International Journal of Database Theory and Application
Vol.9, No.10 (2016), pp.87-98
http://dx.doi.org/10.14257/ijdta.2016.9.10.08

Teaching Resources Scheduling Method and Application of Data Mining Based on Association Rule

Chen Lin[1], Wang Yanxian[1], Tan Yang[1], Fang Song[1] and Liu Yan[1]

HUNAN RADIO & TV UNIVERSITY 410004 China
E-mail: 22206741@qq.com

Abstract

Traditional association rule mining method has the high redundancy of computation. Therefore, this paper proposes a kind of association rule algorithm of minimum single constraint based on post-processing closure operator. Firstly, it proposes association rule mining method of equivalence relation set based on closure operator constraint rule. It can meet the above minimum single constraint, maximum support and confidence coefficient threshold value effectively. In addition, it can divide constraint rule set into disjointed equivalence rule class. Secondly, it gives answers to questions and necessary and sufficient conditions of specific rule class existence. It can reduces redundant computation effectively and improve computation efficiency. At last, it verifies the validity of proposed algorithm through experimental contrast of standard test set.

Keywords*: Post-processing; Closure operator; Minimum single constraint; Association rules*

1. Introduction

In the premise of reducing storage and execution time, in order to realize more rapid response to the demand of users, constraint based data mining technology is widely studied [1]. At the beginning stage of algorithm research, related literatures utilizes data mining algorithm of original constraint. A typical case is frequent item set discovery constrained with the lowest frequent in transaction database. Minimum confidence coefficient is another basic constraint condition based association rule of frequent item set [2].

For traditional association rule algorithm, given model $T = (O, A, R)$, when T is big, the operating efficiency of algorithm is not high. In addition, in the situation of support and confidence coefficient constraint, users are difficult to position the interested rule subset rapidly [3]. In order to solve this problem, many literatures propose more complicated constraints. They make association rule directly relate to real demand of users so as to reduce the cost of data mining. For example, literature proposes monotonous and anti-monotonous constraints. They are expressed as C_m and C_{am} respectively; Literature [5] proposes minimum association algorithm based on two constraint conditions; Literature [6] proposes association rule based on both sides of item set. It's a kind of tri-phase data mining rule; Literature [7] proposes convertible constraint concept based on FP-growth algorithm; Literature [8] proposes data compression technique. It computes covetable constraint step by step; Literature [9] proposes multi-dimensional constraint solution algorithm design of association rule, *etc.*

Based on the thought of the above literatures, this paper raises improvement from the following three aspects:

The first is to propose association rule mining method based on closure operator constraint rule equivalence relation set. It can meet the above minimum single constraint, maximum support and confidence coefficient threshold value effectively. It can divide

ISSN: 2005-4270 IJDTA

constraint rule set $ARS_{\supseteq L_0,\supseteq R_0}(s_0,s_1,c_0,c_1)$ into disjointed equivalence rule class. The second is to give the solution of questions and necessary and sufficient conditions of specific rule's existence. It can reduce redundant computation effectively and improve computation efficiency. The third is to propose a new kind of constraint rule. It has the following advantages: (1) it can observe the structure of constraint rule more clearly; (2) it can eliminate repetition and avoid redundancy; (3) it can delete unnecessary constraint inspection and realize rapid extraction of constraint rule $L'\supseteq L_0$ and $R'\supseteq R_0$.

2. Question Description

Before describing question, the paper firstly gives some used common concepts and related terms. Given model $T=(O,\mathrm{A},R)$. Of them, O is nonempty set limited by object, A is object attribute and R is binary relation of $O\times A$. The cardinal number of O and A are expressed as $n=|O|$ and $m=|A|$ respectively. Set $X\subseteq \mathrm{A}$ is item set; support set of item set X can be expressed as $supp(X)$. Let s_0 and s_1 be maximum and minimum support threshold value. Of them, $0<1/n\le s_0\le s_1\le 1$, $n=|O|$. Nonempty item set A is frequent item set, meeting $s_0\le \sup p(A)\le s_1$ ($s_1=1$, then get traditional concept of frequent item set). For any frequent item set S', extract nonempty true subset L' from S': ($\varnothing\ne L'\subset S'$), and $R'\equiv S'/L'$. Then create rule $L'\to R'$ according to L', or according to support and confidence coefficient R' (or L', S'), get from the following formula [10]:

$$\begin{cases} supp(r)\equiv supp(L') \\ conf(r)\equiv supp(S')/supp(L') \end{cases} \quad (1)$$

Minimum and maximum confidence coefficient threshold values can be expressed as c_0 and c_1. Of them, $0<c_0\le c_1\le 1$. If it meets $c_0\le conf(r)$ and $s_0\le supp(r)$, rule r is called association rule in traditional method. All association rule set can be expressed as:

$$ARS(s_0,c_0)\equiv\begin{Bmatrix} r:L'\to R' \mid \varnothing\ne L',R'\subseteq \mathrm{A}, \\ L'\cap R'=\varnothing,\ S'\equiv L'+R', \\ s_0\le supp(r),c_0\le conf(r) \end{Bmatrix} \quad (2)$$

In the formula, $ARS(s_0,c_0)$ is association rule. The current researches on this question mainly focus on constraint support, confidence coefficient and subset, *etc*. For dual character of rule additional constraint, L_0, $R_0\subseteq \mathrm{A}$. The purpose is to find all association rules r : $L'\to R'$, so their confidence coefficient and support conform to condition: $s_0\le supp(r)\le s_1$, $c_0\le conf(r)\le c_1$. Item set constraint: $L'\supseteq L_0$, $R'\supseteq R_0$ is called minimum single constraint. Question form can be described as the following:

$$\begin{aligned} &ARS_{\supseteq L_0,\supseteq R_0}(s_0,s_1,c_0,c_1)\equiv \\ &\{r:L'\to R'\in ARS(s_0,s_1,c_0,c_1) \mid L'\supseteq L_0,R'\supseteq R_0\} \end{aligned} \quad (3)$$

Of them:

$$ARS(s_0,s_1,c_0,c_1)\equiv\begin{Bmatrix} r:L'\to R'\in ARS(s_0,c_0) \\ \mid supp(r)\le s_1,conf(r)\le c_1 \end{Bmatrix} \quad (4)$$

In order to discuss restricted problem, if $s_1=c_1=1$, $L_0=R_0=\varnothing$, it could be got association rule mining question $ARS(s_0,c_0)$ in traditional sense. The mined rule can be applied to all kinds of important fields like network traffic, *etc*.

3. Association Rules of Minimum Single Constraint

3.1. Rough Classification

In order to reduce repetition of candidate solution, set up equivalence rule with disjointed partition. Based on closure operator, propose equivalence relation of frequent item set $FS(s_0,s_1)$ and $ARS(s_0,s_1,c_0,c_1)$.

Definition 1: (equivalence relation of frequent item set $FS(s_0,s_1)$ and $ARS(s_0,s_1,c_0,c_1)$): (a) $\forall A,B \in FS(s_0,s_1)$, $B \Leftrightarrow h(A)=h(B)$, $A \square A$; (b) $\forall r_k : L_k \to R_k \in ARS(s_0,s_1,c_0,c_1)$, $r_2 \Leftrightarrow [h(L_1)=h(L_2), h(L_1+R_1)=h(L_2+R_2)]$, $k=1,2$, $r_1 \square r$ [11].

Let closure operator $FCS(s_0,s_1)=FS(s_0,s_1)\cap CS$. For $\forall L \in FCS(s_0,s_1)$, utilize equivalence relation formula $[L]_A \equiv \{\varnothing \neq L' \subseteq L, h(L')=L\}$ to express frequent item set equivalence class with the same closure L. $\forall L,S \in FCS(s_0,s_1)$, $\varnothing \neq L \subseteq S$, $supp(S)/supp(L) \in [c_0,c_1]$, equivalence class of all rules $r: L' \to R'$, then $h(L')=L$ and $h(L'+R')=S$ can be expressed as:

$$AR(L,S) \equiv \{r{:}L' \to R' \in ARS(s_0,s_1,c_0,c_1) \\ \left| L' \in [L]_A, S' \equiv L'+R' \in [S]_A \right. \} \tag{5}$$

Deduction 1 (rough partition) rough partition form of constraint rule set is as follows:

$$ARS_{\supseteq L_0,\supseteq R_0}(s_0,s_1,c_0,c_1)= \\ \sum\nolimits_{(L,S)\in NFCS(s_0,s_1,c_0,c_1)} AR_{\supseteq L_0,\supseteq R_0}(L,S) \tag{6}$$

In the formula, $AR_{\supseteq L_0,\supseteq R_0}(L,S) \equiv \{r{:}L' \to R' \in AR(L,S) \ \left| L' \supseteq L_0, R' \supseteq R_0^{(t)} \right. \}$

Based on Deduction1, generate post-processing algorithm PMM of $ARS_{\supseteq L_0,\supseteq R_0}(s_0,s_1,c_0,c_1)$. But, for limit value, $ARS_{\supseteq L_0,\supseteq R_0}(s_0,s_1,c_0,c_1)=\varnothing$, or for frequent closed itemset, $(L,S) \in NFCS(s_0,s_1,c_0,c_1)$, $ARS_{\supseteq L_0,\supseteq R_0}(L,S)=\varnothing$. When $\varnothing \neq AR_{\supseteq L_0,\supseteq R_0}(L,S)$, and $AR_{\supseteq L_0,\supseteq R_0}(L,S) \subseteq AR(L,S)$, the possibility of cardinal number of $AR(L,S)$ is still high and many redundant rules are existed, as is shown in Example 1.

Example 1: for data set T and corresponding closed itemset shown in Diagram 1, select minimum support threshold value $s_0=0.28$, maximum support threshold value $s_1=0.28$, maximum and minimum confidence coefficient threshold value are: $c_1=0.9$, $c_0=0.4$.

(a) Constraint Group 1: $L_0=c$, $R_0=f$. PMM Algorithm firstly generates $\left|ARS(s_0,s_1,c_0,c_1)\right|=134$ rules. But, through detection limit L_0 and R_0, it can be known, for arbitrary rule of 12 classes of $NFCS(s_0,s_1,c_0,c_1)$, all $AR(L,S)=\varnothing$, therefore, $ARS_{\supseteq L_0,\supseteq R_0}(s_0,s_1,c_0,c_1)=\varnothing$. Define $NFCS(s_0,s_1,c_0,c_1)$, the form is as follows:

$$NFCS(s_0,s_1,c_0,c_1) \equiv \{(L,S) \in CS^{\,2} \left| S \in FCS(s_0,s_1), \right. \\ \varnothing \neq L \subseteq S, supp(S)/supp(L) \in [c_0,c_1]\} \tag{7}$$

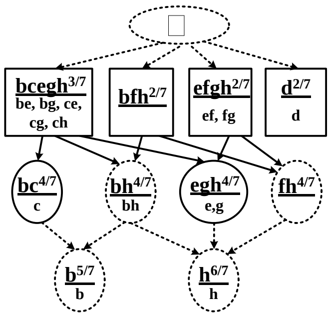
(b) Constraint Group 2: $L_0=h$, $R_0=b$, PMM algorithm firstly generates 134 rules. Through detection limit, it can be got $\left|ARS_{\supseteq L_0,\supseteq R_0}(s_0,s_1,c_0,c_1)\right|=19$ rules of four groups of rule classes (L,S) of $NFCS(s_0,s_1,c_0,c_1)$. Four groups of rule classes are $(egh,bcegh)$, $(h,bcegh)$, (fh,bfh) and (h,bh). For remaining $\left|NFCS(s_0,s_1,c_0,c_1)\right|-4=11$ groups of rule classes (L,S), algorithm will produce corresponding $\left|ARS(s_0,s_1,c_0,c_1) \setminus AR_{\supseteq L_0,\supseteq R_0}(s_0,s_1,c_0,c_1)=115\right.$ redundant candidate rules in total, then $AR_{\supseteq L_0,\supseteq R_0}(L,S)=\varnothing$. Consider class $(bc,bcegh) \in NFCS(s_0,s_1,c_0,c_1)$, then 21 candidate rules can be enumerated according to PMM algorithm. Even so, through detection limit condition $L_0 \subseteq L'$ and $R_0 \subseteq R'$, $ARS_{\supseteq L_0,\supseteq R_0}(s_0,s_1,c_0,c_1)=\varnothing$ also can be got.

(3) Constraint Group 3: $L_0 = f$, $R_0 = h$, only two pairs of 4 rules $\left(L^1 = fh, S^1 = efgh\right)$, $\left(L^2 = fh, S^2 = bfh\right)$ meet $AR_{\supseteq L_0,\supseteq R_0}\left(L^i, S^i\right) = \varnothing$, $i = 1,2$. For $\left(L^1 = fh, S^1 = efgh\right)$, 9 candidate rules are produced by $AR\left(L^1, S^1\right)$, only 3 rules meet limit $AR\left(L^1, S^1\right) = \{f \to eh, f \to egh, f \to gh\}$, the remaining 6 rules are redundant candidate rules.

In order to overcome these shortcomings, it is needed to find limit set L and (L,S), let $\left|ARS_{\supseteq L_0,\supseteq R_0}(s_0,s_1,c_0,c_1)\right| \neq \varnothing$ be necessary and sufficient condition. Therefore, we have another expression $AR^{+}_{\supseteq L_0,\supseteq R_0}(L,S)$ of $AR_{\supseteq L_0,\supseteq R_0}(L,S)$, and get better partition $ARS_{\supseteq L_0,\supseteq R_0}(s_0,s_1,c_0,c_1)$.

Trans	Items
1	bcegh
2	bcegh
3	bcd
4	defgh
5	bcefgh
6	bfh
7	fh

(a)



(b)

Figure 1. Data Set Sample

3.2. Constraint Rule Non-Empty Sufficient and Necessary Condition

Firstly, it gives nonempty necessary and sufficient conditions of constraint rule.

Theorem 1: (necessary condition) nonempty necessary condition of $ARS_{\supseteq L_0,\supseteq R_0}(s_0,s_1,c_0,c_1)$ and $AR_{\supseteq L_0,\supseteq R_0}(L,S)$ and equivalence expression of $AR_{\supseteq L_0,\supseteq R_0}(L,S)$.

(a) Necessary condition of $ARS_{\supseteq L_0,\supseteq R_0}(s_0,s_1,c_0,c_1) \neq \varnothing$:

If $r: L' \to R' \in ARS(s_0,s_1,c_0,c_1) \neq \varnothing$, then $(L,S) \in NFCS(s_0,s_1,c_0,c_1)$, $r \in AR(L,S) \neq \varnothing$. Of them, $L = h(L')$, $S = h(L'+R')$. Meet the following necessary condition: (H_1) $L_0 \cap R_0 = \varnothing$, $s_0^* \le s_1^*$, $supp\left(S_0^*\right) \ge s_0^*$, $supp(A) \le s_1^*$. Of them: $S_0^* \equiv L_0 + R_0$, $s_0^* \equiv \max\left(s_0; c_0.supp(C_1)\right)$, $s_1^* \equiv \min\left(s_1; c_1.supp(L_0)\right)$.

(b) Necessary condition of $AR_{\supseteq L_0,\supseteq R_0}(L,S) \neq \varnothing$. For $(L,S) \in NFCS(s_0,s_1,c_0,c_1)$, necessary condition of $r: L' \to R' \in AR(L,S) \neq \varnothing$ is $S \in FCS_{S \supseteq S_0^*}\left(s_0^*, s_1^*\right)$, $L' \in FS_{C_0 \subseteq L_{C_1}}$, $L_{C_1} \in FCS_{C_0 \subseteq C_1}\left(s_0', s_1'\right)$, $R' \in FS(S \setminus L')_{L', R_0^* \subseteq R_1^*}$. Of them, $FCS_{S \supseteq S_0^*}\left(s_0^*, s_1^*\right) \equiv \left\{S \in FCS\left(s_0^*, s_1^*\right) \middle| S \supseteq S_0^*\right\}$, $s_0' \equiv s_0'(S) \equiv supp(S)/c_1$, $s_1' \equiv s_1'(S) \equiv \min\left(1; supp(S)/c_0\right)$, $L_{C_1} \equiv L \cap C_1$, $FCS_{C_0 \subseteq C_1}\left(s_0', s_1'\right) \equiv \left\{L_{C_1} \equiv L \cap C_1\right\}$, $FS(S \setminus L')_{L', R_0^* \subseteq R_1^*} \equiv \left\{R' \supseteq R_0^* \middle| \varnothing \neq R' \subseteq R_1^*\right\}$.

(c) Equivalence expression of $AR_{\supseteq L_0,\supseteq R_0}(L,S)$: for $\forall (L,S) \in NFCS(s_0,s_1,c_0,c_1) \neq \varnothing$, then $FS_{C_0 \subseteq L_{C_1}} \neq \varnothing$ and $AR^{+}_{\supseteq L_0,\supseteq R_0}(L,S) = AR_{\supseteq L_0,\supseteq R_0}(L,S)$.

Deduction 2 (necessary and sufficient conditions) nonempty necessary and sufficient conditions of $ARS_{\supseteq L_0,\supseteq R_0}(s_0,s_1,c_0,c_1)$:

(a) In H_1, if at least one condition is not met, then $ARS_{\supseteq L_0,\supseteq R_0}(s_0,s_1,c_0,c_1) = \varnothing$.

(b) $r: L' \to R' \in ARS_{\supseteq L_0,\supseteq R_0}(s_0,s_1,c_0,c_1) \neq \varnothing \iff \exists\ (L,S) \in NFCS_{\supseteq L_0,\supseteq R_0}(s_0,s_1,c_0,c_1)$, $L' \in FS_{C_0 \subseteq L_{C_1}}$, $R' \in FS(S \setminus L')_{L', R_0^* \subseteq R_1^*}$, $r: L' \to R' \in AR^{+}_{\supseteq L_0,\supseteq R_0}(L,S) \neq \varnothing$.

Proof: In deduction 2(a) and (b), "$\Rightarrow$" can be got according to Theorem 1(a); In deduction 2(b), "$\Rightarrow$" can be got according to $AR_{\supseteq L_0,\supseteq R_0}(L,S) \subseteq ARS_{\supseteq L_0,\supseteq R_0}(s_0,s_1,c_0,c_1)$ and $AR^{+}_{\supseteq L_0,\supseteq R_0}(L,S) \subseteq AR_{\supseteq L_0,\supseteq R_0}(L,S)$.

3.3. Minimum Single Constraint Smooth Partition

Theorem 2: (constraint rule smooth partition) if limit condition H_1 meets, then get:

$$ARS_{\supseteq L_0,\supseteq R_0}(s_0,s_1,c_0,c_1) = \sum\nolimits_{(L,S)\in NFCS_{\supseteq L_0,\supseteq R_0}(s_0,s_1,c_0,c_1)} AR^{+}_{\supseteq L_0,\supseteq R_0}(L,S) \tag{8}$$

Through above rules, gain candidate class (L,S) from $NFCS(s_0,s_1,c_0,c_1)$ ($\supseteq NFCS_{\supseteq L_0,\supseteq R_0}(s_0,s_1,c_0,c_1)$), the quantity is still big and there are many redundant candidates that cannot meet constraint. The algorithm is shown in Pseudo code 1. The algorithm is mainly to find out frequent closed itemset $FCS_{C_0\subseteq C_1}(s_0',s_1')$; $FCS_{C_0\subseteq C_1}(s_0',s_1') = MFCS\_FL(LCG, S_0^*, A, s_0^*, s_1^*)$. Of them, $LCG \equiv \{(S,supp(S),G(S)) | (S,supp(S)) \in LC\}$.

Pseudo code 1: Constraint rule smooth partition

$FCS_{C_0\subseteq C_1}(s_0',s_1') = MFCS\_FL(LCG, C_0, C_1, s_0', s_1')$:

$FCS_{C_0\subseteq C_1}(s_0',s_1')$: $= \varnothing$;
if $(s_0' > s_1') \,||\, (C_0 \subseteq C_1) \,||\, (supp(C_0) < s_0') \,||\, (s_1' < supp(C_1))$ then
  return $\varnothing$;
endif
for each $((L, supp(L), G(L)) \in LCG^{s})$ ***do***
  if $(s_0' \le supp(L) \le s_1' \,\&\, L \supseteq C_0)$ ***then***
    if $(\exists L_i \in G(L) \,\&\, L_i \subseteq C_1)$ ***then***
      $L_{C_1} = L \cap C_1$; $G_{C_1}(L) = \{L_i \in G(L) | L_i \subseteq C_1\}$;
      $FCS_{C_0\subseteq C_1}(s_0',s_1') := FCS_{C_0\subseteq C_1}(s_0',s_1') + (L_{C_1}, supp(L_{C_1}))$;
    endif
  endif
endfor
return $FCS_{C_0\subseteq C_1}(s_0',s_1')$;

4. Equivalence Class Association Rules

4.1. Expansion Class Structure

Let $Y, X, Z_0 \subseteq A$, $Y \neq \varnothing$, $Y \cap X \neq \varnothing$, $Z_0 \subseteq Y$, then $FS(Y)_{X\supseteq Z_0} \equiv \{R' \supseteq Z_0 | \varnothing \neq R' \subseteq Y\}$ and suppose the following conditions $R_{\min} = Minimal$ $R_k \equiv S_k \setminus (X + Z_0)$, $S_k \in G(X+Y)$, $R_U^k \equiv \bigcup_{R_j \in R_{\min}, j \le k} R_j$. When $k \ge 1$, $R_{U,k} = R_U^{k-1} \setminus R_k$; when $k = 0$, $R_{U,k} = \varnothing$. Of them, $R_k \in R_{\min}$, $R_{-,k} \equiv Y \setminus (Z_0 + R_U^k)$.

It is noted that when $Y = \varnothing \,||\, Z_0 \neq \varnothing$, $FS(Y)_{X,\supseteq Z_0} \neq \varnothing$. Else:

$$\begin{aligned} FS^*(Y)_{X,\supseteq Z_0} \equiv \{ & R' \equiv Z_0 + R_k + R_k' + R_k^{\sim} | R_k \in R_{\min}, \\ & R_k' \subseteq R_{U,k}, R_k^{\sim} \subseteq R_{-,k}, (R_j \not\subset R_k + R_k', \forall R_j \in R_{\min} : \\ & 1 \le j < k)^{(*)}, R' \neq \varnothing \} \end{aligned} \tag{9}$$

Deduction 3 (expansion class structure) specific expression of $FS(Y)_{X,\supseteq Z_0}$, $\forall X, Y, Z_0 \subseteq A$: $Y \neq \varnothing$, $Y \cap X = \varnothing$, $Z_0 \cap X = \varnothing$, $Z_0 \subseteq Y$, then:

(a) $FS^*(Y)_{X,\supseteq Z_0} \neq \varnothing$ 。 (H_2) ;

(b) $FS(Y)_{X,\supseteq Z_0} = FS^*(Y)_{X,\supseteq Z_0}$ 。

For particular value of Y, X, Z_0 in $FS(Y)_{X\supseteq Z_0}$, it can be got $FS_{C_0 \subseteq L_{C_1}}$ and the structure of $FS(S \setminus L')_{L',R_0^* \subseteq R_1^*}$. It will be described in the following section.

4.2. $FS_{C_0 \subseteq L_{C_1}}$ and $FS\ (S \setminus L')_{L',R_0^* \subseteq R_1^*}$ Expression

Suppose class $(L,S) \in NFCS_{\supseteq L_0,\supseteq R_0}(s_0,s_1,c_0,c_1)$, $S \in FCS_{\supseteq S_0^*}(S_0^*,S_1^*)$, $L_{C_1} \in FCS_{C_0 \in C_1}(s_0',s_1')$, $L' \in FS_{C_0} \in L_{C_1}$, because $L_{C_1} \in FCS_{C_0 \in C_1}(s_0',s_1')$, then get $G_{C_1}(L) \neq \varnothing$ and $\exists L_i \in G(L)$, meet $\varnothing \subset L_i \subset L_{C_1} \,||\, L_{C_1} \neq \varnothing$.

Lemma 1: $\forall L, C_1 \subseteq A$, if $L_{C_1} \equiv L \cap C_1 \neq \varnothing \,||\, G_{C_1}(L) \neq \varnothing$, then $G(L_{C_1}) = G_{C_1}(L)$.

(a) When $Y \equiv L_{C_1}$, $X \equiv \varnothing \,||\, Z_0 \equiv C_0$. Owing to $Y \equiv L_{C_1} \in FCS_{C_0 \subseteq C_1}(s_0',s_1')$, then get $L \supseteq C_0$, $C_0 \subseteq L_{C_1}$ and $L_{C_1} \neq \varnothing$, and

$$
\begin{aligned}
FS(L_{C_1})_{\varnothing,\supseteq C_0} &= \{L' \supseteq C_0 \,|\, \varnothing \neq L' \subseteq L_{C_1}, h(L') \\
&= h(L_{C_1})\} \equiv FS_{C_0 \subseteq L_{C_1}}
\end{aligned} \tag{10}
$$

According to expression $R_{\min}$ and $FS^*(Y)_{X,\supseteq Z_0}$ in Section 3.1 and Lemma 1, it can be got $K_{\min} \equiv Min\{K_i \equiv L_i \setminus C_0, L_i \in G_{C_1}\{L\}\}$, $K_U^i \equiv \bigcup_{K_k \in K_{\min}, k \leq i} K_k$. If $i \leq 1$, then $K_{U,i} \equiv K_U^{i-1} \setminus K_i$; if $i = 0$, then $K_{U,i} \equiv \varnothing$ and $K_{-,i} \equiv L_{C_1} \setminus (C_0 + K_U^i)$, thus, the following result can be got:

$$
\begin{aligned}
FS^*_{C_0 \subseteq L_{C_1}} &\equiv \{L' \equiv C_0 + K_i + K_i' + K_i^{\sim} \,|\, K_i \in K_{\min}, \\
&K_i' \subseteq K_{U,i}, K_i^{\sim} \subseteq K_{-,i}, (K_k \not\subset K_i + K_i', \forall K_k \in K_{\min} : \\
&1 \leq k < i)^{(**)}, L' \neq \varnothing\}
\end{aligned} \tag{11}
$$

Owing to $L_{C_1} \neq \varnothing$ and according t Deduction 3(b), it can be got $FS^*_{C_0 \subseteq L_{C_1}} \neq \varnothing$.

(b) When $Y \equiv R_1^* = S \setminus L'$, $X \equiv L'$, $Z_0 \equiv R_0^* = R_0$ and suppose $S \setminus L' \neq \varnothing$, then according to the above definition, it can be got:

$$
\begin{aligned}
&FS(S \setminus L')_{L',\supseteq R_0^*} \equiv FS(S \setminus L')_{L',R_0^* \subseteq R_1^*} = \\
&\left\{R' \supseteq R_0^* \,\middle|\, \varnothing \neq R' \subseteq R_1^*, h(L' + R') = s\right\}
\end{aligned} \tag{12}
$$

Suppose $R_{\min} = Minimal$ { $R_k \equiv S_k \setminus (L' + R_0^*)$ $S_k \in G(S)$ }, $R_U^k \equiv \bigcup_{R_j \in R_{\min}, j \leq k} R_j$, when $k \geq 1$, $R_{U,k} = R_U^{k-1} \setminus R_k$; when $k = 0$, $R_{U,k} = \varnothing$. Of them, $R_k \in R_{\min}$, $R_{-,k} \equiv S \setminus (L' + R_0^* + R_U^k)$, the following result can be got:

$$
\begin{aligned}
FS^*(S \setminus L')_{L',R_0^* \subseteq R_1^*} &\equiv \{R' \equiv R_0^* + R_k + R_k' + R_k^{\sim} \,|\, R_k \\
&\in R_{\min}, R_k' \subseteq R_{U,k}, R_k^{\sim} \subseteq R_{-,k}, (R_j \not\subset R_k + R_k', \forall R_j \\
&\in R_{\min} : 1 \leq j < k)^{(*)}, R' \neq \varnothing\}
\end{aligned} \tag{13}
$$

According to $S \setminus L' \neq \varnothing$ and Deduction 3(a), the following result can be got: $FS^*(S \setminus L')_{L',R_0^* \subseteq R_1^*} \neq \varnothing$.

According to Deduction 3, the following result can be got:

Deduction 4: for equivalence class $AR^+_{\supseteq L_0,\supseteq R_0}(L,S)$, $\forall (L,S) \in NFCS_{\supseteq L_0,\supseteq R_0}(s_0,s_1,c_0,c_1)$, existing:

(a) $FS^*(S \setminus L')_{L',R_0^* \subseteq R_1^*}$ and $FS^*_{C_0 \subseteq L_{C_1}}$ have different elements;

(b) $FS(S \setminus L')_{L',R_0^* \subseteq R_1^*} = FS^*(S \setminus L')_{L',R_0^* \subseteq R_1^*}$, $FS_{C_0 \subseteq L_{C_1}} = FS^*_{C_0 \subseteq L_{C_1}}$, $FS^*_{C_0 \subseteq L_{C_1}} \neq \varnothing$;

(c) $\forall L' \in FS^*_{C_0 \subseteq L_{C_1}}$, there is $FS^*(S \setminus L')_{L',R_0^* \subseteq R_1^*} \neq \varnothing$ existed, it is equal to $S \setminus L' \neq \varnothing$.

3.3 $AR^+_{\supseteq L_0,\supseteq R_0}(L,S)$ structure expression

For $\forall (L,S) \in NFCS_{\supseteq L_0,\supseteq R_0}(s_0,s_1,c_0,c_1)$, all of them:

$$\begin{cases} AR^{+}_{\supseteq L_0,\supseteq R_0}(L,S) \equiv \{r : L' \to R' \big| L' \in FS^{*}_{C_0 \subseteq L_{C_1}} \\ FS^{*}_{C_0 \subseteq L_{C_1}}(S) \equiv \left\{L' \in FS^{*}_{C_0 \subseteq L_{C_1}} \big| S \setminus L' \neq \varnothing\right\} \end{cases} \tag{14}$$

According to Deduction 4, Deduction 5 can be got:

Deduction 5: nonempty necessary and sufficient conditions of $AR^{+}_{\supseteq L_0,\supseteq R_0}$. For $\forall (L,S) \in NFCS_{\supseteq L_0,\supseteq R_0}(s_0,s_1,c_0,c_1)$, there is:

(a) $AR^{+}_{\supseteq L_0,\supseteq R_0}(L,S) = AR^{*}_{\supseteq L_0,\supseteq R_0}(L,S)$;

(b) $AR^{*}_{\supseteq L_0,\supseteq R_0}(L,S) \neq \varnothing \Leftrightarrow FS^{*}_{C_0 \subseteq L_{C_1}}(S) \neq \varnothing$;

(c) $AR^{*}_{\supseteq L_0,\supseteq R_0}(L,S) = \sum_{L' \in FS^{*}_{C_0 \subseteq L_{C_1}}} \{r : L'R' : R' \in \ FS^{*}(S \setminus L')_{L', R_0^{*} \subseteq R_1^{*}}\}$ 。

According to Theorem 2 and Deduction3, Theorem 3 can be got:

Theorem 3: suppose it meet H_1 limit, the following result can be got:

$$\begin{cases} ARS_{\supseteq L_0,\supseteq R_0}(s_0,s_1,c_0,c_1) = \\ \sum_{(L,S) \in NFCS_{\supseteq L_0,\supseteq R_0}(s_0,s_1,c_0,c_1)} AR^{*}_{\supseteq L_0,\supseteq R_0}(L,S) \\ AR^{*}_{\supseteq L_0,\supseteq R_0}(L,S) = \sum_{L', FS^{*}_{C_0 \subseteq L_{C_1}}} \{r : L' \\ \quad \to R' : R' \in FS^{*}(S \setminus L')_{L', R_0^{*}, R_1^{*}}\} \end{cases} \tag{15}$$

Algorithm process is shown as Pseudo code 2.

Pseudo code 2: Constraint rule generation

$AR^{*}_{\supseteq L_0,\supseteq R_0}(L,S) MAR-Minsc\left(C_0, L_{C_1}, G(L_{C_1}), R_0^{*}, S, G(S)\right)$:

$AR^{*}_{\supseteq L_0,\supseteq R_0}(L,S) = \varnothing$; $FS^{*}(Y)_{X \supseteq Z_0}$;
for each ($L' \in FS^{*}_{C_0 \subseteq L_{C_1}}$ && $S \setminus L' \neq \varnothing$)***do***
  $R_1^{*} = S \setminus L'$; $FS^{*}(S \setminus L')_{L', R_0^{*} \subseteq R_1^{*}}$;
  for each $R' \in FS^{*}(S \setminus L')_{L', R_0^{*} \subseteq R_1^{*}}$ ***do***
    $AR^{*}_{\supseteq L_0,\supseteq R_0}(L,S) = AR^{*}_{\supseteq L_0,\supseteq R_0}(L,S) + \{r : L' \to R'\}$;
  endfor
endfor
return $AR^{*}_{\supseteq L_0,\supseteq R_0}(L,S)$;

$ARS_{\supseteq L_0,\supseteq R_0}(s_0,s_1,c_0,c_1) MAR-Minsc1(s_0,s_1,c_0,c_1,L_0,R_0,LCG)$

$ARS_{\supseteq L_0,\supseteq R_0}(s_0,s_1,c_0,c_1) = \varnothing$;
if $(s_0 > s_1 \| c_0 > c_1 \| L_0 \cap R_0 \neq \varnothing)$ do
  return $\varnothing$;
end
$S_0^{*} = L_0 \cup R_0$; $C_0 = L_0$; $C_1 = A \setminus R_0$; $R_0^{*} = R_0$;
$s_0^{*} \equiv \max(s_0; c_0.supp(C_1))$, $s_1^{*} \equiv \max(s_1; c_1.supp(L_0))$;
$FCS_{\supseteq S_0^{*}}(s_0^{*}, s_1^{*}) = MFCS\_FL(LCG, S_0^{*}, A, s_0^{*}, s_1^{*})$;
for each $\left((S, supp(S), G(S)) \in FCS_{\supseteq S_0^{*}}(s_0^{*}, s_1^{*})\right)$ do
  $s_0' = supp(S)/c_1$; $s_1' = \min(1; supp(S)/c_0)$;
  for each $\left((L_{C_1}, supp(L_{C_1}), G_{C_1}(L)) \in FCS_{C_0 \subseteq C_1}(s_0', s_1')\right)$ do
    $AR^{*}_{\supseteq L_0,\supseteq R_0}(L,S) = MAR-Minsc\left(C_0, L_{C_1}, G_{C_1}(L), R_0^{*}, S, G(S), S\right)$;
    $ARS_{\supseteq L_0,\supseteq R_0}(s_0,s_1,c_0,c_1) = ARS_{\supseteq L_0,\supseteq R_0} + AR^{*}_{\supseteq L_0,\supseteq R_0}(L,S)$;
  end for
end for

From Pseudo code 2, it can be seen that in the above candidate rule generation process, it can avoid the generation of redundancy, so $MAR-Minsc$ algorithm is more efficient.

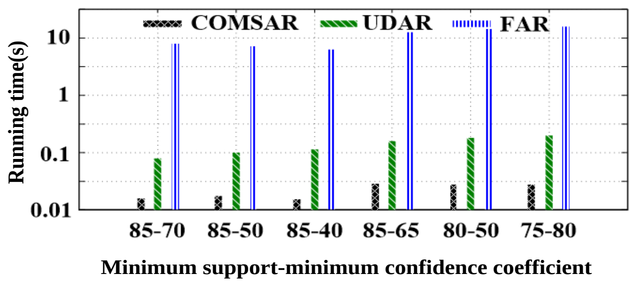
5. Experiment and Analysis

Experimental hardware setting: CPU i3-2440k 2.2GHz and CPU i3-2440k 2.2GHz. in order to verify the effectiveness of proposed algorithm, select contrast algorithm: fuzzy association rule algorithm (FAR) in Literature [12] and uniform distribution uncertain association rule algorithm (UDAR) Literature [13]. In order to evaluate the performance of proposed algorithm, make contrast of the performance of COMSAR algorithm and two kinds of contrast algorithms in Literature [12] and Literature [13]. Meanwhile select 5 kinds of benchmark database used in Literature [14] as test object. Of them, Connect, Mushroom, Pumsb and Chess are density data set. They have many high support of long frequent itemset. T10I4D100K is complex sparse dataset. Benchmark test dataset information is shown in Diagram 1.

For every test database, threshold value s_1 and c_1 are 0.9 and 0.95. For given minimum confidence coefficient (MC) and (MS), the value range of limit condition L_0 and R_0 is $4\%|A^F| \sim 22\%|A^F|$, value interval is $2\%|A^F|$. For $L_0's$ and $R_0's$, select 10 groups of (L_0, R_0), every (L_0, R_0) includes service item extracted from A^F. The experiment in the above database shows, the forecast output result of FAR, UDAR and COMSAR is uniform. Contrast result in runtime is as shown in Diagram 2-6.

Table 1. Benchmark Test Dataset Information

Dataset	Itemset quantity	Total record	Average length
Connect	132	67894	42
Mushroom	124	8345	26
Pumsb	7236	49126	72
Chess	87	3358	36
T10I4D100K	1127	100976	44



Minimum Support-Minimum Confidence Coefficient

Figure 2. Connect Database Runtime

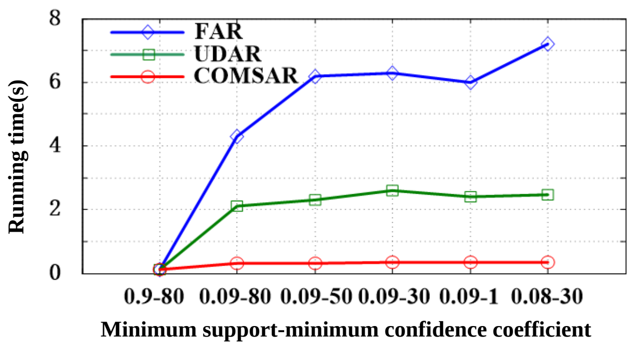


Figure 3. Mushroom Database Runtime

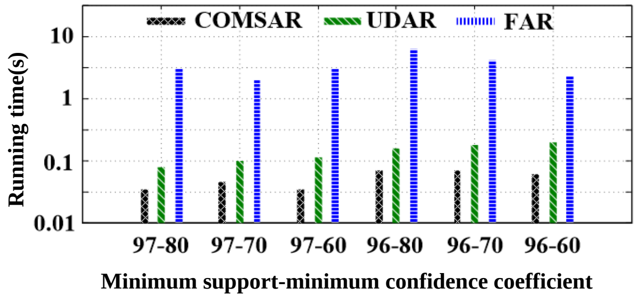


Figure 4. Pumsb Database Runtime

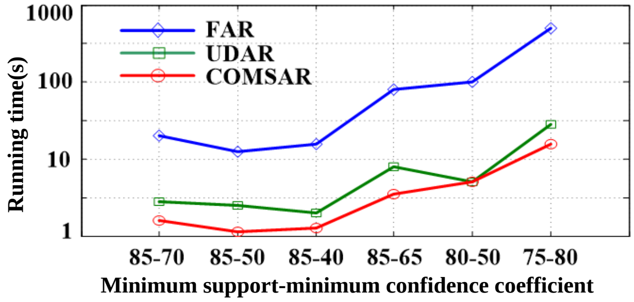


Figure 5. Chess Database Runtime

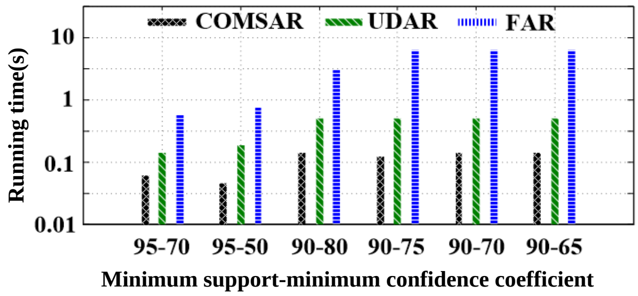


Figure 6. T10I4D100K Database Runtime

Diagram 2 to Diagram 6 show contrast result of runtime of FAR, UDAR and COMSAR in five kinds of test database Connect, Mushroom, Pumsb, Chess and T10I4D100K. These databases include density database and artificial sparse database. From contrast result, the operation efficiency of COMSAR algorithm in the above five database is superior to FAR and UDAR. It realizes the purpose of solving redundant problem at the beginning of algorithm design. It improves computation efficiency of association rule algorithm effectively.

Table 2. Time Reduction (%)

DB-MS-MC	R-ovPP2	R-ovPP1	RR-PP2	RR-PP1
Co-97-80	45.51	0.46	94.51	98.72
Co-96-60	43.26	0.27	99.81	99.93
Ch-85-50	33.61	6.85	99.93	99.94
Ch-80-35	20.37	4.71	99.92	99.94
P-95-80	48.24	1.52	98.58	99.61
P-90-65	21.12	0.13	99.88	99.95

M-30-70	20.42	0.26	99.97	99.97
M-25-45	17.75	0.28	99.94	99.97

Table 2 gives several groups of simulation result of time reduction. The related symbols in the Table are defined as the following: DB-MS-MC expresses database-minimum support-minimum confidence coefficient combination; R-ovPP2 expresses the rate of runtime of COMSAR algorithm and that of UDAR algorithm; R-ovPP1 expresses that rate of runtime of COMSAR algorithm and that of FAR algorithm; RR-PP2 expresses the rate of redundant rule quantity that UDAR algorithm cannot meet constraint condition and total rules; RR-PP1 expresses that the rate of redundant rule quantity that FAR cannot meet constraint condition and total rules.

Experimental result in Table 2 shows that, compared with FAR and UDAR, COMSAR algorithm is superior to contrast algorithm in operation speed. The main reason can be seen from two rows RR-PP2 and RR-PP1. The main reason that FAR 和 UDAR operate slowly is too many redundant rules of algorithm.

6. Conclusion

In the view of solving redundant rule problem of association rule algorithm, this paper applies post-processing minimum single constraint method to realize pre-judgment and pre-processing of association rules, reduces the impact of redundant rule on operation efficiency of algorithm. Through making contrast with FAR and UDAR on five kinds of test databases of Connect, Mushroom, Pumsb, Chess and T10I4D100K, it shows high operation efficiency of COMSAR algorithm proposed in this paper.

References

[1] Z. Lv, T. Yin and Y. Han, "WebVR-web virtual reality engine based on P2P network", Journal of Networks, vol. 6, no. 7, **(2011)**, pp. 990-998.
[2] D. Jiang, Z. Xu and Z. Lv, "A multicast delivery approach with minimum energy consumption for wireless multi-hop networks", Telecommunication Systems, **(2015)**, pp. 1-12.
[3] Y. Lin, J. Yang and Z, Lv, "A self-assessment stereo capture model applicable to the internet of things", Sensors, vol. 15, no. 8, **(2015)**, pp. 20925-20944.
[4] Y. Liang, "Satisfaction with Economic and Social Rights and Quality of Life in a Post-Disaster Zone in China: Evidence From Earthquake-Prone Sichuan", Disaster Medicine and Public Health Preparedness, vol. 9, no. 2, pp. 111-118.
[5] Y. Liang, "Correlations Between Health-Related Quality of Life and Interpersonal Trust: Comparisons Between Two Generations of Chinese Rural-to-Urban Migrants, Social Indicators Research, vol. 123, no. 3, pp. 677-700.
[6] Y. Liang and P. Lu, "Medical insurance policy organized by Chinese government and the health inequity of the elderly: longitudinal comparison based on effect of New Cooperative Medical Scheme on health of rural elderly in 22 provinces and cities", International Journal for Equity in Health, DOI:10.1186/1475 -9276-13-37, vol. 13, no. 37, **(2014)**, pp. 1-15.
[7] J. Yang, J. Zhou and Z. Lv, "A Real-Time Monitoring System of Industry Carbon Monoxide Based on Wireless Sensor Networks", Sensors, vol. 15, no. 11, **(2015)**, p. 29535-29546.
[8] D. Jiang, X. Ying and Y. Han, "Collaborative multi-hop routing in cognitive wireless network", Wireless Personal Communications, vol. 86, no. 2, **(2016)**, pp. 901-923.
[9] Z. Lv, A. Halawani and S. Feng, "Multimodal hand and foot gesture interaction for handheld device", ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), vol. 11, no. 1, **(2014)**.
[10] J. Hu and Z. Gao, "Distinction immune genes of hepatitis-induced heptatocellular carcinoma", Bioinformatics, vol. 28, no. 24, **(2012)**, pp. 3191-3194.
[11] J. Hu, Z. Gao and W. Pan, "Multiangle Social Network Recommendation Algorithms and Similarity Network Evaluation", Journal of Applied Mathematics, vol. 2013, **(2013)**.
[12] J. Hu and Z. Gao, "Modules identification in gene positive networks of hepatocellular carcinoma using Pearson agglomerative method and Pearson cohesion coupling modularity", Journal of Applied Mathematics, **(2012)**.

[13] Y. Geng, J. Chen, R. Fu, G. Bao and K. Pahlavan, "Enlighten Wearable Physiological Monitoring systems: On-Body RF Characteristics Based Human Motion Classification Using a Support Vector Machine", no. 99, pp. 1-16.
[14] X. Song and Y. Geng, "Distributed Community Detection Optimization Algorithm for Complex Networks", Journal of Networks, vol. 9, no. 10, pp. 2758-2765.
[15] K. Pahlavan, P. Krishnamurthy and Y. Geng, "Localization Challenges for the Emergence of the Smart World", Access, IEEE, vol. 3, no. 1, pp. 1-11.
[16] J. He, Y. Geng, Y. Wan, S. Li and K. Pahlavan, "A cyber physical test-bed for virtualization of RF access environment for body sensor network", Sensors Journal, IEEE, vol. 13, no. 10, **(2013)**, pp. 3826-3836.

Authors

Chen Lin, was born in Xiangtan, Hunan, China, 1980. She graduated from Hunan Normal University in 2003 and graduated from computer science and technology. In 2012, she was awarded a master's degree in engineering from the University of Electronic Science and technology. Her research interest is mainly in the area of Computer Software. She has published several research papers in scholarly journals in the above research areas and has participated in several books. She is currently a lecturer the College of Hunan Network Engineering Vocational College, Changsha, china.

