American Journal of Computer Science and Information Engineering
2018; 5(1): 1-8
http://www.aascit.org/journal/ajcsie
ISSN: 2381-1110 (Print); ISSN: 2381-1129 (Online)

American Association for Science and Technology

An Approach to Interval-Valued Complex Information Mining

Yunfei Yin[1, 2], Huan Liu[2], Xuesong Feng[2], Chunmei Ning[2]

[1]Key Lab. of Dependable Service Computing in Cyber Physical Society of Ministry of Education, Chongqing, P. R. China
[2]College of Computer Science, Chongqing University, Chongqing, P. R. China

Email address
yinyunfei@cqu.edu.cn (Yunfei Yin)

Citation
Yunfei Yin, Huan Liu, Xuesong Feng, Chunmei Ning. An Approach to Interval-Valued Complex Information Mining. *American Journal of Computer Science and Information Engineering.* Vol. 5, No. 1, 2018, pp. 1-8.

Received: June 7, 2017; **Accepted**: July 30, 2017; **Published**: February 12, 2018

Abstract: It is a difficult issue for complex information mining, especially for interval-valued information mining. In this paper, an iterative interval-valued mining model is proposed that classifies the interval-valued information into three types, viz., "Interval-Value", "Interval-Interval" and "Interval-Matrix". As to the "Interval-Value" information, the "Netting"→"Type-I clustering"→"Type-II clustering" method is adopted to handle; as to the "Interval-Interval" information, the interval medium clustering method is adopted to handle; as to the "Interval-Matrix", the matrix threshold clustering method is adopted to handle. Finally, based on these three interval-valued clustering methods, the iterative data mining model has been designed. The motivation is to mine the interval-valued association rules from the interval-valued complex information. By the experimental study in the typical interval-valued information fields, the experimental results show the effectiveness and efficiency of the models and algorithms.

Keywords: Data Mining, Interval-Valued, Clustering, Model of Design

1. Introduction

Interval-valued is a sort of calculation pattern, which can present the accurate range of an objective value, and is widely used in aeronautics, military and astronomy. By iterative data mining, the real-time dynamically changing data can be mined. Since the data in real-life are always in changing, the iterative data mining has more practical significance.

In this paper, the motivation is to study the features of interval-valued clustering, and further explore the iterative data mining method based on interval-valued clustering.

The related researches include: (1) the research on interval-valued model; (2) the research on dynamic data mining; (3) the research on iterative algorithm.

The research on interval-valued model aims at the model design based on interval-valued. The typical research includes the “interval-valued based on Interval Order Relations” [1] and the “stereo matching algorithm based on interval-valued fuzzy sets” [2]. Liu et al. proposed a real-coded genetic algorithm with ranking selection to solve the mixed integer constrained optimization problem, called by “interval-valued model based on Interval Order Relations” [1]. Sheng et al. developed a stereo matching based on interval-valued fuzzy sets [2]. In addition, Zhang et al. proposed the interval-valued algorithm based on Network Flow to solve the multi-echelon, spare-part inventory management problem [3]; Ghiyasvand also proposed an interval and fuzzy data model to solve the minimum cost flow problem [4]. These researches provide the enormous impetus for the study of interval-valued models.

Dynamic data mining research aims at mining the dynamic data. The typical research includes the “proximate dynamic model for data mining” [5] and the “data mining model in dynamic environments” [6]. Yin proposed a proximate dynamic model to conduct the processing of dynamic data, and conducted the experimental validations by the dynamic datasets of aeronautics, stock and medicine [5]. Matsubara et al. combine the neural network algorithm and the association rules mining algorithm to conduct dynamic data mining [6]. In addition, Wu et al. constructed a support vector machine model to conduct data mining [7]; Duncan et al. also used the knowledge discovery and data mining technology via a dynamic web site [8].

Iterative algorithm design aims at the design of high-efficient mining algorithms, the typical research includes the “distributed iterative algorithm” [9] and the “iterative algorithm with matrix operations" [10], where Kibriyaand Ramon analyzed the convergence of the distributed algorithms, and proposed the Time Invariant Convergence-Optimal Quantizer algorithm and the Time Varying Convergence-Optimal Quantizer algorithm [10]. In addition, Plonkaet al. designed an iterative Fourier-transform algorithm for the design of diffractive optical elements [11].

The features of interval-valued are combined to propose three interval-valued clustering algorithms. Further, apply the interval-valued algorithms to the complex information mining, and design an iterative data mining method based on interval-valued clustering.

The organization of the paper is as follows: in the next section we will introduce the method of interval-valued clustering, including the "Interval-Value" clustering, the "Interval-Interval" clustering, and the "Interval-Matrix" clustering. In Section 3, the iterative data mining method based on the interval-valued clustering is explored. Section 4 is the experiment study. Section 5 is the conclusion and the future research direction.

2. Interval-Valued Clustering

Interval-valued clustering has three patterns, viz.,<interval; value>, <interval; interval> and <interval; matrix>.

2.1. "Interval-Value" Clustering

Suppose x_1, x_2, … x_n is *n* objects, whose values are interval-valued. The similarity matrix is as follows:

$$R(r_{i,j}) = \begin{bmatrix} 1 & & & \\ [t_{21}^-, t_{21}^+] & 1 & & \\ \ldots & \ldots & \ldots & \\ [t_{n1}^-, t_{n1}^+] & [t_{n2}^-, t_{n2}^+] & \ldots & 1 \end{bmatrix} \quad (1)$$

Where r_{ij} is any element of the matrix *R*, and $r_{ij} = [\, t_{ij}^-, t_{ij}^+ \,]$; t_{ij}^- is the lower boundary of the interval, and t_{ij}^+ is the upper boundary of the interval.

$R(r_{ij})$ is asymmetry matrix and r_{ij} is the similarity of x_i and x_j, where i, j = 1, 2 … n.

Suppose λ_0 is the threshold of Interval-Value clustering, and then, the Interval-Value clustering method is designed as "Netting"→"Type-I clustering"→"Type-II clustering".

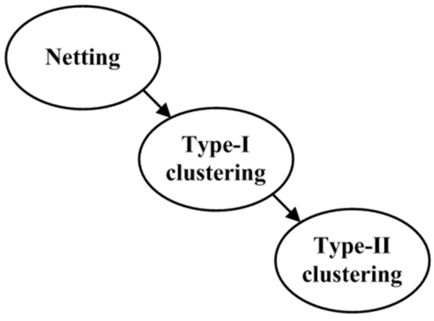

Figure 1. *Interval-Value Clustering.*

(1) Netting. If $\lambda_0 \in [\, t_{ij}^-,\ t_{ij}^+ \,]$, then $[\, t_{ij}^-,\ t_{ij}^+ \,]$ is replaced by "#" in matrix *R*; if $t_{ij}^- > \lambda_0$, then $[\, t_{ij}^-,\ t_{ij}^+ \,]$ is replaced by "×" in matrix *R*; and if $t_{ij}^+ < \lambda_0$, then $[\, t_{ij}^-,\ t_{ij}^+ \,]$ is replaced by a space in matrix *R*.

(2) Type-I clustering. For each "×", find the corresponding objects in the diagonal, and cluster them into one set. See also Figure 2.

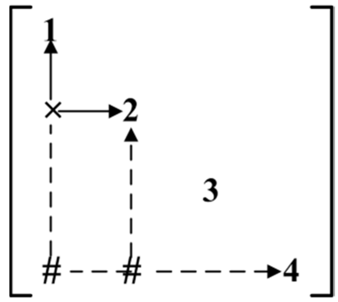

Figure 2. *Interval-Value Clustering.*

(3) Type-II clustering. For each "#", find the corresponding objects in the diagonal, and cluster them into one set. See also Figure 2.

In Figure 2, there are 4 objects marked by 1, 2, 3 and 4. And there are one "×" and two "#". Where, the "×"relates to the object 1 and the object 2; the former "#" relates to the object 1 and the object 4, and the latter "#" relates to the object 2 and the object 4.

Type-I clustering can clearly discern the objects, whereas type-II clustering cannot clearly discern the objects. For example, {1, 2} is a clear set in Figure 2, whereas both {1, 4} and {2, 4} are all loose sets for they cannot clearly cluster the objects.

Definition 1. Call "×" as node and "#" as similar node, where "Node" corresponds to the clear clustering, and "similar node" corresponds to the loose clustering.

Suppose A = {x_1, x_2… x_n}, the similar interval between the objects x and x_i is $[\, t_i^-, t_i^+ \,]$. Let $\alpha_A = \min\{(\, t_i^+ - \lambda_0)/(\, t_i^+ - t_i^- \,)\}$, where λ_0 is the threshold of Interval-Value clustering. If $\alpha_A \geq$ 0.5, then x belongs to A.

Suppose $\alpha = \max\{\alpha_{A1}, \alpha_{A2} \ldots \alpha_{Aj} \ldots \alpha_{Am}\}$, where A1, A2 …$A_m$ denote *m* sets. If $\alpha = \alpha_{Aj} \geq 0.5$, then x belongs to A_j; else if $\alpha = \alpha_{Aj} < 0.5$, then x is a separated set.

Example 1. Suppose x_1, x_2, … x_n is *n* objects, and the similarity matrix is:

$$R = \begin{bmatrix} 1 & & & & \\ [0.85, 0.9] & 1 & & & \\ [0.3, 0.66] & [0.7, 0.8] & 1 & & \\ [0.2, 0.67] & [0.4, 0.6] & [0.7, 0.72] & 1 & \\ [0.7, 0.98] & [0.7, 0.9] & [0.1, 0.7] & [0.22, 0.5] & 1 \end{bmatrix}.$$

And $\lambda_0 = 0.8$ is the threshold of Interval-Value clustering.

Then, object x_1 and object x_2 consists of "×"; object x_2 and object x_3 consists of "#" and object x_2 and object x_5 also consist of "#". According to the Interval-Value clustering method, $\{x_1, x_2\}$ is a clear set denoted by A1, and $\{x_2, x_3, x_5\}$ is a loose set denoted by B. For x_2 belongs to the clear set A1, x_3 and x_5 are required to be evaluated whether they belong to the clear set A1.

So, the evaluation process is as follows:

For object x_3: $\alpha_{A1} = \max\{\min\{(0.66-0.8)/(0.66-0.3), (0.8-0.8)/(0.8-0.7)\}\} = -0.42$

For object x_5: $\alpha_{A1} = \max\{\min\{(0.98-0.8)/(0.98-0.7), (0.9-0.8)/(0.9-0.7)\}\} = 0.5$

According to the Interval-Value clustering method, x_5 belongs to A1, viz., A1= $\{x_1, x_2, x_5\}$, and x_3, x_4 are separated sets denoted by A2 = $\{x_3\}$ and A3 = $\{x_4\}$, respectively.

So, A1 = $\{x_1, x_2, x_5\}$, A2 = $\{x_3\}$ and A3 = $\{x_4\}$.

2.2. "Interval-Interval" Clustering

Suppose $x_1, x_2, \ldots x_n$ is n objects, whose values are interval-valued. Suppose λ_0 is the threshold of Interval-Interval clustering whose value is interval-valued. Obviously, $<x_1, x_2, \ldots x_n; \lambda_0>$areinterval-interval pairs and called Interval-Interval clustering pattern.

Suppose A = $\{x_1, x_2, \ldots x_n\}$ and $\lambda_0 = [\lambda_0^-, \lambda_0^+]$. The similar interval between the object x and the object x_i is $[t_i^-, t_i^+]$.

(1) If $[\lambda_0^-, \lambda_0^+] \cap [t_{ij}^-, t_{ij}^+] \neq \Phi$, then $[t_{ij}^-, t_{ij}^+]$ is replaced by "#"and called a similar node in Figure 2; if $\lambda_0^+ < t_{ij}^-$, then $[t_{ij}^-, t_{ij}^+]$ is replaced by "×"and called a node in Figure 2; and if $\lambda_0^- > t_{ij}^+$, then $[t_{ij}^-, t_{ij}^+]$ is replaced by a space in Figure 2.

(2) The similarity of the object x belonging to A is defined as:

$$[\alpha_A^-, \alpha_A^+] = \min\{[\alpha_1^-, \alpha_1^+], [\alpha_2^-, \alpha_2^+]\ldots[\alpha_i^-, \alpha_i^+]\ldots[\alpha_n^-, \alpha_n^+]\} \tag{2}$$

Where,

$$\alpha_i^- = \begin{cases} 1 + \dfrac{t_i^+ - \lambda_0^+}{t_i^+ - t_i^-} \log_2^{\frac{t_i^+ - \lambda_0^+}{t_i^+ - t_i^-}} & if \quad t_i^- \le \lambda_0^+ < t_i^+ \\ 0 & otherwise \end{cases},$$

$$\alpha_i^+ = \begin{cases} 1 + \dfrac{t_i^+ - \lambda_0^-}{t_i^+ - t_i^-} \log_2^{\frac{t_i^+ - \lambda_0^-}{t_i^+ - t_i^-}} & if \quad t_i^- \le \lambda_0^- < t_i^+ \\ 0 & otherwise \end{cases}.$$

(3) If the similarities of object x belonging to the set A1, A2…, and Am are $[\alpha_{A1}^-, \alpha_{A1}^+]$, $[\alpha_{A2}^-, \alpha_{A2}^+]$…, and $[\alpha_{Am}^-, \alpha_{Am}^+]$, respectively, then, let $\alpha = \max\{(\alpha_{A1}^- + \alpha_{A1}^+)/2, (\alpha_{A2}^- + \alpha_{A2}^+)/2 \ldots (\alpha_{Am}^- + \alpha_{Am}^+)/2\}$.

If $\alpha = (\alpha_i^- + \alpha_i^+)/2 \ge 0.5$, then object belongs to Ai; else if$\alpha = (\alpha_i^- + \alpha_i^+)/2 < 0.5$, then x is a separated set.

Example 2. In Example 1, if $\lambda_0 = [0.7, 0.8]$ that is the threshold of Interval-Interval clustering, the Interval-Interval clustering process is as follows.

Object x_1 and object x_2 consists of "×"; object x_2 and object x_3, object x_3 and object x_4, object x_1 and object x_5, and object x_2 and object x_5 consist of "#", respectively.

According to the Interval-Interval clustering method, $\{x_1, x_2\}$ is a clear set denoted by A1, and $\{x_2, x_3, x_4, x_5\}$ is a loose set denoted by B. For x_2 belongs to the clear set A1, x_3, x_4 and x_5 are required to be evaluated whether they belong to the clear set A1 or not.

So, the evaluation process is as follows:

For object x_3, $\alpha_{A1} = \max\{\min\{[0, 0], [0, 1]\}\} = \max\{[0, 0]\} = 0$

For object x_4, $\alpha_{A1} = \max\{\min\{[0, 0], [0, 0]\}\} = \max\{[0, 0]\} = 0$

For object x_5, $\alpha_{A1} = \max\{\min\{[0.5, 1], [0.59, 1]\}\} = \max\{[0.59, 1]\} = 0.795$

According to the Interval-Interval clustering method, x_5 belongs to A1, viz., A1= $\{x_1, x_2, x_5\}$, and x_3, x_4 are separated sets denoted by A2 = $\{x_3\}$ and A3 = $\{x_4\}$, respectively.

Therefore, A1 = $\{x_1, x_2, x_5\}$, A2 = $\{x_3\}$ and A3 = $\{x_4\}$.

2.3. "Interval-Matrix" Clustering

Interval-Matrix clustering is the pattern that the threshold of Interval-Matrix clustering is a matrix. Suppose $x_1, x_2, \ldots x_n$ is n objects whose values are interval-valued, andλ_0 is the threshold of Interval-Matrix clustering whose value is matrix. Obviously, $<x_1, x_2, \ldots x_n; \lambda_0>$ is called Interval-Matrix clustering pattern.

The threshold of Interval-Matrix clustering has the following form:

$$\lambda_0(\lambda_{i,j}) = \begin{bmatrix} 1 & & & \\ \lambda_{2,1} & 1 & & \\ \ldots & \ldots & \ldots & \\ \lambda_{n,1} & \lambda_{n,2} & \ldots & 1 \end{bmatrix} \tag{3}$$

Where, any element $\lambda_{i,j}$ of the matrix λ_0 denotes the clustering threshold between the object i and the object j.

Suppose A = $\{x_1, x_2, \ldots x_n\}$ and the similar interval between the object x and the object x_i is $[t_i^-, t_i^+]$.

(1) If $\lambda_{i,j} \in [t_{ij}^-, t_{ij}^+]$, then $[t_{ij}^-, t_{ij}^+]$ is replaced by "#" and called a similar node; if $\lambda_{i,j} < t_{ij}^-$, then $[t_{ij}^-, t_{ij}^+]$ is replaced by "×"and called a node; and if $\lambda_{i,j} > t_{ij}^+$, then $[t_{ij}^-, t_{ij}^+]$ is replaced by a space.

(2) The similar interval between the objects x and x_i is $[t_i^-, t_i^+]$; and the clustering threshold between the object x and the object x_i is $\lambda_{0,i}$.

(3) The similarity of the object x belonging to A is $\alpha_A = \min\{(t_i^+ - \lambda_{0,i}) / (t_i^+ - t_i^-)\}$. If $\alpha_A \ge 0.5$, then x belongs to A.

(4) $\alpha = \max\{\alpha_{A1}, \alpha_{A2}, \ldots\alpha_{Aj}, \ldots\alpha_{Am}\}$, where A1, A2 … Am are *m* sets. If $\alpha = \alpha_{Aj} \geq 0.5$, then x belongs to A_j; else If $\alpha = \alpha_{Aj} < 0.5$, then x is a separated set.

Example 3. In Example 1 and Example 2, if the threshold of Interval-Matrix clustering is:

$$\lambda_0\,(\lambda_{i,\,j}) = \begin{bmatrix} 1 & & & & \\ 0.8 & 1 & & & \\ 0.7 & 0.8 & 1 & & \\ 0.8 & 0.7 & 0.8 & 1 & \\ 0.65 & 0.49 & 0.8 & 0.8 & 1 \end{bmatrix}.$$

The clustering for the objects $x_1, x_2, \ldots x_n$ should use the Interval-Matrix clustering method.

Obviously, object x_1 and object x_2 consists of "×"; object x_1 and object x_5, and object x_2 and object x_5 also consist of "×"; and object x_2 and object x_3 consists of "#". According to the Interval-Matrix clustering method, $\{x_1, x_2, x_5\}$ is a clear set denoted by A1, and $\{x_2, x_3\}$ is a loose set denoted by B. For x_2 belongs to the clear set A1, x_3 is required to be evaluated whether it belongs to the clear set A1.

So, the evaluation process is as follows:

For object x_3: $\alpha_{A1} = \max\{\min\{(0.66 - 0.7)/(0.66 - 0.3), (0.8 - 0.8)/(0.8 - 0.7)\}\} = -0.12$

According to the Interval-Matrix clustering method, x_3 does not belong to A1, viz., A1= $\{x_1, x_2, x_5\}$, A2 = $\{x_3\}$ and A3 = $\{x_4\}$.

So, A1 = $\{x_1, x_2, x_5\}$, A2 = $\{x_3\}$ and A3 = $\{x_4\}$.

3. Interval-Valued Clustering for Complex Information Mining

3.1. Model Descriptions

The complex information mining model based on interval-valued clustering is shown in Figure 3.

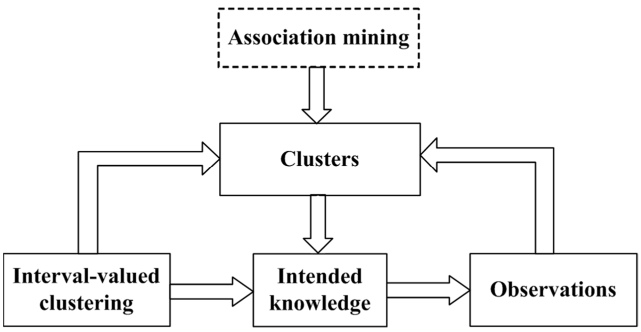

Figure 3. *Complex Information Mining Model.*

In Figure 3, "interval-valued clustering" denotes that the interval-valued clustering methods are adopted to cluster the objects into different clusters; "Clusters" denotes the generated classes; "Intended knowledge" denotes that the association mining methods are adopted to discover the associations among the clusters and finally form the intended knowledge; "Association mining" denotes the association mining methods; and "Observations" denotes the intended knowledge is observed and revised according to the domain knowledge.

Remarks

(1) Given *n* objects $x_1, x_2, \ldots x_n$, use the interval-valued clustering methods to cluster them into *m* classes $A_1, A_2, \ldots A_m$.

(2) According to $A_1, A_2, \ldots A_m$, transfer the related transaction dataset into the form of "class-data set":

$$\text{For any } x, \text{ if } x \in A_i, \text{ then replace } x \text{by} A_i. \tag{4}$$

For example, 16, 28, 87 $\Rightarrow$ A_1, A_2, A_3, where $16 \in A_1$, $28 \in A_2$ and $87 \in A_3$.

(3) Remove the redundant information in the dataset.

(4) Mine the association relation among $A_1, A_2, \ldots A_m$ in the dataset.

(5) According to the domain knowledge check the relation among $A_1, A_2, \ldots A_m$, merge the clear relation sets, and further revise the clusters $A_1, A_2, \ldots A_m$.

(6) Repeat the above steps (2) – (5) until the final clusters and association relation meet the actual situation.

3.2. Model Designs

For database DB, each row of DB represents an object and each column of DB represents one attribute of the object. Extract the attributes with computation feature of all the objects' attributes, and regard them as the computation attributes of interval-valued, and adopt the following formula to compute the similarities between the objects:

$$S_{\min}(O_i, O_j) = [\sum_{k=1}^{M}(O_{i,k}^{\min} - O_{j,k}^{\max})^2\,]^{-0.5},\ S_{\max}(O_i, O_j) = [\sum_{k=1}^{M}(O_{i,k}^{\max} - O_{j,k}^{\min})^2\,]^{-0.5} \tag{5}$$

Where, $[S_{\min}(O_i, O_j), S_{\max}(O_i, O_j)]$ is the similarity of the object o_i and the object o_j. $O_{i,k}^{\min}$ is the minimal value of the object O_i in k^{th} attribute, and $O_{i,k}^{\max}$ is the maximal value of the object O_i in k^{th} attribute, and the rest may be deduced by the same token. M is the number of common attributes of the object O_i and the object O_j.

Algorithm 1. Interval-valued Clustering Algorithm.

Input: S = $\{o_1, o_2, \ldots o_n\}$, the set of n objects; M, the number of attributes of the object; λ_0, the threshold of Interval-Value clustering.

Output: $[S_{\min}, S_{\max}]$, the minimal similarities between the objects and the maximal similarities of the objects; R, the similarities matrix.

Method:

(1) for any element o_i in S {
(2) for any element o_j in S {
(3) S1 = 0; S2 = 0;
(4) for k = 1 to M {
(5) xx1 ←*min*(o_i, k);// get the minimal value on the k^{th}

attribute related to oi
(6) yy1 ←*max*(o_i, k);// get the maximal value on the k^{th} attribute related to oi
(7) xx2 ←*min*(o_j, k);// get the minimal value on the k^{th} attribute related to oj
(8) yy2 ←*max*(o_j, k);// get the maximal value on the k^{th} attribute related to oj
(9) S1 += (xx1 – yy2) * (xx1 – yy2); S2 += (yy1 – xx2) * (yy1 – xx2);
(10) }
(11) R(S_{min}, S_{max}) ← [sqrt(S1), sqrt(S2)];
(12) } }
(13) for any element r_{ij} = [t_{ij}^-, t_{ij}^+] in R {
(14) ifλ_0 in r_{ij}
(15) r_{ij}←"#";
(16) else if $\lambda_0 < t_{ij}^-$
(17) r_{ij}←"×";
(18) else if $\lambda_0 > t_{ij}^+$
(19) r_{ij}←""; // r_{ij} is replaced by a space
(20) }
(21) for each "×" in R {
(22) A ←find the corresponding objects in the diagonal of R; }
(23) for each "#" in R {
(24) B ←find the corresponding objects in the diagonal of R;}
(25) for each *b* in B {
(26) *a*← compute the similarity related A under the condition λ_0;
(27) if $a \geq 0.5$
(28) A ←*b*;
(29) else *b* as a separated set;

Remarks. In Algorithm 1, firstly travel the *n* objects, and select any two objects from them to conduct the comparisons and compute their minimal similarity and maximal similarity. By *min*(o_i, k), compute the minimum of object o_i on the k^{th} attribute; by *max*(o_i, k), compute the maximum of object o_i on the k^{th} attribute, and others may be deduced by the same token. And then, the similarities between all the objects (including the maximal similarity and the minimal similarity) are stored in the similarities matrix R. Finally, conduct the interval-valued clustering process, as shown in the steps 13) – 29).

Suppose o_1, o_2, …are classified into A_1, A_2, … A_m, and then, according to the following algorithm, conduct the processing of interval-valued data mining.

Algorithm 2. Interval-valued Data Mining Algorithm.

Input: A = {A_1, A_2, … A_m}, the set of m classes; S = {o_1, o_2, … o_n}, the set of n objects; min_Supp, the threshold of Support; min_Conf, the threshold of Confidence; NUM, the number of repeating execution.

Output: AR, the association rules.

Method:
(1) for any element A_i in A {
(2) for any element o_j in S {
(3) if $o_j \in A_i$
(4) o_j← sign of A_i; // Replace o_j by the sign of A_i
(5) }}
(6) (6) for each element l_i in S { //Remove the redundant information in S;
(7) for each element o_j in l_i {
(8) if o_j first occurs
(9) mark o_j
(10) else delete o_j
(11) }}
(12) *AprioriL*(S,min_Supp, min_Conf); // *AprioriL* see also Ref.[14, 15]
(13) Cnt++;
(14) repeat (1) – (13) until Cnt > NUM;

Remarks. In Algorithm 2, firstly according to the elements in set A, transform the set S into the form of "class-data set", viz., for any element *x* in S, if $x \in A_i$, then replace *x* by A_i. To be followed up, conduct the reduction for the repeating elements in set S. Finally, call the algorithm*AprioriL*(S, min_Supp, min_Conf) to mine the association rules among A_1, A_2, …; *AprioriL*(S, min_Supp, min_Conf) is a data mining algorithm [12, 13, 14, 15], where S is the "class-data set", min_Supp is the threshold of Support, and min_Conf is the threshold of Confidence. NUM is the number of iterations.

4. Experiments

In this section, we will firstly present the experiment settings and steps, and then, the experimental results will be stated which will be able to validate the performance of the proposed model.

The experiments were performed on a ThinkPad T420s with Intel CORE i5-2520M Dual CPU and 4GB of RAM. The operating system was MS Windows XP Professional SP3, and the development tool was MS VStudio.

The experimental steps are as follows:
(1) Read the set of objects;
(2) Scan each attribute of the object, and work out the taking-value interval of each attribute;
(3) According to formula 5, compute the similarities between objects;
(4) According to Algorithm 1, cluster the objects into *m* classes;
(5) Replace the object by the class that the object belongs to, and form the set of "class-data";
(6) In the set of "calss-data", use *AprioriL* algorithm to mine the association rules;
(7) According to the domain knowledge, check the generated association relations between "classes";
(8) Combine the similar association relations of "classes";
(9) Apply the revised "classes" to the set of objects, and iteratively conduct the mining of association rules between "classes".

4.1. Experiment I: Mining Research of Aeronautical Objects Dataset

Aeronautical object dataset is provided by the bibliography [16, 17]. This dataset describes the behavioral streams generated by two fighters during the dogfight, where each data unit is a three-dimension vector, viz., the owner of behavior (number of the fighters), behavior (number of the maneuvers) and the time tag, shown as follows:

```
struct Time_Owner_Actions {
long OwnerID;//number of the fighter
int Act;// number of the maneuvers
int Time;// the time tag
```

In this experiment, the aeronautical object dataset being used includes 13140 records. Where, "number of the fighters" involves in two taking values, viz., 0 and 1; "number of the maneuvers" involves in 13 Basic Fighter Maneuvers (BFMs) that are "pursuit", "break pull", "right pull", "left pull", "break descend", "fighting turn", "high Yo-Yo", "low Yo-Yo", "half-loop roll", "flick half roll", "quick hover", "speed-up turn", "level flight", encoded by 1, 2, …. 13. "time tag" encodes the time slice, viz., 0001, 0002, …

Firstly read the dataset of aeronautical objects into memory, scan the involved owner ID, act Number and the time Number, and compute the intervals of their taking values. Table 1 shows the taking-value intervals of the involved attributes in the dataset of aeronautical objects.

Table 1. *Taking-value intervals of the aeronautical objects dataset.*

Attributes	Descriptions	Taking-value intervals
OwnerID	Number of fighters	OwnerID∈ [1, 2] and OwnerID∈ Z
Act	Number of maneuvers	Act∈ [1, 13] and Act∈ Z
Time	Time tag	Time∈ Z

In Table 1, "Attributes" denotes the involved attributes of each aeronautical object; "Descriptions" denotes the descriptions for the corresponding attribute; "Taking-value intervals" denotes the taking-value of the corresponding attribute.

According to Formula 5, compute the similarities between the aeronautical objects.

According to Algorithm 1, cluster the objects into different set. Table 2 shows the results of interval-valued clustering.

Table 2. *Interval-valued clustering results under different thresholds (Aeronautical objects dataset).*

λ_0	Number of sets	λ_0	Number of sets
0.1	1018	0.15	1018
0.2	225	0.25	225
0.3	225	0.35	225
0.4	221	0.45	82
0.5	82	0.55	82
0.6	82	0.65	74
0.7	72	0.75	6
0.8	4	0.85	1
0.9	1	0.95	1

In Table 2, "λ_0" denotes the threshold of interval-valued clustering, and "Number of sets" denotes the clustering number of aeronautical objects dataset under different thresholds.

Then, use the sets in Table 2 to replace the elements in the set. For example, A = {Object 1, Object 2, Object 5}, and then, in the aeronautical objects dataset, replace the object 1, object 2 and object 5 with A.

Arrange the aeronautical objects dataset. Reduce the redundant elements and get the set of "class-data".

For example:

Object 1, Object 2, Object 3
Object 4, Object 5
Object 1, Object 5

Where A = {Object 1, Object 2, Object 5}, B = {Object 3, Object 4}. And then:

A, A, B(reduction) →A, B
B, A(reduction) →A, B
A, B(reduction) →A, B

Use the association rule mining algorithm *AprioriL* to mine the set of "class-data". For example, A → B.

Finally, combine the similar association rules. Merge the similar classes, reduce the redundant and conduct the mining iteratively. For example, Class A is similar with Class C, and then, we can replace C with A, replace C→D with A→D and rid A→C.

Figure 4 shows in the function *AprioriL*, as the difference of min_Supp, the number of the mined association rules is also different.

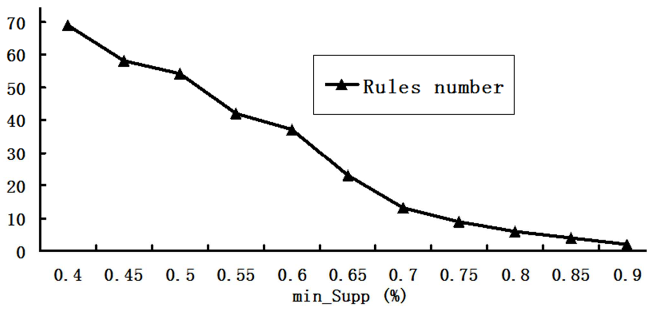

Figure 4. *Number of the mined rules under the different threshold min Supp.*

Figure 5 shows in the function *AprioriL*, as the difference of min_Conf, the number of mined association rules is also different.

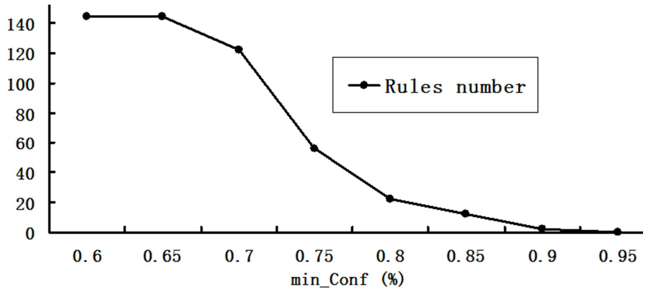

Figure 5. *Number of the mined rules under the different threshold min Conf.*

Figure 4 and Figure 5 show, by the method of interval-valued clustering, in the aeronautical objects dataset, the qualified (min_Supp, min_Conf) association rules can be mined.

By the experiments, we find some typical aeronautical objects association rules. Where, "low-speed Yo-Yo → pursuit" denotes that, if the maneuver low-speed Yo-Yo is conducted, then it implies the maneuver pursuit will be conducted; "left pull → right pull" denotes that, if the left pull is conducted, then the next maneuver will possibly be right pull; "half-loop turn → level fly" denotes that, the maneuver half-loop turn is conducted, then it implies the maneuver level fly will be conducted; “quick hover → fighting turn” denotes that, the maneuver quick hover is conducted, then it implies the maneuver fighting turn will be conducted.

In the field of aeronautics, these mined association rules are all with good reference value.

4.2. Experiment Two: Mining Research of Large Stock Dataset

Large stock dataset is from the software *DataAnalyzer* (http://www.fxj.net.cn/), and the dataset includes 20530 records, 54 attributes. And the missing values are filled with the mean-value of the same type of data; partial seemingly contradictory records are revised and deleted directly; non-numerical attributes are conducted the processing of “sampling”, normalizing qualification and encoding.

Firstly, read the large dataset into memory. Scan the taking-values of each attribute and make them into interval-valued.

To follow up, according to Formula 5, compute the similarities between records.

And then, according to Algorithm 1, cluster different records into different classes, and thus get the set of “class-data”.

Table 3 is the interval-valued clustering result of the large dataset under different clustering thresholds.

Table 3. *Interval-valued clustering results under different thresholds (stock dataset).*

λ_0	Number of classes	λ_0	Number of classes
0.1	753	0.15	753
0.2	445	0.25	445
0.3	36	0.35	31
0.4	13	0.45	1

In Table 3, “λ_0” is the threshold of interval-valued clustering, and when $\lambda_0 > 0.45$ the number of clustering is not larger than 1.

The classes in Table 3 are used to replace the association objects in the original dataset. For example, if A = {o_1, o_2, o_3}, then o_1, o_2 and o_3 will be replaced by A.

For the revised large dataset, the reduction process is conducted, and the purpose is to rid the redundant elements.

As mentioned above, the algorithm *AprioriL* is used to mine the “classes” association rules.

Finally, according to the domain knowledge, the similar classes and the similar association rules in the meaning are combined.

Repeat the replacements of classes, association rules mining and the combination of classes, and the process is iteratively conducted.

Figure 6 is the case of the final mined association rules as the changes of the Support thresholds when the iteration is terminated.

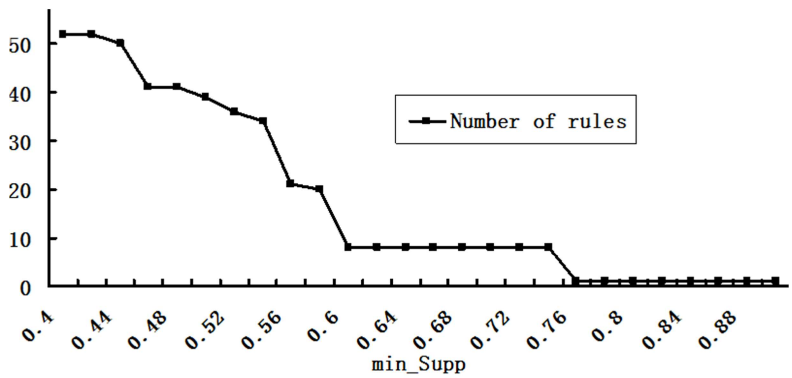

Figure 6. *Number of the mined association rules when the iteration is terminated (min_Supp).*

In Figure 6, when the threshold of Support changes from 0.4 to 0.9, the number of the rules discovered by the iterative data mining is shown. Where, the appearing flat line segment means the clustering is stabilized in certain number, and the mined number of rules is not affected by the threshold of Support in this case.

Figure 7 is the case when the iteration is terminated the final mined association rules change as the threshold of Confidence changes.

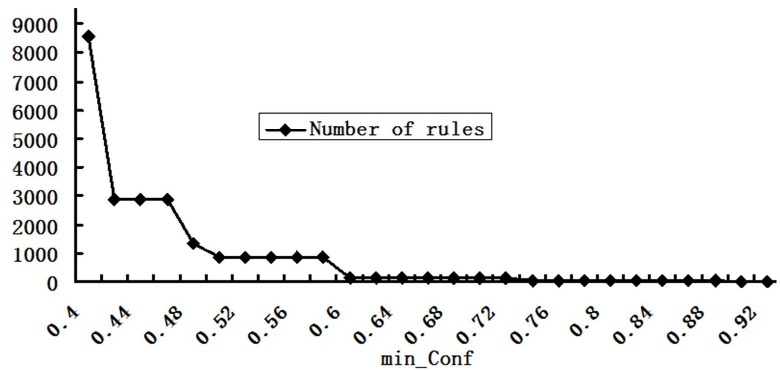

Figure 7. *Number of the mined association rules when the iteration is terminated (min_Conf).*

In Figure 7, it shows the case of the number of the mined association rules when the iteration is terminated. Where, different thresholds of Confidence produce certain effect to the number of the mined association rules: in min_Conf = 0.6 ~ 0.7 and min_Conf = 0.78 ~ 0.84, the mining generates two stable "points", viz., the number of the mined rules isstabilized in 144 and 47, respectively.

5. Conclusion

We have introduced a data mining method based on interval-valued clustering, as well as three interval-valued clustering methods. This sort of data mining method and the clustering methods are with significance for handling complex information. The train of thought for the method introduction is features analysis, properties research, explanation by examples, and design of algorithms. Finally, we conducted the experiment study for the proposed method in aeronautical objects dataset and large stock synthesized dataset. Experimental results show that the method in this paper has more advantage than the common methods for mining the complex information. The future research direction of this paper is to enhance the real-time and the efficiency of this method for mining the dynamically generated association rules.

Acknowledgements

The research was supported by Basic Research on the Frontier and Application of Chongqing City (cstc2015jcyjA40006); Postgraduate Educational Grant of Chongqing City (yjg20163062); and Educational Grant of Chongqing University (2016Y25).

References

[1] Y. J Liu, C. Y. Liang, F. Chiclana, J. Wu. A trust induced recommendation mechanism for reaching consensus in group decision making, *KNOWLEDGE-BASED SYSTEMS*, vol. 119, pp. 221-231, 2017.

[2] C. Y. Sheng, J. Zhao, W. Wang. Map-reduce framework-based non-iterative granular echo state network for prediction intervals construction, *NEUROCOMPUTING*, vol. 222, pp. 116-126, 2017.

[3] Y. Y. Zhang, T. R. Li, C. Luo. Incremental updating of rough approximations in interval-valued information systems under attribute generalization, *Information Sciences*, vol. 373, pp. 461-475, 2016.

[4] M. Ghiyasvand. A New Approach for Solving the Minimum Cost Flow Problem With Interval and Fuzzy Data, *International Journal Of Uncertainty Fuzziness And Knowledge-based Systems*, vol. 19, no. 1, pp. 71-88, 2011.

[5] Y. F. Yin, G. H. Gong, and L. Han. A Framework for Interval-valued Information System, *International Journal of Systems Science*, vol. 43, no. 9, pp. 1603-1622, 2012.

[6] Y. Matsubara, Y. Sakurai, C. Faloutsos. Ecosystem on the Web: non-linear mining and forecasting of co-evolving online activities, *WORLD WIDE WEB-INTERNET AND WEB INFORMATION SYSTEMS*, vol. 20, no. 3, pp. 439-465, 2017.

[7] B. Wu, X. K. Zhou, Q. Jin, F. H. Lin, H. Leung. Analyzing Social Roles Based on a Hierarchical Model and Data Mining for Collective Decision-Making Support, vol. 11, no. 1, pp. 356-365, 2017.

[8] D. F. Duncan, H. C. Kum, E. C. Weigensberg, F. A. Flair, and C. J. Stewart. Informing Child Welfare Policy and Practice Using Knowledge Discovery and Data Mining Technology via a Dynamic Web Site, *Child Maltreatment*, vol. 13, no. 4, pp. 383-391, 2008.

[9] Y. Cui, and V. K. N. Lau. Convergence-Optimal Quantizer Design of Distributed Contraction-Based Iterative Algorithms With Quantized Message Passing, *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5196-5205, 2011.

[10] A. M. Kibriya, and J. Ramon. Nearly Exact Mining of Frequent Trees in Large Networks, *Data Miningand Knowledge Discovery*, vol. 27, no. 3, pp. 478-504, 2013.

[11] G. Plonka, K. Wannenwetsch. A sparse fast Fourier algorithm for real non-negative vectors, *JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS*, vol. 321, pp. 532-539, 2017.

[12] Y. F. Yin, G. H. Gong, and L. Han. Theory and Techniques of Data Mining in CGF Behavior Modeling, *Science China Information Sciences*, vol. 54, no. 4, pp. 717-731, 2011.

[13] Y. F. Yin, G. H. Gong, and L. Han. A Weighted Dynamic Information SystemsReduction Method, *Intelligent Automation & Soft Computing*, DOI: 10.1080/10798587.2013.828907, 2013.

[14] N. Thanh-Long, V. Bay, V. Snasel. Efficient algorithms for mining colossal patterns in high dimensional databases, *KNOWLEDGE-BASED SYSTEMS*, vol. 122, pp. 75-89, 2017.

[15] Y. F. Yin, G. H. Gong, and L. Han. Experimental Study on Fighters Behaviors Mining, *Expert Systems With Applications*, vol. 38, no. 5, pp. 5737-5747, 2011.

[16] Y. F. Yin, and H. C Guan. Dynamic Software Testing and Evaluation with State Space Method, *Journal of Testing and Evaluation*, vol. 41, no. 3, pp. 403-408, 2013.

[17] Y. F. Yin, X. N. Wang, H. C. Guan, Y. F. Zeng, and B. L. Zhang. Online Joint Control Approach to FormationFlying Simulation, *IEEE Aerospace and Electronic Systems*, vol. 29, no. 6, pp. 24-36, 2014.