16th World Congress of the International Fuzzy Systems Association (IFSA)
9th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT)

Using imprecise user knowledge to reduce redundancy in Association Rules

Julio Diaz[1] **Carlos Molina**[2] **M. Amparo Vila**[3]

[1]University of Informatics Sciences, Havana, Cuba
[2]Jaen University, Jaen Spain
[3]Granada University, Granada Spain

Abstract

Redundancy is a handicap in association rules. It becomes a limitation to use rules models in order to support the decision-making process. A technique based on user knowledge has been proposed recently , which aims at eliminating redundancy. However, it ignores the imprecise nature of knowledge. In this paper, the notion of knowledge redundancy is generalized and a method to propagate the user certainty over derivate rules is developed. Certainty factor models are used. Obtained results have shown a model reduction of 50% with previous knowledge below 3%. This method improves the efficiency of association rules and the use of discovered association rules.

Keywords: Association Rules, Imprecise Knowledge, Knowledge Based Redundancy

1. Introduction

Association rules deal with the discovery of relationships and correlations between attributes in large databases. They are one of the most applied and studied data mining models. However, several problems make difficult to use them in order to support decision-making process:

- A large number of rules are discovered.
- Not all discovered rules are interesting.
- Rule discovery algorithms have a high computational cost.

Several investigations have been conducted to reduce the negative impact of these problems. A major line of research aims at improving the performance of mining algorithms. Most of them focus on developing new methods to prune the search space or define more efficient data structures and more effective organization [1], [2], [3], [4].

Another important research line has studied the output reduction problem following two main approaches:

- The generation of compact representation called bases [5], [6], [7], [8].
- The use of constraints [9], [10], inside the extraction algorithms to limit the rules that must appear in the final model.

The last approach requires extraction algorithms to be executed for each user with different constraints. This condition overload the computational cost.

Several researches have focused on defining objective interest measures [11], [12], [13], [14]. However, researchers are increasingly paying attention to subjective measures based on user controlled factors.

The presence of redundant rules is a factor that negatively affects association rules model's size. *An association rule is redundant if it conveys the same information, or less general information, than the information conveyed by another rule of the same usefulness and the same relevance* [15]. Most papers about redundancy in association rules try to obtain bases using some criteria associated with rules' structure and objective measures. Such approaches allow evaluating rule's relevancy but it is practically impossible to assess the usefulness. This is why it depends on the specific problem and the user knowledge about the domain.

If a rule in the model is known by the user, or can be derived directly from what he/she already knows, then the rule would be useless. Each user may have different knowledge's level about the domain, so the previous known rules' set can be different for each user. A definition of redundancy in association rules was proposed in [1]. It focused on user's ability to express rules as domain prior knowledge and specific task to perform. Besides, an algorithm to detect and eliminate redundant rules was developed. However, it does not take into account how imprecise human's knowledge is.

This paper aims to generalize the prior knowledge redundancy elimination method in order to work with imprecise knowledge. It will produce models which are closer to users. The way different extraction objective measures affects the redundancy elimination procedure is also verified.

The paper is organized as follows: Section 2 discusses knowledge based redundancy in association rules. In section 3, problem statement is presented. In section 4, knowledge redundancy is generalized and the certainty factor models is used in order to propagate certainty over derivate rules. In section 5, the solution is tested with three datasets. Section 6 presents conclusions reached.

2. Knowledge based redundancy in association rules

Let I e a particular itemset and T a set of transactions, each of them subset of I. An association rule is an implication of the form $X \rightarrow Y$ whose meaning is the involvement of both (X and Y) in some transaction in T which also fulfils: $X, Y \subset I$, $X \cap Y = \emptyset$ and $X, Y \neq \emptyset$.

The quality of the rule is evaluated using different metrics. The most widespread among them are support, which represents the probability of appearing in $T : X \cup Y$, and confidence whose meaning is the conditional probability $p(Y|X)$.

2.1. Prior knowledge redundancy

Knowledge based redundancy was defined in [1]. Let S be an association rules set and S_c a set of previous known rules, S and S_c are defined over the same domain. An association rule $R : X \rightarrow Y$ is redundant respect to S_c if there is a rule $R^{'} : X^{'} \rightarrow Y^{'}$ and fulfils some of the following conditions:

1. $X^{'} \subseteq X \wedge Y^{'} \subseteq Y$
2. $X^{'} \subseteq X \wedge \exists R^{''} : X^{''} \rightarrow Y^{''} \in S_c \wedge X^{''} \subseteq Y' \wedge Y \subseteq Y^{''}$
3. $X^{'} \subseteq X \wedge Y \subseteq Y^{'}$
4. $X^{'} \subseteq X \wedge Y^{'} \subseteq X$
5. $X^{'} \subseteq Y \wedge Y^{'} \subseteq Y$

Two algorithms were proposed. Both use an inference mechanism based on Armstrong's axioms. Despite Armstrong's axioms cannot be used to infer association rules because they do not guarantee the confidence threshold for inferred rules [16]. In this case, they are used to assess if a rule has prior knowledge redundancy. Due to the support's descendant closure property, each reduced rule fulfils the support threshold and no new rules are generated.

The first algorithm is used to identify and eliminate redundant items in a rule (see figure 1) and the second one is used to determine if the entire rule is redundant (see algorithm 2).

2.2. Previous knowledge

Prior knowledge, as stated in this paper, consists of relationships between attributes of the domain, which are previously known by the user. They are the result of expert's experience at the work area. Therefore, this knowledge is considered to be in somehow more comprehensive than the extracted rules of a particular dataset, which contains partial information. User can represent previous knowledge in different ways such as semantic networks, ontologies, and some others.

Previous knowledge is incorporated to the model using association rules format, based on the fact that the expert is interested in association rules discovering. For example, an expert knows customers with high income ([*income*].[*high*]) pay their loans in time and therefore their applications must be approved. This knowledge can be represented as the association rule [*income*].[*high*] $\rightarrow$ [*loan*].[*yes*].

An important element to be considered is the imprecise nature of expert's knowledge. This requires adding a degree of certainty into user knowledge representation. The knowledge expressed in the previous examples is represented in a more natural way as 80% of high income customers pay their loans in time. The association rules representing this knowledge requires a value of certainty [*income*].[*high*] $\rightarrow$ [*loan*].[*yes*] certainty= 0.8.

Require: Set of previous knowledge rules S_c
A rule R_i in form $X \rightarrow Y$
Ensure: Rewritten rule $R_i^{'}$

```
F = S_c ∪ R_i
for all item A ∈ X do
    if (X − {A})^+ overF = (X − {A})^+ over((F − {R_i}) ∪ (X − {A}) → Y) then
        R_i = X − {A} → Y
    end if
end for
if conf(R_i) ≥ treshold then
    prune(R_i)
end if
for all item W ∈ Y do
    if (X^+ overF = X^+ over((F − {R_i}) ∪ X → Y − {W}) then
        R_i = X → Y − {W}
    end if
end for
return R_i
```

Figure 1: Algorithm for rewriting rules

Require: Set of previous knowledge rules S_c
A rule R_i in form $X \rightarrow Y$
Ensure: Boolean value to indicate if rule must be pruned

```
F = S_c
if Y ∈ (X^+)overF then
    return true
else
    return false
end if
```

Figure 2: Algorithm for punning rules

3. Problem statement

Let D be a database, A a technique for association rules mining over D and S_c a previous knowledge representation containing the degree of certainty for each rule. The rules set R contains the mined rules. A subset $R^{'}$ of R contains the rules that can be derived from S_c, so they are redundant rules. It is necessary to pay attention to two important facts:

1. The same rules' model R may have different

redundant models $R^{'}$ associated to users with different previous knowledge.
2. The user's knowledge can be modified within the process; therefore determining the redundant rules is a dynamic an interactive procedure.

The set of potential interesting rules is $\{R - R^{'}\}$. It usually is remarkably smaller than R, so it is desired to show only these rules.

The problem of removing redundancy from the prior knowledge of association rules is defined as: given a set of association rules R and user previous knowledge S_c find the set of non-redundant rules at a given time.

Some techniques are used to deal with certainty degree in user's knowledge, in order to solve two concrete situations:

1. The certainty propagation from user's knowledge to derived rules.
2. A certainty threshold to consider a derived rule as a non-redundant one.

4. Knowledge based redundancy elimination in presence of imprecise knowledge

It is possible to develop a rule-based system, from redundancy conditions given in knowledge based redundancy definition, to determine whether a rule is redundant or not. Each condition stated in the definition will be a rule in the rule-based system:

1. If $X^{'} \subseteq X \wedge Y^{'} \subseteq Y$ then $X \rightarrow Y$ is redundant.
2. If $X^{'} \subseteq X \wedge \exists R^{''} : X^{''} \rightarrow Y^{''} \in S_c \wedge X^{''} \subseteq Y' \wedge Y \subseteq Y^{''}$ then $X \rightarrow Y$ is redundant.
3. If $X^{'} \subseteq X \wedge Y \subseteq Y^{'}$ then $X \rightarrow Y$ is redundant.
4. If $X^{'} \subseteq X \wedge Y^{'} \subseteq X$ then $X \rightarrow Y$ is redundant.
5. If $X^{'} \subseteq Y \wedge Y^{'} \subseteq Y$ then $X \rightarrow Y$ is redundant.

The certainty factor (CF) model [17] is proposed to manage the certainty propagation. It defines a group of functions to combine the CF of rules:

- Parallel combination

$$CF_{AB} = \begin{cases} CF_A + CF_B - CF_A * CF_B & CF_A, CF_B \geq 0 \\ CF_A + CF_B + CF_A * CF_B & CF_A, CF_B \leq 0 \\ \frac{CF_A + CF_B}{1 - min(|CF_A|, |CF_B|)} & other \end{cases}$$

- Serial combination

$$CF_{AB} = \begin{cases} CF_A * CF_B & CF_A \geq 0 \\ 0 & CF_A \leq 0 \end{cases}$$

- Conjunction of evidence
  $CF_{A\&B} = min(CF_A, CF_B)$
- Disjunction of evidence
  $CF_{A|B} = max(CF_A, CF_B)$

Parallel combination is used to combine the CF of two rules with the same hypothesis. Serial combination is used to combine two rules when the hypothesis of one rule is the evidence of the other rule.

TID	Items
1	[unemployed.no], [income.high], [loan.no]
2	[balance.high], [income.high], [loan.yes]
3	[unemployed.no], [balance.high], [income.high], [loan.yes]
4	[balance.high], [loan.yes]
5	[unemployed.no], [balance.high], [income.high], [loan.yes]

Table 1: Transactional database

Conjunction of evidence is used when the rule's evidence is a conjunction of elements. The disjunction of evidence is used when the rule's evidence is a disjunction of elements.

The rule's CF value is the multiplication of rule evidence by rule CF. Propagation of CF is illustrated in the following example:

Table 1 represents a set of transaction, the first column contains a transaction identifier and the second one keeps a list with items in the transaction.

Assuming prior knowledge $S_c = \{[balance.high] \rightarrow [income.high]cf = 0.8, [income.high] \rightarrow [loan.yes]cf = 0.75\}$. With a threshold of 40% for support and confidence the rule $\{[balance.high][income.high] \rightarrow [loan.yes]\}$ is part of the association rule model. To check if $\{[balance.high][income.high] \rightarrow [loan.yes]\}$ is redundant respect to S_c the rule-based system is constructed.

The condition for rule 1 is satisfied because $[income.high] \subseteq [balance.high][income.high]$ and $[loan.yes] \subseteq [loan.yes]$ so the rule: if $[income.high] \rightarrow [loan.yes]$ then $[balance.high][income.high] \rightarrow [loan.yes]$ is redundant is added to the rule-based system. The condition for rule 2 is also fulfilled because $[balance.high] \subseteq [balance.high][income.high], [income.high] \subseteq [income.high]$ and $[loan.yes] \subseteq [loan.yes]$ so the rule if $[balance.high] \rightarrow [income.high] \wedge [income.high] \rightarrow [loan.yes]$ then $[balance.high][income.high] \rightarrow [loan.yes]$ is redundant. The condition for rule 3 is also accomplished because $[income.high] \subseteq [balance.high][income.high]$ and $[loan.yes] \subseteq [loan.yes]$, so the rule if $[income.high] \rightarrow [loan.yes]$ then $[balance.high][income.high] \rightarrow [loan.yes]$ is redundant and added to de rule-based system. The condition for rule 4 is satisfied because $[balance.high] \subseteq [balance.high][income.high]$ and $[income.high] \subseteq [balance.high][income.high]$ so the rule if $[balance.high] \rightarrow [income.high]$ then $[balance.high][income.high] \rightarrow [loan.yes]$ is redundant and added to the rule-based system. Rule 1 and 3 in the rule-based system are the same so only one of them appears in the system.

1. If $[income.high] \rightarrow [loan.yes])$ then

([$balance.high$][$income.high$] $\rightarrow$ [$loan.yes$] is redundant

2. If [$balance.high$] $\rightarrow$ [$income.high$] $\wedge$ [$income.high$] $\rightarrow$ [$loan.yes$] then [$balance.high$][$income.high$] $\rightarrow$ [$loan.yes$] is redundant
3. If [$balance.high$] $\rightarrow$ [$income.high$]then [$balance.high$][$income.high$] $\rightarrow$ [$loan.yes$] is redundant

The rule 2 in the rule-based system is formed by the conjunction of rule 1 and 3, so this rule is not considered to calculate certainty propagation because it can duplicated the evidence effect over de hypothesis.

Total certainty is guarantee for each rule in the rule-based system, but the evidence of the rules may be uncertain, so the CF will be equivalent to the CF of the evidence. The parallel combination of rule 1 and 3 is used to calculate the certainty propagation.

$CF_{(1,3)} = CF_1 + CF_3 - CF_1 * CF_3 = 0.95$

The threshold problem requires a commitment among rules' quantity in the final model and the possible loss of information. If the threshold is very low most of the rules will be pruned. It includes those with quality metric's value (confidence, certainty factor) greater than the CF of the redundancy. The compromise option is to set the cutoff threshold at the same level of the quality metric for each individual rule. Therefore, a rule will be redundant if the CF of the rule is equal or greater than the quality measureâs valued used in the extraction algorithm.

4.1. Algorithm in order to eliminate Knowledge Based Redundancy with imprecise knowledge

The logic developed to eliminate redundancy in association rules respect to imprecise knowledge contains the following states:

1. Determining the rule's redundancy using a modified version of algorithm in figure 1.
2. Establishing the rules in prior knowledge that provide evidence to redundancy. This operation is performed by a modification to the closure computation algorithm.
3. Combining the CF for each evidence

Algorithm in figure 3 performs these steps.

The function $cf\_closure()$ computes the antecedent in R_i closure and finds the rules in S_c supporting redundancy in R_i because they are used to add elements to the closure, its operation is showed in figure 4.

5. Process application

The effectiveness of the proposal was tested applying the redundancy elimination process over three datasets available at the UCI dataset repository

Require: Set of previous knowledge rules S_c
A rule R_i in form $X \rightarrow Y$
Ensure: Boolean: true if R_i is redundant false in other case

```
redundant = false
F = R_c ∪ {R_i}
for all item A ∈ X do
    if (X−{A})^+ overF = (X−{A})^+ over((F−{R_i}) ∪ (X−{A}) → Y) then
        redundant = true
    end if
end for
for all item W ∈ Y do
    if (X^+ overF = X^+ over((F−{R_i}) ∪ X → Y−{W}) then
        redundant = true
    end if
end for
if redundant = true then
    R_e = cf_closure()
    cf = combine(R_e)
    if cf < conf(R_i) then
        redundant = false
    end if
end if
return redundant
```

Figure 3: Imprecise knowledge based redundancy elimination

```
for i = 0; i < S_c.size; i++ do
    if S_c[i].antecedent ⊆ closure then
        closure = closure ∪ S_c.consequent
        cf_rules.add(S_c[i])
        i = −1
    end if
end for
return cf_rules
```

Figure 4: Function $cf\_closure()$

[18]. The first one, Adult, with data about USA census, the second one, Mushroom, with hypothetical data about mushrooms and the last one BC with data associated to breast cancer.

For all experiments, a prior knowledge S_c with four rules was used. The association rules were obtained using two different quality metrics confidence and certainty factor.

Table 2 presents the general characteristics of each experiment.

1. Column 1 holds the experiment id to match in table 3.
2. Column 2 contains the dataset name.
3. Column 3 contains the support threshold used in frequent itemset mining.
4. Column 4 holds the quality metric used.
5. Column 5 contains the total number of mined rules.

ID	Data	Support	Metric	Rules
1	Adult	0.25	conf: 0.5	3136
2	Adult	0.25	cf: 0.5	704
3	Adult	0.25	conf: 0.6	2469
4	Adult	0.25	cf: 0.6	579
5	Adult	0.25	conf: 0.7	1999
6	Adult	0.25	cf: 0.7	422
7	Adult	0.3	conf: 0.5	1804
8	Adult	0.3	cf: 0.5	369
9	Adult	0.3	conf: 0.6	1426
10	Adult	0.3	cf: 0.6	325
11	Adult	0.3	conf: 0.7	1148
12	Adult	0.3	cf: 0.7	422
13	Adult	0.35	conf: 0.5	849
14	Adult	0.35	cf: 0.5	79
15	Adult	0.35	conf: 0.6	683
16	Adult	0.35	cf: 0.6	77
17	Adult	0.35	conf: 0.7	531
18	Adult	0.35	cf: 0.7	64
19	BC	0.25	conf: 0.5	5939
20	BC	0.25	cf: 0.5	3978
21	BC	0.25	conf: 0.6	5498
22	BC	0.25	cf: 0.6	2626
23	BC	0.25	conf: 0.7	4297
24	BC	0.25	cf: 0.7	1493
25	BC	0.3	conf: 0.5	5207
26	BC	0.3	cf: 0.5	3606
27	BC	0.3	conf: 0.6	4879
28	BC	0.3	cf: 0.6	2423
29	BC	0.3	conf: 0.7	3909
30	BC	0.3	cf: 0.7	1402
31	BC	0.35	conf: 0.5	3070
32	BC	0.35	cf: 0.5	2293
33	BC	0.35	conf: 0.6	2961
34	BC	0.35	cf: 0.6	1636
35	BC	0.35	conf: 0.7	2514
36	BC	0.35	cf: 0.7	993
37	Mush	0.25	conf: 0.5	92089
38	Mush	0.25	cf: 0.5	51159
39	Mush	0.25	conf: 0.6	76706
40	Mush	0.25	cf: 0.6	42066
41	Mush	0.25	conf: 0.7	61963
42	Mush	0.25	cf: 0.7	31320
43	Mush	0.3	conf: 0.5	78998
44	Mush	0.3	cf: 0.5	48390
45	Mush	0.3	conf: 0.6	68759
46	Mush	0.3	cf: 0.6	40776
47	Mush	0.3	conf: 0.7	57634
48	Mush	0.3	cf: 0.7	30351
49	Mush	0.35	conf: 0.5	13908
50	Mush	0.35	cf: 0.5	6369
51	Mush	0.35	conf: 0.6	12031
52	Mush	0.35	cf: 0.6	5397
53	Mush	0.35	conf: 0.7	9822
54	Mush	0.35	cf: 0.7	4380

Table 2: General information

Table 3 presents the information associated to knowledge based redundancy elimination.

1. Column 1 holds the experiment id to match in table 2.
2. Column 2 contains the quantity of pruned rules.
3. Column 3 holds the quantity of redundant rules not pruned, because their CF are lower than the quality metric.
4. Column 4 contains the ratio between the pruned rules and the total number of rules.
5. Column 5 contains the ratio between the not enough certainty rules and the total number of rules.

In Figure 5, Figure 6 and Figure 7 pruned levels achieved in CF metric are presented for Adult, BC and Mush dataset respectively. In Figure 8, Figure 9 and Figure 10 pruned levels for confidence metric are presented. In both cases, the pruned ratio is calculated for support's values of 0.25, 0.3 and 0.35.

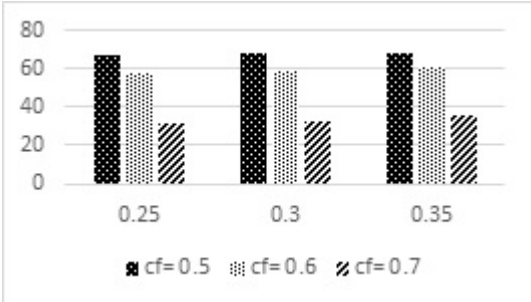

Figure 5: Pruned ratio for Adult dataset with CF

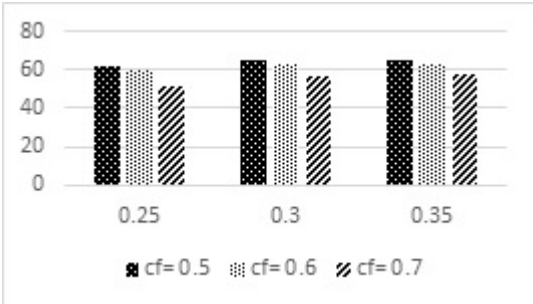

Figure 6: Pruned ratio for BC dataset with CF

Two facts are relevant from these results. The first fact is that pruning level around 50% is achieved with relatively few knowledge, only four rules, and the second ones is the highest pruning ratios are achieved with the lowest support values.

The quality metricâs influence over the pruned ratio was evaluated by the ANOVA single factor, with 5% error value. The independent variable was quality metric (confidence and certainty factor with values 0.5, 0.6 and 0.7) and the dependent variable was the pruned ratio for the different support val-

ID	Pruned	Not enougth certainty	Pruned Ratio	Not enougth certainty ratio
1	1945	472	62	15
2	471	136	66.9	19.3
3	1566	296	63.4	11.9
4	395	102	68.2	17.6
5	1305	242	65.2	12.1
6	289	82	68.4	19.4
7	1064	267	58.9	14.8
8	212	84	57.4	22.7
9	846	181	59.3	12.6
10	191	69	58.7	21.2
11	683	158	59.4	13.3
12	153	56	61.1	22
13	417	143	49.11	16.8
14	25	25	31.6	31.6
15	355	86	51.9	12.5
16	25	23	32.4	29.8
17	267	77	50.2	14.5
18	23	21	35.9	32.8
19	3331	1569	56	26.4
20	2483	682	62.4	17.4
21	3184	1313	57.9	23.9
22	1717	348	65.4	13.3
23	2627	783	61.1	18.2
24	965	174	64.6	11.7
25	2768	1406	53.1	27
26	2160	639	59.9	17.7
27	2680	1204	54.9	24.7
28	1532	336	63.2	13.9
29	2287	741	58.5	19
30	877	171	62.6	12.2
31	1427	728	46.4	23.7
32	1184	398	51.6	17.3
33	1429	655	48.3	22.1
34	925	216	56.7	13.2
35	1262	474	50.2	18.9
36	568	109	57.2	11
37	45282	30243	49.1	32.8
38	23264	19158	45.4	37.4
39	36583	26320	47.7	34.3
40	18159	16677	43.2	39.6
41	28239	22715	45.6	36.7
42	12086	13599	38.6	43.4
43	39780	25245	50.3	31.9
44	23280	16700	48.1	34.5
45	33862	22642	49.2	32.9
46	19202	14558	47.1	35.7
47	27515	19799	47.7	34.4
48	13511	11603	44.5	38.2
49	6541	4633	47.0	33.3
50	2724	2418	42.7	37.9
51	5483	4154	45.6	34.5
52	2193	2184	40.6	40.5
53	4245	2645	43.2	37.1
54	1637	1905	37.4	43.5

Table 3: Redundancy elimination

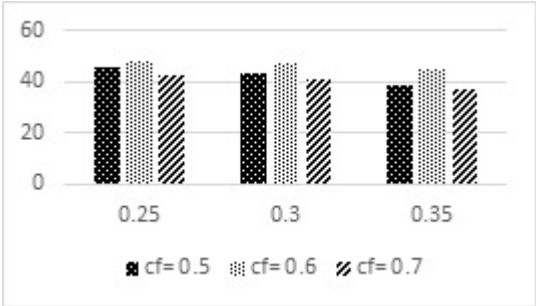

Figure 7: Pruned ratio for Mush dataset with CF

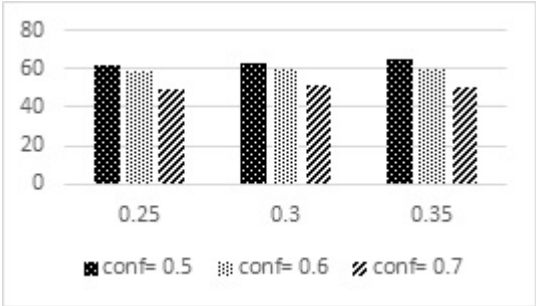

Figure 8: Pruned ratio for Adult dataset with confidence

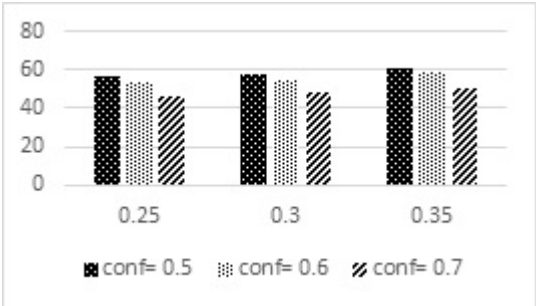

Figure 9: Pruned ratio for BC dataset with confidence

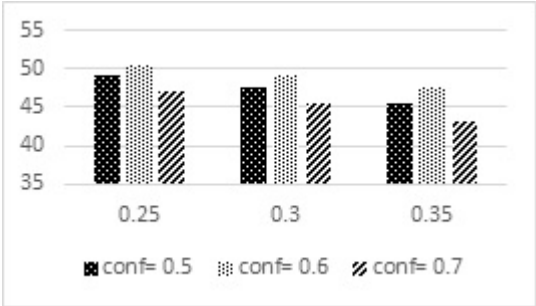

Figure 10: Pruned ratio for Mush dataset with confidence

ues (0.25, 0.3, 0.35). The other variables were controlled.

ANOVA was executed for all three datasets (Adult, Mush and BC) and three quality metric values (0.5, 0.6 and 0.7). The previous knowledge quality was invariable for all cases. The ANOVA results for *Adult*_05, *Adult*_06,

Sources	SS	df	MS	F	F crit
Between groups	33.1	1	33.1	0.17	7.7
Within groups	758	4	189.5		
Total	791.1	5			

Table 4: ANOVA for *Adult*_05 dataset

Sources	SS	df	MS	F	F crit
Between groups	39	1	39	0.2	7.7
Within groups	755.8	4	188.9		
Total	794.8	5			

Table 5: ANOVA for *Adult*_06 dataset

*Adult*_07, *BC*_05, *BC*_06, *BC*_07, *Mush*_05, *Mush*_06 and *Mush*_07 are presented in table 4, table 5, table 6, table 7, table 8, table 9, table 10, table 11 and table 12 respectively.

In all cases, the value of F is lower than the critical value of F. Therefore the hypothesis that there is not significant differences between quality metrics is accepted. The validity of this statement is partial and it might need more experimentation, but the presented results are good enough to support future works in this research line.

6. Conclusion

The notion of knowledge base redundancy in association rules has been generalized. It provides the user the possibility to express knowledge with a degree of certainty. This knowledge representation is closer to reality and it increases userâs expressive power to represent his/her knowledge about the focus domain. At this stage, non-redundant association rules models can be obtained, which makes easier user decision-making process.

7. Acknowledgements

This research is partial founding by the research project P10-TIC-6109 and TIC-1582 by Junta de Andalucia (Spain)

Sources	SS	df	MS	F	F crit
Between groups	14.7	1	14.7	0.08	7.7
Within groups	695.5	4	173.9		
Total	710.6	5			

Table 6: ANOVA for *Adult*_07 dataset

Sources	SS	df	MS	F	F crit
Between groups	56.4	1	56.4	2	7.7
Within groups	112.4	4	28.1		
Total	168.8	5			

Table 7: ANOVA for *BC*_05 dataset

Sources	SS	df	MS	F	F crit
Between groups	97.6	1	97.6	4.37	7.7
Within groups	89.1	4	22.2		
Total	186.7	5			

Table 8: ANOVA for *BC*_06 dataset

Sources	SS	df	MS	F	F crit
Between groups	35.5	1	35.5	1.5	7.7
Within groups	94.1	4	23.5		
Total	129	5			

Table 9: ANOVA for *BC*_07 dataset

Sources	SS	df	MS	F	F crit
Between groups	17.34	1	17.34	3.44	7.7
Within groups	20.16	4	5.04		
Total	37.5	5			

Table 10: ANOVA for *Mush*_05 dataset

Sources	SS	df	MS	F	F crit
Between groups	22.4	1	22.4	3.2	7.7
Within groups	27.9	4	6.98		
Total	50.3	5			

Table 11: ANOVA for *Mush*_06 dataset

Sources	SS	df	MS	F	F crit
Between groups	42.6	1	42.6	4.37	7.7
Within groups	39	4	9.75		
Total	81.6	5			

Table 12: ANOVA for *Mush*_07 dataset

References

[1] J. Diaz, C. Molina, M.A Vila, A Model for Redundancy Reduction in multidimentional Association Rules, In IADIS European Conference on Data Mining, Prague, 2013.

[2] A. S. Sadh, N. Shukla, Association Rules Optimization: A Survey, Int. J. Adv. Comput. Res. IJACR, vol. 3, 2013.

[3] M. Dhanabhakyam, M. Punithavalli, A survey on Data mining algorithm for market basket analysis, Glob. J. Comput. Sci. Technol., vol. 11, n.o 11, 2011.

[4] C. R. Valencio, F. T. Oyama, P. S. Neto, R. C. G. de Souza, Comparative Study of Algorithms for Mining Association Rules: Traditional Approach versus Multi-relational Approach, Parallel and Distributed Computing, Applications and Technologies (PDCAT), 2011 12th International Conference on, 2011, pp. 275â280.

[5] M. J. Zaki, Mining non-redundant association rules, Data Min. Knowl. Discov., vol. 9, n.o 3, pp. 223â248, 2004.

[6] G. Li, H. J. Hamilton, Basic association rules, en Fourth SIAM International Conference on Data Mining, 2004.

[7] H. Liu, L. Liu, H. Zhang, A fast pruning redundant rule method using Galois connection, Appl. Soft Comput., vol. 11, n.o 1, pp. 130â137, 2011.

[8] Y. Xu, Y. Li, G. Shaw, Reliable representations for association rulesÂ», Data Knowl. Eng., vol. 70, n.o 6, pp. 555â575, 2011.

[9] E. Baralis, L. Cagliero, T. Cerquitelli, P. Garza, Generalized association rule mining with constraints, Inf. Sci., vol. 194, pp. 68â84, 2012.

[10] R. Srikant, Q. Vu, R. Agrawal, Mining Association Rules with Item Constraints, en KDD, 1997, vol. 97, pp. 67â73.

[11] L. Geng, H. J. Hamilton, Interestingness measures for data mining: A survey, ACM Comput. Surv. CSUR, vol. 38, n.o 3, p. 9, 2006.

[12] P. Lenca, P. Meyer, B. Vaillant, S. Lallich, On selecting interestingness measures for association rules: User oriented description and multiple criteria decision aid, Eur. J. Oper. Res., vol. 184, n.o 2, pp. 610â626, 2008.

[13] K. McGarry, A survey of interestingness measures for knowledge discovery, Knowl. Eng Rev., vol. 20, n.o 1, pp. 39â61, 2005.

[14] P.-N. Tan, V. Kumar, J. Srivastava, Â«Selecting the right objective measure for association analysisÂ», Inf. Syst., vol. 29, n.o 4, pp. 293â313, 2004.

[15] Y. Bastide, N. Pasquier, R. Taouil, G. Stumme, L. Lakhal, Mining minimal non-redundant association rules using frequent closed itemsets, in Computational LogicâCL 2000, Springer, 2000, pp. 972â986.

[16] J. L. Balcazar, Redundancy, deduction schemes, and minimum-size bases for association rules, Log. Methods Comput. Sci., vol. 6, n.o 2-3, pp. 1-33, 2010.

[17] D. Heckerman, The certainty-factor model, Encycl. Artif. Intell., pp. 131â138, 1992.

[18] C. Blake, C. Merz, UCI Repository of Machine Learning Databases. University of California, Department of Information and Computer Science, Irvine, CA, 1998.