$$\Pr\left(D_1^{(t+1)} \middle| R_3^{(t+1)}\right) > \Pr\left(D_1^{(t)} \middle| R_3^{(t)}\right). \quad (Theorem \quad 4(3))$$

- Update the three-way decisions of $D_2^{(t+1)}$:
  We can also estimate the trend of conditional probability as follows:

$$\Pr\left(D_2^{(t+1)} \middle| R_1^{(t+1)}\right) > \Pr\left(D_2^{(t)} \middle| R_1^{(t)}\right); \quad (Theorem \quad 1(2))$$
$$\Pr\left(D_2^{(t+1)} \middle| R_2^{(t+1)}\right) = \Pr\left(D_2^{(t)} \middle| R_2^{(t)}\right); \quad (Theorem \quad 1(1))$$
$$\Pr\left(D_2^{(t+1)} \middle| R_3^{(t+1)}\right) < \Pr\left(D_2^{(t)} \middle| R_3^{(t)}\right). \quad (Theorem \quad 3(1))$$

Step 3 of SS3WD

- Firstly, we will find the elements of $D_1^{(t+1)}$.
  According to Corollary 2, since $R_1^{(t)} \subseteq BND_{(\alpha,\beta)}\left(D_1^{(t)}\right)$ and $\beta < \Pr\left(D_1{}^{(t+1)} \middle| R_1{}^{(t+1)}\right) = 1/2 < \alpha$, we have $BND_{(\alpha,\beta)}(D_1^{(t+1)}) = BND_{(\alpha,\beta)}\left(D_1^{(t)}\right) - \{\underline{x}\} = R_1^{(t)} \cup R_3^{(t)} - \{\underline{x}\} = R_1^{(t+1)} \cup R_3^{(t)}$. According to Corollary 3, since $R_3^{(t)} \subseteq$

$BND_{(\alpha,\beta)}\left(D_1^{(t)}\right)$ and $\Pr\left(D_1{}^{(t+1)}\middle|R_3{}^{(t+1)}\right) = 3/4 \geq \alpha$, we have $POS_{(\alpha,\beta)}\left(D_1^{(t+1)}\right) = POS_{(\alpha,\beta)}\left(D_1^{(t)}\right) \cup R_3^{(t+1)} = R_3^{(t+1)}$, $BND_{(\alpha,\beta)}\left(D_1^{(t+1)}\right) = BND_{(\alpha,\beta)}\left(D_1^{(t)}\right) - R_3^{(t)} = R_1^{(t+1)} \cup R_3^{(t)} - R_3^{(t)} = R_1^{(t+1)}$.

- Secondly, we will find the elements of $D_2^{(t+1)}$.
  According to Corollary 3, since $R_1^{(t)} \subseteq NEG_{(\alpha,\beta)}\left(D_2^{(t)}\right)$ and $\beta < \Pr\left(D_2{}^{(t+1)}\middle|R_1{}^{(t+1)}\right) = 1/2 < \alpha$, we have $BND_{(\alpha,\beta)}\left(D_2^{(t+1)}\right) = BND_{(\alpha,\beta)}\left(D_2^{(t)}\right) \cup R_1^{(t+1)} = R_1^{(t+1)}$, $NEG_{(\alpha,\beta)}\left(D_2^{(t+1)}\right) = NEG_{(\alpha,\beta)}\left(D_2^{(t)}\right) - R_1^{(t)} = R_3^{(t)}$. According to Corollary 2, since $R_3^{(t)} \subseteq NEG_{(\alpha,\beta)}\left(D_2^{(t)}\right)$ and $\Pr\left(D_2{}^{(t+1)}\middle|R_3{}^{(t+1)}\right) = 1/4 < \beta$, we have $NEG_{(\alpha,\beta)}\left(D_2^{(t+1)}\right) = NEG_{(\alpha,\beta)}\left(D_2^{(t)}\right) \cup \{\overline{x}\} = R_3^{(t)} \cup \{\overline{x}\} = R_3^{(t+1)}$.
  In summary, the three-way decision regions are updated by real time stream computing as follows:

$$D_1^{(t+1)}: \begin{cases} POS_{(\alpha,\beta)}\left(D_1^{(t+1)}\right) = R_3^{(t+1)}; \\ BND_{(\alpha,\beta)}\left(D_1^{(t+1)}\right) = R_1^{(t+1)}; \\ NEG_{(\alpha,\beta)}\left(D_1^{(t+1)}\right) = R_2^{(t+1)}. \end{cases} \quad \text{and} \quad D_2^{(t+1)}: \begin{cases} POS_{(\alpha,\beta)}\left(D_2^{(t+1)}\right) = R_2^{(t+1)}; \\ BND_{(\alpha,\beta)}\left(D_2^{(t+1)}\right) = R_1^{(t+1)}; \\ NEG_{(\alpha,\beta)}\left(D_2^{(t+1)}\right) = R_3^{(t+1)}. \end{cases}$$

Step 4 of SS3WD

Accordingly, at time $t+1$ in Example 1, we can make three-way decisions of acceptance, deferment and rejection, respectively according to Definition 5. Meanwhile, we need to update the conditional probability of R_1 and R_3 because of object variation. Since equivalence class of C_2 remains unchanged, we will update the conditional probability of D_i, $i = 1, 2$ w.r.t. R_j, $j = 1, 3$

$$\Pr(D_1^{(t+1)}|R_1^{(t+1)}) = (2-1)/(3-1) = 1/2; \Pr(D_2^{(t+1)}|R_1^{(t+1)}) = 1/(3-1) = 1/2;$$
$$\Pr(D_1^{(t+1)}|R_3^{(t+1)}) = (2+1)/(3+1) = 3/4; \Pr(D_2^{(t+1)}|R_1^{(t+1)}) = 1/(3+1) = 1/4;$$

From the above-mentioned example, we can observe that it is possible to immediately estimate the trend of conditional probability and compute the three-way decisions given variation of the objects. Hence, the stream computing update theory and algorithm are proved to be effective.

5.2. Experimental analysis

In this section, we conduct experiments to verify the efficiency and robustness of SS3WD. To accomplish it, we design altogether three experiments. The first experiment endeavored to demonstrate the stream computing efficiency, whereas the second managed to testify the performance on real big dataset. The third experiment, however, attempted to seek the relationship between thresholds setting and computation cost. All experiments are performed on a computer with Intel® Core™ 2 Duo CPU E7500 and 4 GB of memory, running Microsoft Windows 7. Methods are programmed in Java with MyEclipse and JDK 1.6.0. We select eight datasets from the University of California Irvine (UCI) Machine Learning Database Repository (http://archive.ics.uci.edu/ml/datasets) as benchmarks to assess the computational cost respectively. They are "Skin–NoSkin", "Letter", "Magic", "Shuttle", "IRIS", "Zoo", "Haberman", and "Breast-cancer". Detailed characteristics of the eight datasets are shown in Table 5. For datasets which contains numerical attribute, we simply divide them into 10 equal bins so that equivalence class can be generated.

Experiment 1. *Comparison of computational time on small dataset.*

In the first experiment, we intend to show the overall performance of SS3WD *w.r.t.* IL3WD by using the metric of accumulated time. To reduce the impact of accidental factors, all experiments are executed 10 times. As for thresholds (α, β), we arbitrarily assign them as $(\alpha, \beta) = (0, 75, 0.3)$ in the first two experiments so that three disjoint regions can be determined.

Assuming that the memory space can keep at most 100 records, we examine the performance of both algorithms as gradually increasing the proportion of data stored at time t for aforementioned eight datasets. To achieve it, we select ten different sizes of records $(10, 20, 30, \cdots, 100)$. Experimental results of two algorithms for updating three-way decision regions are shown in Fig. 4(a)–(h) respectively. The time in Fig. 4 includes the effort of finding variations in equivalence class of both conditional attributes and decision classes, estimating changes of conditional probability and updating three-way decisions for all possible cases, whereas time for calculating original equivalence class is excluded.

It can be observed from Fig. 4 that both algorithms are almost linearly increased as objects continuously processed in memory. SS3WD always performs faster than IL3WD. However, the gradient is much different and SS3WD is much smaller than IL3WD. Algorithm speedup is another essential metric to evaluate the efficiency in previous work concerning incremental learning [29]. We present an analytical comparison between the SS3WD and Luo [29]. The speedup is defined

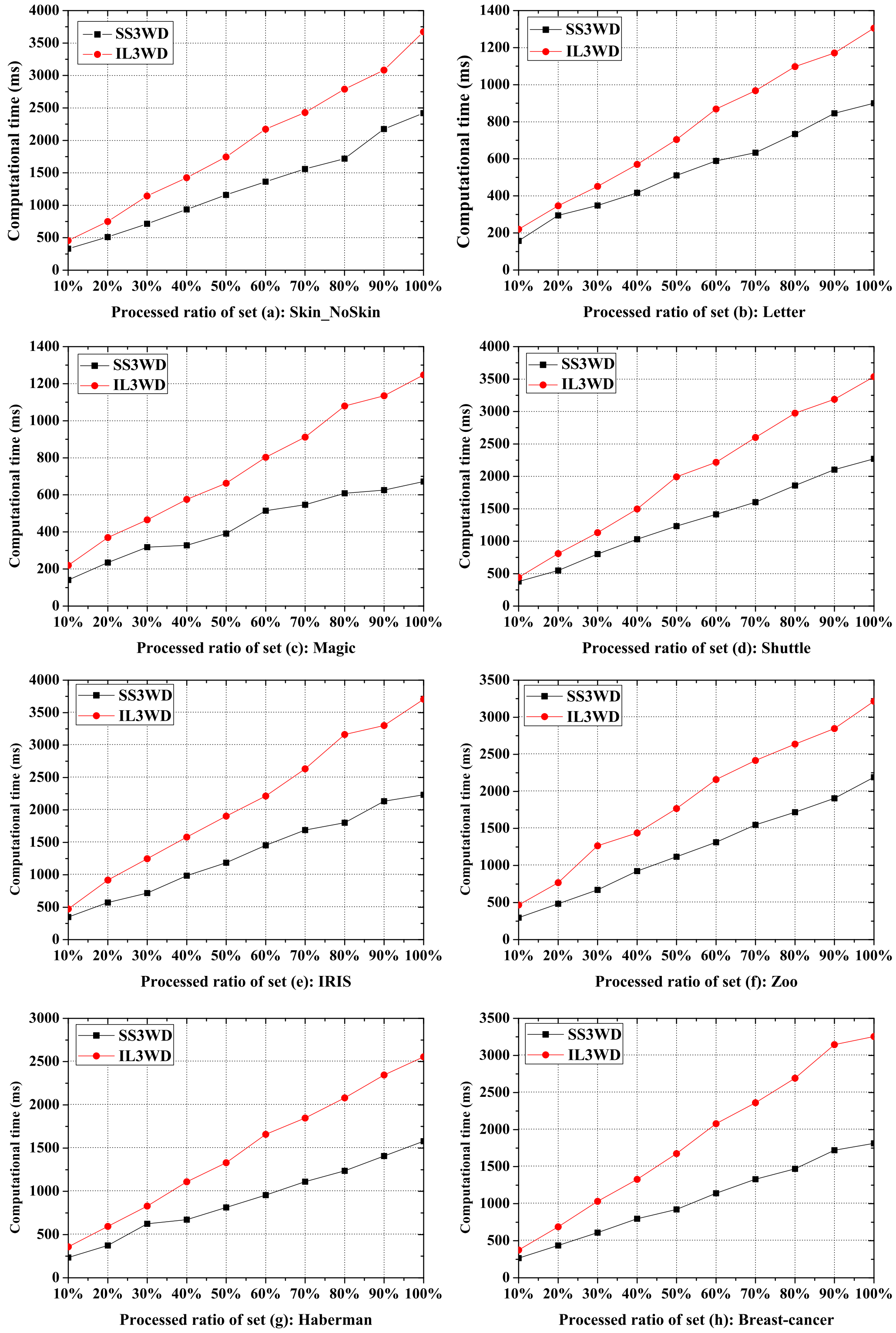

Fig. 4. Elapsed time between IL3WD and SS3WD on eight UCI datasets.

Table 5
The basic information of the eight datasets.

No.	Data sets name	Samples	Features	Concepts count
1	Skin–NoSkin	64486	3	2
2	Letter	20000	16	12
3	Magic	19020	10	2
4	Shuttle	58000	9	5
5	IRIS	60750	8	3
6	Zoo	60600	17	7
7	Haberman	55080	3	2
8	Breast-cancer	60860	10	4

Table 6
Speedup of SS3WD versus IL3WD in eight UCI datasets.

DataSets	Size of processed ratio									
	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Skin_NoSkin	1.378	1.466	1.597	1.521	1.503	1.594	1.558	1.624	1.382	1.518
Letter	1.401	1.716	1.296	1.367	1.378	1.475	1.529	1.496	1.392	1.456
Magic	1.563	1.574	1.462	1.756	1.696	1.559	1.667	1.773	1.550	1.857
Shuttle	1.163	1.478	1.411	1.452	1.616	1.567	1.622	1.601	1.175	1.557
IRIS	1.362	1.606	1.739	1.603	1.603	1.522	1.557	1.754	1.362	1.667
Zoo	1.576	1.585	1.884	1.557	1.581	1.644	1.562	1.534	1.560	1.468
Haberman	1.528	1.584	1.328	1.652	1.636	1.732	1.662	1.682	1.512	1.618
Breast-cancer	1.418	1.572	1.695	1.665	1.815	1.823	1.776	1.831	1.432	1.794

Table 7
Experimental results examined from time t to $t+5$ on eight UCI datasets.

Data set	Algorithm	t	t_1	t_2	t_3	t_4	t_5	mean±std	p-value
Skin_NoSkin	IL3WD	0.076	0.062	0.063	0.059	0.058	0.060	0.063 ± 0.007	0.0022
	SS3WD	0.055	0.043	0.040	0.039	0.039	0.038	0.042 ± 0.006	
Letter	IL3WD	0.110	0.087	0.075	0.071	0.070	0.072	0.081 ± 0.015	0.0931
	SS3WD	0.079	0.074	0.058	0.052	0.051	0.049	0.061 ± 0.013	
Magic	IL3WD	0.110	0.093	0.078	0.072	0.066	0.067	0.081 ± 0.017	0.0173
	SS3WD	0.071	0.059	0.053	0.041	0.039	0.043	0.053 ± 0.012	
Shuttle	IL3WD	0.074	0.068	0.063	0.062	0.066	0.062	0.066 ± 0.005	0.0108
	SS3WD	0.063	0.046	0.045	0.043	0.041	0.039	0.046 ± 0.009	
IRIS	IL3WD	0.079	0.076	0.069	0.062	0.066	0.062	0.066 ± 0.005	0.0079
	SS3WD	0.058	0.048	0.040	0.041	0.040	0.040	0.044 ± 0.007	
Zoo	IL3WD	0.078	0.064	0.070	0.060	0.059	0.060	0.065 ± 0.008	0.0079
	SS3WD	0.050	0.040	0.037	0.039	0.037	0.036	0.040 ± 0.005	
Haberman	IL3WD	0.065	0.054	0.050	0.050	0.048	0.050	0.053 ± 0.006	0.0043
	SS3WD	0.043	0.034	0.038	0.031	0.030	0.029	0.034 ± 0.005	
Breast-cancer	IL3WD	0.063	0.057	0.057	0.055	0.056	0.058	0.058 ± 0.003	0.0079
	SS3WD	0.044	0.036	0.034	0.033	0.031	0.032	0.035 ± 0.005	

as $\frac{T_i}{T_s}$, with T_s be the computation time of the SS3WD and T_i be the computation time of the IL3WD. From Table 6 we can see that 1 and 2 is the lower bound and upper bound of speedup, and it is coincide with the analysis of algorithm complexity.

Additionally, average time differences reach maximal when 90 percent of data stored in memory. To better demonstrate the superiority of SS3WD, we check the time cost one by one starting from the time when 90 percent of data is processed. We run this experiment ranges from time t to $t+5$ to test its stability. Results of IL3WD and SS3WD are described in Table 7. The values before "±" presents the average time to execute either IL3WD or "SS3WD algorithm" on given dataset, whereas the values after "±" gives the standard deviation of executing time. The unit of average time is measured by 10^{-3} seconds.

It can be observed that almost half time is reduced as compared to IL3WD. The Wilcoxon signed-rank test conducted at a 5% significance level in eight UCI datasets reveals that the acceleration is statistically significant. Therefore, SS3WD is particularly efficient for knowledge extraction of massive data. Moreover, the smaller variance to average value reveals in different datasets reveals that SS3WD is quite robust and scalable.

Experiment 2. *Comparison of computational time in big dataset.*

To verify the performance of the proposed algorithm SS3WD, we select a real big data set from UCI (http://archive.ics.uci.edu/ml/machine-learning-databases/kddcup99-mld/kddcup99.html), which contains nearly 5 million records, 1 decision

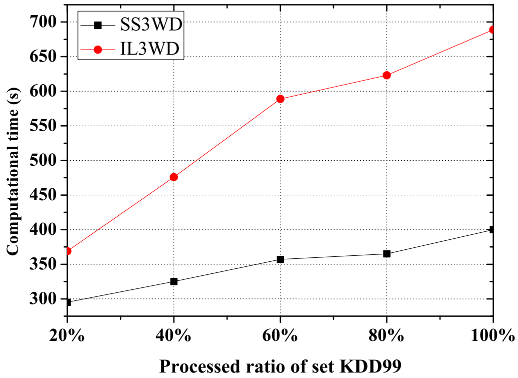

Fig. 5. The average elapsed times between IL3WD and SS3WD on KDD99.

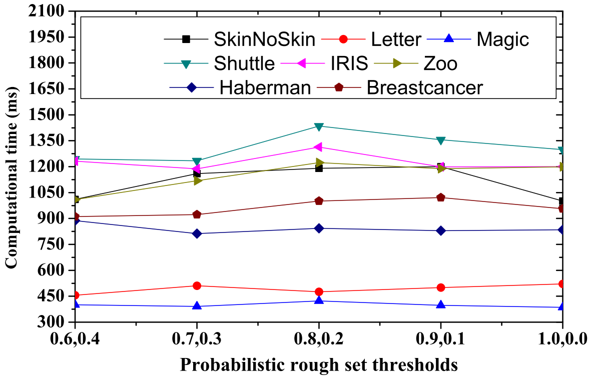

Fig. 6. Time cost of different (α, β) with SS3WD method on UCI dataset.

attributes and 41 condition attributes. This dataset is used to identify the intrusion behavior on Internet. Analogously, we regard it as streaming data and execute on both SS3WD and IL3WD. The average accumulated time is reported in every 20 percent. Accordingly, we have five pairwise results, as described in Fig. 5.

From Fig. 5 we can observe that SS3WD outperforms IL3WD significantly. The time cost of IL3WD is nearly twice when all data are computed. Since the magnitude of big data is easily to reach PB or ZB, reducing half signifies that we can complete the same task with less space but more advanced resources.

Experiment 3. *Computational time with different thresholds.*

The selection of thresholds may also impact the performance of both algorithms since the equivalence class may be different. Since the difference of time cost in Fig. 4 and Fig. 5 is rather dramatic, herein we only present the accumulated average time cost of the datasets appeared in Fig. 4. Inspired by thresholds of α and β discussed in Ref, we randomly select five pairs which observe the hypothesis $\alpha + \beta = 1$. They are $(\alpha, \beta) = (0.6, 0.4), (0.7, 0.3), (0.8, 0.2), (0.9, 0.1)$ and $(1, 0)$ respectively. Detailed change trend of computation time on SS3WD are given in Fig. 6 respectively.

The results displayed in Fig. 6 reveals that thresholds have minor fluctuations on performance. Moreover, the absolute computing time seems to be more affected by the size of original dataset as compared to concept number $|D|$ and attribute number $|C|$. Such result illustrates that in most cases, $O(|U^{(t)}/C||C|)$ occupies most of calculations in both SS3WD and IL3WD. The conditional probability variation in equivalence class do not necessarily yield to variations of three-way regions, and it is the reason why time cost may fluctuate with different thresholds. Furthermore, variations in thresholds will determine the decision quality, which is beyond the scope of present paper. How to find an optimal pair of threshold is still an open issue to be investigated.

Remark 3. Herein we will discuss the reason why throughout the experiments no significant speedup is observed. As declared in assertion ii) (see page 14), drastic speedup is probable if the immigration object and emigration object share the condition values. We know that in benchmark such case is not frequently appeared, and unfortunately it is only a necessary condition. The distribution in certain equivalence class should be around either α or β so that incremental learning should update twice, and the final region disposal keep consistent as original. In this case, only new object $\bar{x}$ is required to assign the affiliation of region in SS3WD. Considering the limitation of memory, it demands decision class of the identical equivalence class data nearby should not deviate the threshold much. It is thus an event with small probability. Besides, speedup of accumulated time is significant unless the aforementioned case occurs frequently in certain subset. Therefore, it is also not likely to appear in the Fig. 6. All experiments can support the remaining assertions, thus stating that stream computing learning method is efficient.

6. Conclusion

Stream computing paradigm is a new computing paradigm in the era of big data which embraces simplified calculation and advanced resource scheduling. In this paper, we firstly discriminate concept between incremental learning method and stream computing learning method. Secondly, we integrate three-way decisions theory to approximate the real-time concept variations. Thirdly, conditional probabilities are employed as an indicator to determine region affiliation of objects generated by equivalence relation, whereas the dynamism of conditional probability is determined by simultaneously immigration and emigration of object. Theoretically, the correctness is guaranteed by four theorems and three corollaries. Finally, experiments shows that proposed algorithm SS3WD is not only faster, but also more robust.

There are several directions to be investigated for applying rough sets on stream computing learning method. Firstly, we can explore the integration of three-way decisions and stream computing learning method in some challenging field such as multi-source and spatio-temporal interoperated scenario. This signifies that not only rough set, but also other tools such as formal concept analysis, shadowed sets as well as fuzzy sets can be considered. Secondly, hierarchical granular structure on stream computing learning method should be studied for tasks of concept drift. Thirdly, we can replace the binary relations based on the data properties and derive different types of stream computing learning method under the umbrella of three-way decisions.

Acknowledgements

Authors would like to thank the anonymous reviewers for their constructive comments and valuable suggestions. This work is supported by the National Natural Science Foundation of China (Nos. 61273304, 61673301) and Specialized Research Fund for the Doctoral Program of Higher Education of China (No. 20130072130004).

References

[1] J. Błaszczyński, R. Słowiński, Incremental induction of decision rules from dominance-based rough approximations, Electron. Notes Theor. Comput. Sci. 82 (4) (2003) 40–51.
[2] C.L.P. Chen, C.Y. Zhang, Data-intensive applications, challenges, techniques and technologies: a survey on Big Data, Inf. Sci. 275 (11) (2014) 314–347.
[3] H.M. Chen, T.R. Li, S. Qiao, A rough set based dynamic maintenance approach for approximations in coarsening and refining attribute values, Int. J. Intell. Syst. 25 (10) (2010) 1005–1026.
[4] H.M. Chen, T.R. Li, R. Da, Maintenance of approximations in incomplete ordered decision systems while attribute values coarsening or refining, Knowl.-Based Syst. 31 (7) (2012) 140–161.
[5] H.M. Chen, T.R. Li, R. Da, A rough set-based incremental approach for updating approximations under dynamic maintenance environments, IEEE Trans. Knowl. Data Eng. 25 (99) (2013) 274–284.
[6] H.M. Chen, T.R. Li, C. Luo, S.J. Horng, G.Y. Wang, A decision-theoretic rough set approach for dynamic data mining, IEEE Trans. Fuzzy Syst. 23 (6) (2015) 1958–1970.
[7] H.C. Chen, R.H.L. Chiang, V.C. Storey, Business intelligence and analytics: from big data to big impact, MIS Q. 36 (4) (2012) 1165–1188.
[8] M. Chen, S.W. Mao, Y.H. Liu, Big data: a survey, Mob. Netw. Appl. 19 (2) (2014) 171–209.
[9] J. Chen, Y.P. Zhang, S. Zhao, Multi-granular mining for boundary regions in three-way decision theory, Knowl.-Based Syst. 91 (2016) 287–292.
[10] R.T. Das, K.K. Ang, C. Quek, ieRSPOP: a novel incremental rough set-based pseudo outer-product with ensemble learning, Appl. Soft Comput. 46 (2016) 170–186.
[11] X.F. Deng, Y.Y. Yao, A multifaceted analysis of probabilistic three-way decisions, Fundam. Inform. 132 (3) (2014) 291–313.
[12] P. Dey, S. Dey, S. Datta, Dynamic discreduction using rough sets, Appl. Soft Comput. 11 (5) (2011) 3887–3897.
[13] E. Elsebakhi, F. Lee, E. Schendel, A. Haque, N. Kathireason, T. Pathare, N. Syed, R. AI-Ali, Large-scale machine learning based on functional networks for biomedical big data with high performance computing platforms, J. Comput. Sci. 11 (2015) 69–81.
[14] S. Eskandari, M.M. Javidi, Online streaming feature selection using rough sets, Int. J. Approx. Reason. 69 (2016) 35–57.
[15] I.A.T. Hashem, V. Chang, N.B. Anuar, K. Adewole, I. Yaqoob, A. Gani, E. Ahmed, H. Chiroma, The role of big data in smart city, Int. J. Inf. Manag. 36 (5) (2016) 748–758.
[16] B.Q. Hu, Three-way decisions space and three-way decisions, Inf. Sci. 281 (2014) 21–52.
[17] B.Q. Hu, H. Wong, K.F.C. Yiu, The aggregation of multiple three-way decision spaces, Knowl.-Based Syst. 98 (2016) 241–249.
[18] X.Y. Jia, Z. Tang, W. Liao, On an optimization representation of decision-theoretic rough sets paradigm, Int. J. Approx. Reason. 55 (1) (2014) 156–166.
[19] G.M. Lang, Q.G. Li, M.J. Cai, T. Yang, Characteristic matrixes-based knowledge reduction in dynamic covering decision information systems, Knowl.-Based Syst. 85 (2015) 1–26.
[20] G.M. Lang, Q.G. Li, T. Yang, An incremental approach to attribute reduction of dynamic set-valued information systems, Int. J. Mach. Learning Cybern. 5 (5) (2014) 775–788.
[21] J.Y. Liang, F. Wang, C.Y. Dang, A group incremental approach to feature selection applying rough set technique, IEEE Trans. Knowl. Data Eng. 26 (2) (2014) 294–308.
[22] J.H. Li, C.C. Huang, J.J. Qi, Y.H. Qian, W.Q. Liu, Three-way cognitive concept learning via multi-granularity, Inf. Sci. 378 (2017) 244–263.
[23] S.Y. Li, T.R. Li, D. Liu, Incremental updating approximations in dominance-based rough sets approach under the variation of the attribute set, Knowl.-Based Syst. 40 (1) (2013) 17–26.
[24] S.Y. Li, T.R. Li, J. Hu, Update of approximations in composite information systems, Knowl.-Based Syst. 83 (2015) 138–148.
[25] T.R. Li, R. Da, W. Geert, A rough sets based characteristic relation approach for dynamic attribute generalization in data mining, Knowl.-Based Syst. 20 (5) (2007) 485–494.

[26] D. Liu, T.R. Li, R. Da, An incremental approach for inducing knowledge from dynamic information systems, Fundam. Inform. 94 (2) (2009) 245–260.
[27] D. Liu, T.R. Li, J.B. Zhang, A rough set-based incremental approach for learning knowledge in dynamic incomplete information systems, Int. J. Approx. Reason. 55 (8) (2014) 1764–1786.
[28] D. Liu, T.R. Li, J.B. Zhang, Incremental updating approximations in probabilistic rough sets under the variation of attributes, Knowl.-Based Syst. 73 (2015) 81–96.
[29] C. Luo, T.R. Li, H.M. Chen, H. Fujita, Y. Zhang, Efficient updating of probabilistic approximations with incremental objects, Knowl.-Based Syst. 109 (2016) 71–83.
[30] C. Luo, T.R. Li, Y. Zhang, H. Fujita, Matrix approach to decision-theoretic rough sets for evolving data, Knowl.-Based Syst. 99 (2016) 123–134.
[31] X.F. Meng, C. Xiang, Big data management: concepts, techniques and challenges, J. Comput. Res. Dev. 50 (1) (2013) 146–169 (in Chinese).
[32] G. Napoles, I. Grau, E. Papageorgiou, R. Bello, K. Vanhoof, Rough cognitive networks, Knowl.-Based Syst. 91 (2016) 46–61.
[33] H.T.H. Nguyen, J. Cao, Trustworthy answers for top-k queries on uncertain Big Data in decision making, Inf. Sci. 318 (10) (2015) 73–90.
[34] Z. Pawlak, Rough sets, Int. J. Comput. Inf. Sci. 11 (1982) 341–356.
[35] Y.H. Qian, J.Y. Liang, C.Y. Dang, Incomplete multigranulation rough set, IEEE Trans. Syst. Man Cybern., Part A, Syst. Hum. 40 (2) (2010) 420–431.
[36] W.H. Shu, H. Shen, Updating attribute reduction in incomplete decision systems with the variation of attribute set, Int. J. Approx. Reason. 55 (3) (2014) 867–884.
[37] D.W. Sun, G.Y. Zhang, W.M. Zheng, Big data stream computing: technologies and instances, J. Softw. 25 (4) (2014) 839–862 (in Chinese).
[38] F. Wang, J.Y. Liang, Y.H. Qian, Attribute reduction: a dimension incremental strategy, Knowl.-Based Syst. 39 (2) (2013) 95–108.
[39] X.D. Wu, X.Q. Zhu, G.Q. Wu, Data mining with big data, IEEE Trans. Knowl. Data Eng. 26 (1) (2014) 97–107.
[40] Y.T. Xu, L.S. Wang, R.Y. Zhang, A dynamic attribute reduction algorithm based on 0–1 integer programming, Knowl.-Based Syst. 24 (8) (2011) 1341–1347.
[41] X.P. Yang, J.T. Yao, Modelling multi-agent three-way decisions with decision-theoretic rough sets, Fundam. Inform. 115 (2–3) (2012) 157–171.
[42] D. Yang, E.A. Rundensteiner, M.O. Ward, Mining neighbor-based patterns in data streams, Inf. Sci. 38 (3) (2013) 331–350.
[43] Y.Y. Yao, Three-way decision: an interpretation of rules in rough set theory, in: International Conference on Rough Sets and Knowledge Technology, 2009, pp. 642–649.
[44] Y.Y. Yao, The superiority of three-way decisions in probabilistic rough sets models, Inf. Sci. 181 (6) (2011) 1080–1096.
[45] Y.Y. Yao, Three-way decisions and cognitive computing, Cogn. Comput. 8 (4) (2016) 543–554.
[46] Y. Yi, D.W. Wang, Soft computing for scheduling with batch setup times and earliness-tardiness penalties on parallel machines, J. Intell. Manuf. 14 (3–4) (2003) 311–322.
[47] H. Yu, Z. Liu, G.Y. Wang, An automatic approach to determine the number of clusters using decision-theoretic rough sets, Int. J. Approx. Reason. 55 (1) (2014) 101–115.
[48] H. Yu, C. Zhang, G.Y. Wang, A tree-based incremental overlapping clustering method using the three-way decision theory, Knowl.-Based Syst. 91 (2016) 189–203.
[49] H. Yu, P. Jiao, Y.Y. Yao, G.Y. Wang, Detecting and refining overlapping regions in complex networks with three-way decisions, Inf. Sci. 373 (2016) 21–41.
[50] A.P. Zeng, T.R. Li, J. Hu, H.M. Chen, C. Luo, Dynamical updating fuzzy rough approximations for hybrid data under the variation of attribute values, Inf. Sci. 378 (2017) 363–388.
[51] J.B. Zhang, T.R. Li, R. Da, Rough sets based matrix approaches with dynamic attribute variation in set-valued information systems, Int. J. Approx. Reason. 53 (4) (2012) 620–635.
[52] J.B. Zhang, T.R. Li, H.M. Chen, Composite rough sets for dynamic data mining, Inf. Sci. 257 (2014) 81–100.
[53] X.H. Zhang, D.Q. Miao, C.H. Liu, M.L. Le, Constructive methods of rough approximation operators and multigranulation rough sets, Knowl.-Based Syst. 91 (2016) 114–125.
[54] B. Zhou, Multi-class decision-theoretic rough sets, Int. J. Approx. Reason. 55 (1) (2014) 211–224.