Dynamic Online learning Algorithm For Three-way decision

Gaojie Cheng [1,2] , Jianmo Liu [1,2,3] , Jianfeng Xu[1,2,3a] ,Xiqiu Wang [1,2]

[1] Software School,of Nanchang University，Jiangxi Nanchang 330029
[2] Collaborative Innovation Center for Economics crime investigation and prevention technology，Jiangxi Province
[3] Information Engineering School of Nanchang University，Jiangxi Nanchang 330029

Abstract. Three-way decision is an important theory for solving uncertain problems. Online computing is a new dynamic Stream computing form . How to execute three-way decision quickly in online computing is a challenging topic. In this paper, Online computing process is divided into incremental computing portion and decreasing computing portion. And a three-way decision dynamic incremental and decreasing learning algorithm for online computing is proposed. Firstly, the dynamic incremental and decreasing learning models is studied for stream computing based on probabilistic rough set . Then, the logical reasoning for three-way decision regions changing are discussed based on the dynamic incremental and decreasing learning models. And a novel dynamic online learning algorithm for three-way decision online computing is proposed based on the above theory. Finally, the experiment by UCI data set show that the proposed algorithms are superior than classical static three-way decision method in time efficiency.

1 Instruction

Three-way decision is one of the important uncertainty decision theory[1]. In recent years, Three-way decision theory was applied to many application fields, like spam filtering, text emotion and image recognition [2][3][4].These application have proved the superiority of three-way decision. The probabilistic rough set dynamic computation is another important research field. Stream data[5] computing is a novel online computing era for dynamic big data. The main characteristic of Online Stream computing is that the dynamic data through memory and CPU by sliding window without external memory cache. From the view of memory, we can find that the essence of Online Stream computing that can be implemented both in incremental computing and decreasing computing in memory [6]. With the development of dynamic Online Stream computing platform, such as Twitter, Kafka, Storm, YahooS4, the importance of dynamic Online Stream computing became more prominent.

In a dynamic system, incremental learning method learns new knowledge constantly from new samples and the previous learned knowledge, without relearning all the data. Thus, incremental learning can reduce the demand of time and space, and more able to be meet the actual requirements. Incremental learning of rough set [7] has been extensively studied in recent years, the main research contents involve approximation [8], attribute reduction[9] and decision rule extracting [10]. However, Online Stream computing is indeed a new research topic in the field of rough set research. In this study, Online Stream computing process is divided into incremental and decreasing computing portion. Then, a dynamic incremental and decreasing learning algorithm for Three-way decision Online Stream computing is proposed.

[a] Corresponding author: jianfeng_x@ncu.edu.cn

2 Basic theory

Probabilistic rough set is the basic prototypes for constructing Three-way decision theory. The following definitions are basic theory of Three-way decision theory [11]. An information system is defined as: $IS=(U,C,D,f)$,where U is a finite non-empty set of data objects, C is condition attribute set, D is decision attribute set, f is the information function from U to C,D . Given two subset $D_j \subseteq U/D$, $[x]\subseteq C/U$ .This probability of $[x]\rightarrow D_j$ may be simply estimated as follows: $\Pr(D_j\,|\,[x]) = |D_j \cap [x]| / |[x]|$

Definition 1: (Three-way decision) Given a set of threshold values, the positive, boundary and negative regions of Three-way decision can be represented respectively as:

$$POS_{(\alpha,\bullet)}(D_j)=\{R_i\in U/R \mid P(D_j\mid R_i)\geq \alpha\}$$

$$BND_{(\alpha,\beta)}(D_j)=\{R_i\in U/R \mid \beta < P(D_j\mid R_i) < \alpha\}$$

$$NEG_{(\bullet,\beta)}(D_j)=\{R_i\in U/R \mid P(D_j\mid R_i)\leq \beta\}$$

Where $0\leq\beta<\alpha\leq 1$.

The corresponding decisions of the three regions can be interpreted as reception, delay, and rejection, respectively:

$DES_{Accept}(R_i \to D_j), for \quad R_i \subseteq POS_{(\alpha,\bullet)}(D_j)$
$(i = 1,2\cdots,m; j = 1,2\cdots,n);$
$DES_{Defer}(R_i \to D_j), for \quad R_i \subseteq BND_{(\alpha,\beta)}(D_j)$
$(i = 1,2\cdots,m; j = 1,2\cdots,n);$
$DES_{Reject}(R_i \to D_j), for \quad R_i \subseteq NEG_{(\bullet,\beta)}(D_j)$

3 The Online Stream computing algorithm for three-way decision

3.1 Online steam computing model analysis

When an object x is added to an information system in memory, the new object is denoted as x_+. After the increase of the information system, the equivalence classes of each condition attribute equivalence class and each decision attribute equivalence class can be updated by the following formula.

$$R_i^{t+1} = \begin{cases} R_i^t \cup \{x_+\} & x_+ \in R_i^t \quad 1 \le i \le m \\ \{x_+\} & x_+ \in R_i^{t+1} \quad i = m+1 \end{cases}$$

$$D_j^{t+1} = \begin{cases} D_j^t \cup \{x_+\} & x_+ \in D_j^t \quad 1 \le j \le n \\ \{x_+\} & x_+ \in D_j^{t+1} \quad j = n+1 \end{cases}$$

When an object x is deleted from the memory information system, the deleted object is denoted as x_-. When the information system is deleted, the equivalence classes of each condition attribute and the equivalence classes of each decision attribute can be updated by the following formula.

$$\begin{cases} R_i^{t+1} = R_i^t - \{x_-\} & x_- \in R_i^t \quad 1 \le i \le m \\ D_j^{t+1} = D_j^t - \{x_-\} & x_- \in D_j^t \quad 1 \le j \le n \end{cases}$$

The superscript t denotes the initial time, and the superscript t+1 denotes the time after the object is deleted and added.

3.2 Single object incremental updating strategy for Three-way decision

For a given decision equivalence class, an object is added, and the positive, negative, and boundary regions of three-way decision are changed as follows.

Theorem 1: Given an information system IS, when $D_j^{t+1} = D_j^t \cup \{x_+\}$ and $R_i^{t+1} = \{x_+\}$, we have
$POS_{(\alpha,\bullet)}(D_j^{t+1}) = POS_{(\alpha,\bullet)}(D_j^t) \cup \{x_+\}$

Theorem 2: Given an information system IS, when $D_j^{t+1} = D_j^t$ and $R_i^{t+1} = \{x_+\}$, we have
$NEG_{(\bullet,\beta)}(D_j^{t+1}) = NEG_{(\bullet,\beta)}(D_j^t) \cup \{x_+\}$

Theorem 3: Given an information system IS, when $D_j^{t+1} = D_j^t \cup \{x_+\}$ and $R_i^{t+1} = R_i^t \cup \{x_+\}$, we have

3.1 If $R_i^t \subseteq POS_{(\alpha,\bullet)}(D_j^t)$, we have
$POS_{(\alpha,\bullet)}(D_j^{t+1}) = POS_{(\alpha,\bullet)}(D_j^t) \cup \{x_+\}$

3.2 If $R_i^t \subseteq BND_{(\alpha,\beta)}(D_j^t)$, then
If $P(D_j^{t+1} \mid R_i^{t+1}) \ge \alpha$, we have
$$\begin{cases} POS_{(\alpha,\bullet)}(D_j^{t+1}) = POS_{(\alpha,\bullet)}(D_j^t) \cup R_i^{t+1} \\ BND_{(\alpha,\beta)}(D_j^{t+1}) = BND_{(\alpha,\beta)}(D_j^t) - R_i^t \end{cases}$$

If $P(D_j^{t+1} \mid R_i^{t+1}) < \alpha$, we have
$BND_{(\alpha,\beta)}(D_j^{t+1}) = BND_{(\alpha,\beta)}(D_j^t) \cup \{x_+\}$

3.3 If $R_i^t \subseteq NEG_{(\bullet,\beta)}(D_j^t)$, then
If $P(D_j^{t+1} \mid R_i^{t+1}) \ge \alpha$, we have
$$\begin{cases} POS_{(\alpha,\bullet)}(D_j^{t+1}) = POS_{(\alpha,\bullet)}(D_j^t) \cup R_i^{t+1} \\ NEG_{(\bullet,\beta)}(D_j^{t+1}) = NEG_{(\bullet,\beta)}(D_j^t) - R_i^t \end{cases}$$

If $\beta < P(D_j^{t+1} \mid R_i^{t+1}) < \alpha$, we have
$$\begin{cases} BND_{(\alpha,\beta)}(D_j^{t+1}) = BND_{(\alpha,\beta)}(D_j^t) \cup R_i^{t+1} \\ NEG_{(\bullet,\beta)}(D_j^{t+1}) = NEG_{(\bullet,\beta)}(D_j^t) - R_i^t \end{cases}$$

If $P(D_j^{t+1} \mid R_i^{t+1}) \le \beta$, we have
$NEG_{(\bullet,\beta)}(D_j^{t+1}) = NEG_{(\bullet,\beta)}(D_j^t) \cup \{x_+\}$.

Theorem 4: Given an information system IS, when $D_j^{t+1} = D_j^t$ and $R_i^{t+1} = R_i^t \cup \{x_+\}$, we have

4.1 If $R_i^t \subseteq POS_{(\alpha,\bullet)}(D_j^t)$, then
If $P(D_j^{t+1} \mid R_i^{t+1}) \ge \alpha$, we have
$POS_{(\alpha,\bullet)}(D_j^{t+1}) = POS_{(\alpha,\bullet)}(D_j^t) \cup \{x_+\}$

If $\beta < P(D_j^{t+1} \mid R_i^{t+1}) < \alpha$, we have
$$\begin{cases} BND_{(\alpha,\beta)}(D_j^{t+1}) = BND_{(\alpha,\beta)}(D_j^t) \cup R_i^{t+1} \\ POS_{(\alpha,\bullet)}(D_j^{t+1}) = POS_{(\alpha,\bullet)}(D_j^t) - R_i^t \end{cases}$$

If $P(D_j^{t+1} \mid R_i^{t+1}) \le \beta$, we have
$$\begin{cases} NEG_{(\bullet,\beta)}(D_j^{t+1}) = NEG_{(\bullet,\beta)}(D_j^t) \cup R_i^{t+1} \\ POS_{(\alpha,\bullet)}(D_j^{t+1}) = POS_{(\alpha,\bullet)}(D_j^t) - R_i^t \end{cases}$$

4.2 If $R_i^t \subseteq BND_{(\alpha,\beta)}(D_j^t)$, then
If $P(D_j^{t+1} \mid R_i^{t+1}) > \beta$, we have
$BND_{(\alpha,\beta)}(D_j^{t+1}) = BND_{(\alpha,\beta)}(D_j^t) \cup \{x_+\}$

If $P(D_j^{t+1} \mid R_i^{t+1}) \le \beta$, we have
$$\begin{cases} NEG_{(\bullet,\beta)}(D_j^{t+1}) = NEG_{(\bullet,\beta)}(D_j^t) \cup R_i^{t+1} \\ BND_{(\alpha,\beta)}(D_j^{t+1}) = BND_{(\alpha,\beta)}(D_j^t) - R_i^t \end{cases}$$

4.3 If $R_i^t \subseteq NEG_{(\bullet,\beta)}(D_j^t)$, we have
$NEG_{(\bullet,\beta)}(D_j^{t+1}) = NEG_{(\bullet,\beta)}(D_j^t) \cup \{x_+\}$.

The proof of these above theorem is brief.

3.3 Single object decreasing updating strategy for Three-way decision

For a given decision equivalence class, an object is deleted, and its positive, negative, and boundary regions change as follows.

Theorem 5: Given an information system IS, when $D_j^{t+1} = D_j^t - \{x_-\}$ and $R_i^{t+1} = R_i^t - \{x_-\}$, we have

5.1 If $R_i^t \subseteq POS_{(\alpha,\bullet)}(D_j^t)$, then
If $P(D_j^{t+1} \mid R_i^{t+1}) \ge \alpha$, we have
$POS_{(\alpha,\bullet)}(D_j^{t+1}) = POS_{(\alpha,\bullet)}(D_j^t) - \{x_-\}$

If $\beta < P(D_j^{t+1} \mid R_i^{t+1}) < \alpha$, we have

$$\begin{cases} BND_{(\alpha,\beta)}(D_j^{t+1}) = BND_{(\alpha,\beta)}(D_j^t) \cup R_i^{t+1} \\ POS_{(\alpha,\bullet)}(D_j^{t+1}) = POS_{(\alpha,\bullet)}(D_j^t) - R_i^t \end{cases}$$

If $P(D_j^{t+1} \mid R_i^{t+1}) \le \beta$, we have

$$\begin{cases} NEG_{(\bullet,\beta)}(D_j^{t+1}) = NEG_{(\bullet,\beta)}(D_j^t) \cup R_i^{t+1} \\ POS_{(\alpha,\bullet)}(D_j^{t+1}) = POS_{(\alpha,\bullet)}(D_j^t) - R_i^t \end{cases}$$

5.2 If $R_i^t \subseteq BND_{(\alpha,\beta)}(D_j^t)$, then

If $P(D_j^{t+1} \mid R_i^{t+1}) > \beta$, we have

$$BND_{(\alpha,\beta)}(D_j^{t+1}) = BND_{(\alpha,\beta)}(D_j^t) - \{x_-\}$$

If $P(D_j^{t+1} \mid R_i^{t+1}) \le \beta$, we have

$$\begin{cases} NEG_{(\bullet,\beta)}(D_j^{t+1}) = NEG_{(\bullet,\beta)}(D_j^t) \cup R_i^{t+1} \\ BND_{(\alpha,\beta)}(D_j^{t+1}) = BND_{(\alpha,\beta)}(D_j^t) - R_i^t \end{cases}$$

5.3 If $R_i^t \subseteq NEG_{(\bullet,\beta)}(D_j^t)$, we have

$$NEG_{(\bullet,\beta)}(D_j^{t+1}) = NEG_{(\bullet,\beta)}(D_j^t) - \{x_-\}.$$

Theorem 6: Given an information system IS, when $D_j^{t+1} = D_j^t$ and $R_i^{t+1} = R_i^t - \{x_-\}$, we have

6.1 If $R_i^t \subseteq POS_{(\alpha,\bullet)}(D_j^t)$, we have

$$POS_{(\alpha,\bullet)}(D_j^{t+1}) = POS_{(\alpha,\bullet)}(D_j^t) - \{x_-\}$$

6.2 If $R_i^t \subseteq BND_{(\alpha,\beta)}(D_j^t)$, then

If $P(D_j^{t+1} \mid R_i^{t+1}) \ge \alpha$, we have

$$\begin{cases} POS_{(\alpha,\bullet)}(D_j^{t+1}) = POS_{(\alpha,\bullet)}(D_j^t) \cup R_i^{t+1} \\ BND_{(\alpha,\beta)}(D_j^{t+1}) = BND_{(\alpha,\beta)}(D_j^t) - R_i^t \end{cases}$$

If $P(D_j^{t+1} \mid R_i^{t+1}) > \beta$, we have

$$BND_{(\alpha,\beta)}(D_j^{t+1}) = BND_{(\alpha,\beta)}(D_j^t) - \{x_-\}$$

6.3 If $R_i^t \subseteq NEG_{(\bullet,\beta)}(D_j^t)$, then

If $P(D_j^{t+1} \mid R_i^{t+1}) \ge \alpha$, we have

$$\begin{cases} POS_{(\alpha,\bullet)}(D_j^{t+1}) = POS_{(\alpha,\bullet)}(D_j^t) \cup R_i^{t+1} \\ NEG_{(\bullet,\beta)}(D_j^{t+1}) = NEG_{(\bullet,\beta)}(D_j^t) - R_i^t \end{cases}$$

If $\beta < P(D_j^{t+1} \mid R_i^{t+1}) < \alpha$, we have

$$\begin{cases} BND_{(\alpha,\beta)}(D_j^{t+1}) = BND_{(\alpha,\beta)}(D_j^t) \cup R_i^{t+1} \\ NEG_{(\bullet,\beta)}(D_j^{t+1}) = NEG_{(\bullet,\beta)}(D_j^t) - R_i^t \end{cases}$$

If $P(D_j^{t+1} \mid R_i^{t+1}) \le \beta$, we have

$$NEG_{(\bullet,\beta)}(D_j^{t+1}) = NEG_{(\bullet,\beta)}(D_j^t) - \{x_-\}.$$

The proof of these above theorem is brief.

4 Online incremental and decreasing learning algorithm for three-way decision

4.1 Algorithm

In the online Stream computing model, the changing data can be existed at the same time in memory. According to the idea of time division multiplexing, the dynamic incremental and decreasing learning algorithm based on Three-way decision are proposed to deal with the online Stream computing problem.

Dynamic incremental and decreasing learning algorithms for online Stream computing as follows:

Algorithm input: The data x_+ and x_-, all decision equivalence classes D_j^t and their corresponding positive region $POS_{(\alpha,\bullet)}(D_j^t)$, boundary region $BND_{(\alpha,\beta)}(D_j^t)$, negative region $NEG_{(\bullet,\beta)}(D_j^t)$ and threshold (α, β).

Algorithm output: The updated data and all decision equivalence classes D_j^{t+2}, the positive region $POS_{(\alpha,\bullet)}(D_j^{t+2})$, the boundary region $BND_{(\alpha,\beta)}(D_j^{t+2})$ and the negative region $NEG_{(\bullet,\beta)}(D_j^{t+2})$.

Step 1: Reduce learning, remove data x_-, and update Three-way decision areas of D_j^t.

Step 1.1: remove the data x_- and choose between theorem 5-6 based on the relation between the x_- and the conditional equivalence class (contained in the 3 decision regions) and the decision equivalence class D_j^t.

Step 1.2: According to the conditions x_- equivalence class R_i^t belongs to the decision domain, and select the sub theorem in the theorem obtained in step 1.1.

Step 1.3: The conditional probability $P(D_j^{t+1} \mid R_i^{t+1})$ is calculated. According to the relation between the conditional probability value and the threshold value, the specific case of the neutron theorem of step 1.2 is selected, and the transformation of the decision region is carried out.

Step 2: Incremental learning, adding data, x_+, and updating Three-way decision areas of D_j^{t+1} after removing data x_-.

Step 2.1: Add data x_+ and choose between theorem 1-4 based on the relation between x_+ and the conditional equivalence class (contained in 3 decision regions) and the decision equivalence class D_j^{t+1}.

Step 2.2: According to the decision region of the conditional equivalence class R_i^{t+1} of x_+, the sub theorem is selected in the theorem obtained in step 2.1.

Step 2.3: Calculate the conditional probability x_+ of $P(D_j^{t+2} \mid R_i^{t+2})$. According to the relation between the conditional probability value and the threshold value, select the specific case of the neutron theorem of step 2.2 and transform the decision domain.

4.2 Time complexity analysis of algorithm

The time complexity analysis of dynamic incremental and decreasing learning algorithm based on Three-way decision is deducted as followed:

Assume there are M data in memory, $M = |U|$. In the decreasing learning stage, the time complexity of reducing learning is $O(M^2 + m + 8)$.

In the incremental learning phase, the complexity of incremental learning time is $O(M^2+2M+m+9)$.

So the time complexity of this new dynamic algorithm is $O(n*(2M^2+2M+2m+17))$, $n=|U/D|$.

Its easy to proof that the time complexity of Three-way decision classical non-incremental learning algorithms is $O(n*(M^3+2M^2+3M))$, $n=|U/D|$.

We can see that the time complexity of dynamic incremental and decreasing learning algorithms is reduced from the M^3 number to the M^2 compared to the classical non-incremental learning algorithm.

5 Experiment and analysis

This chapter uses eight datasets on UCI to test the advantages of the dynamic incremental and decreasing learning algorithm to the classical non-incremental algorithms in terms of time cost. The operating system is windows7 machine configuration for core i7-2670QM processor (2.2GHz frequency), and the allocation of memory for 8G.

The eight datasets used in the experiment were derived from UCI (http://archive.ics.uci.edu/ml/datasets). The basic information of 1 breast cancer, 2 contraceptive method choice, 3 mammographic mass, 4 monk's problems, 5 skin segmentation, 6 thoracic surgery data, 7 Balance Scale and 8 Indian Liver with Patient Dataset is shown in table 1.

Table 1 The basic information of the datasets

Data set	Samples	Features	Concept Count
1	699*100	10	2
2	1473*10	9	3
3	961*100	6	2
4	432*100	7	2
5	245057	4	2
6	470*100	17	2
7	625*100	4	3
8	583*100	10	2

This experiment simulates the inflow and outflow of the memory data. The algorithm process is given as followed: Firstly, the amount of data in memory is fixed. Then, a new data is added into the memory, and other data object is deleted at the same time from the memory. The stream simulation process is repeated until all stream data is computed.

In this experiment, the amount of data stored in the set memory is 1000, and the thresholds α=0.75, β=0.35 respectively. Dynamic incremental and decreasing learning algorithm as algorithm1. Classical non-incremental learning algorithms of Three-way decision as algorithm2 . The two algorithms was done 30000 times and the time cost result is shown in table 2.

Table 2 Time cost for algorithm1 and algorithm2

Data set	Algorithm1	Algorithm2
1	8s	350s
2	14s	600s
3	9s	350s
4	9s	350s
5	10s	400s
6	11s	500s
7	12.5s	450s
8	13s	510s

The time cost of the two algorithms both showed a linear growth trend, and the proposed dynamic learning algorithm has great reduced of the time cost. The experimental results and the time complexity analysis are matched with the time complexity of section

6. Conclusion

With the development of big data, the proportion of stream data in machine learning and big data applications is more evidently. It is of great significance to do research on Online Stream computing based on three-way decisions. In this paper, the incremental and decreasing mode of single object insertion and deletion was analyzed , and then the computing time complexity for the Three-way decision is deducted. The experiments by UCI data sets show that the proposed online computing algorithms are superior than classical rough set three-way decision method in time efficiency.

Acknowledgement

This work is supported by :National Natural Science Foundation of China (61763031)The Opening Project of Collaborative Innovation Center for Economics crime investigation and prevention technology，Jiangxi, China (Nos. JXJZXTCX-023).

References

1. Y Yao . An Outline of a Theory of Three-way decisions.Rough set and Current Trends in Computing. Springer Berlin Heidelberg, 2012:1-17.
2. B Zhou , Y Yao , J Luo. Cost-sensitive three-way email spam filtering. Journal of Intelligent Information Systems, 2014, 42(1):19-45.
3. Z Zhang ,Dq Miao ,J Nie , et al. Sentiment uncertainty measure and classification of negative sentences[J]. Journal of Computer Research & Development, 2015.
4. H Li ,L Zhang ,B Huang , et al. Sequential Three-way decision and granulation for cost-sensitive face recognition. Knowledge-Based Systems, 2016, 91(C):241–251.
5. M D Assunção , R N Calheiros, S Bianchi, et al. Big Data computing and clouds: Trends and futuredirections. Journal of Parallel & Distributed Computing, 2015, s 79–80:3-15.
6. J Xu ,D Miao , Y Zhang , et al. A Three-way decision model with probabilistic rough set for

Online Stream computing . International Journal of Approximate Reasoning, 2017.

7. S Li ,T Li ,J Hu . Update of approximations in composite information systems. Knowledge-Based Systems, 2015, 83(1):138-148.
8. A Zeng ,T Li , J Hu , et al. Dynamical updating fuzzy rough approximations for hybrid data under the variation of attribute values. Information Sciences, 2016, 378(C):363-388.
9. M Li , G Wang . Approximate concept construction with Three-way decision and attribute reduction in incomplete contexts. Knowledge-Based Systems, 2016, 91:165-178.
10. R T Das ,K A Kai ,Q Chai . ieRSPOP: A novel incremental rough set-based pseudo outer-product with learning learning. Applied Soft Computing, 2016, 46:170-186.
11. Y Yao . The superiority of Three-way decision in probabilistic rough set models. Information Sciences, 2011, 181(6):1080-1096.