Combined Pattern Mining: From Learned Rules to Actionable Knowledge*

Yanchang Zhao[1], Huaifeng Zhang[1], Longbing Cao[1], Chengqi Zhang[1], and Hans Bohlscheid[2]

[1] Data Sciences & Knowledge Discovery Research Lab
Centre for Quantum Computation and Intelligent Systems
Faculty of Engineering & IT, University of Technology, Sydney, Australia
{yczhao,hfzhang,lbcao,chengqi}@it.uts.edu.au
[2] Projects Section, Business Integrity Programs Branch,
Centrelink, Australia
hans.bohlscheid@centrelink.gov.au

Abstract. Association mining often produces large collections of association rules that are difficult to understand and put into action. In this paper, we have designed a novel notion of *combined patterns* to extract useful and actionable knowledge from a large amount of learned rules. We also present definitions of combined patterns, design novel metrics to measure their interestingness and analyze the redundancy in combined patterns. Experimental results on real-life social security data demonstrate the effectiveness and potential of the proposed approach in extracting actionable knowledge from complex data.

1 Introduction

The notion of association rules [1] was proposed 15 years ago and is widely used today. However, as large numbers of association rules are often produced by association mining, it can sometimes be very difficult for users to not only understand such rules, but also find them a useful source of knowledge to apply to their business processes. Therefore, to present associations in an interesting and effective way, and in order to find actionable knowledge from resultant association rules, a novel idea of *combined patterns* is proposed. Combined patterns comprise *combined association rules*, *combined rule pairs* and *combined rule clusters*. A combined association rule is composed of multiple heterogeneous itemsets from different datasets, while combined rule pairs and combined rule clusters are built from combined association rules. The proposed combined patterns provide more interesting knowledge and more actionable results than traditional association rules. The contributions of this paper are: 1) a definition of combined

* This work was supported by the Australian Research Council (ARC) Linkage Project LP0775041 and Discovery Projects DP0667060 & DP0773412, and Early Career Researcher Grant RM2007002447 from University of Technology, Sydney, Australia.

W. Wobcke and M. Zhang (Eds.): AI 2008, LNAI 5360, pp. 393–403, 2008.

patterns, including combined rules, combined rule pairs and combined rule clusters; 2) interestingness measures designed for combined patterns; 3) two kinds of redundancy (i.e., rule redundancy and rule pair redundancy) identified for combined patterns; and 4) an experimental evaluation of the proposed technique on real-life data.

2 Related Work

There are often too many association rules discovered from a dataset and it is necessary to conduct post-processing before a user is able to study the rules and identify interesting ones from them. There are many techniques proposed to summarize and/or post-analyze the learned association rules [3,6]. Hilderman et al. proposed to characterize itemsets with information from external databases, e.g., customer or lifestyle data [2]. Their technique works by firstly mining frequent itemsets from transactional data and then partitioning each frequent itemset according to the corresponding characteristic tuple. This method likely results in a large number of rules when many characteristics are involved, with every characteristic having multiple value. Liu and Hsu proposed to rank learned rules by matching against expected patterns provided by user [4]. *Rule_Similarity* and *Rule_Difference* are defined to compare the difference between two rules based on their conditions and consequents, and *Set_Similarity* and *Set_Difference* are defined to measure the similarity between two sets of rules. The learned rules are ranked by the above similarity/difference and then it is up to the user to identify interesting patterns. In another work, Liu et al. proposed to mine for *class association rules* and build a classifier based on the rules [5]. With their rule generator, the rule with the highest confidence is chosen from all the rules having the same conditions but different consequents. Liu et al. also proposed *direction setting rules* to prune and summarize association rules [6]. Chi-square (χ^2) test is used to measure the significance of rules and insignificant ones are pruned. The test is then used again to remove the rules with "expected directions", that is, the rules which are combinations of direction setting rules. Zaïane and Antonie studied strategies for pruning classification rules to build associative classifiers [7]. Their idea selects rules with high accuracy based on the plot of correct/incorrect classification for each rule on the training set. Lent et al. proposed to reduce the number of learned association rules by clustering [3]. Using two-dimensional clustering, rules are clustered by merging numeric items to generate more general rules.

3 The Problem

The example that follows illustrates the target problem. Suppose that there are two datasets, transactional dataset and customer demographic dataset (see Tables 1 and 2), where "Churn" is the behaviour of a customer's switching from a company to another. In the following analysis, campaigns "d" and "e" are ignored to make the result easy to read. The traditional association rules

Table 1. Transactional Data

Customer ID	Campaign/Policy	Churn
1	a,b	Y
1	a	Y
2	a,c	N
2	b,c	Y
2	b,c,d	N
3	a,c,d	Y
3	a,b,e	Y
4	a,b	N
4	c	N
4	b,d	N

Table 3. Traditional Association Rules

Rules	Supp	Conf	Lift
$F \rightarrow Y$	3/10	3/5	1.2
$F \rightarrow N$	2/10	2/5	0.8
$M \rightarrow Y$	2/10	2/5	0.8
$M \rightarrow N$	3/10	3/5	1.2
$a \rightarrow Y$	4/10	4/6	1.3
$a \rightarrow N$	2/10	2/6	0.7
$b \rightarrow Y$	3/10	3/6	1
$b \rightarrow N$	3/10	3/6	1
$c \rightarrow Y$	2/10	2/5	0.8
$c \rightarrow N$	3/10	3/5	1.2

Table 2. Customer Demographic Data

Customer ID	Gender	...
1	F	
2	F	
3	M	
4	M	

Table 4. Combined Association Rules

Rules	Supp	Conf	Lift	$Lift_1$	$Lift_2$	I_{rule}
$F \wedge a \rightarrow Y$	2/10	2/3	1.3	1	1.1	0.8
$F \wedge b \rightarrow Y$	2/10	2/3	1.3	1.3	1.1	1.1
$F \wedge c \rightarrow N$	2/10	2/3	1.3	1.1	1.7	1.4
$M \wedge a \rightarrow Y$	2/10	2/3	1.3	1	1.7	1.3
$M \wedge b \rightarrow N$	2/10	2/3	1.3	1.3	1.1	1.1

Table 5. Combined Rule Pairs

Pairs	Combined Rules	I_{pair}
$\mathcal{P}_1$	$M \wedge a \rightarrow Y$ $M \wedge b \rightarrow N$	1.4
$\mathcal{P}_2$	$F \wedge b \rightarrow Y$ $M \wedge b \rightarrow N$	1.2

discovered are shown in Table 3, and the four rules with lift greater than one are $F \rightarrow Y$, $M \rightarrow N$, $a \rightarrow Y$ and $c \rightarrow N$. If partitioning the whole population into two groups, male and female, based on the demographic data in Table 2, and then mining the two groups separately, some rules are shown in Table 4, where $Lift_1$ and $Lift_2$ denote respectively the lift of the first/second part of the left side, and I_{rule} is the interestingness of the combined rule. The definitions of the three measures will be given in Section 4.2. We can see from Table 4 that more rules with high confidence and lift can be found by combining the rules from two separate datasets.

Although all the rules in Table 4 are of the same confidence and lift, their interestingness are not the same, which is shown by the last column I_{rule}. For example, for the first rule in Table 4, $F \wedge a \rightarrow Y$, its interestingness I_{rule} is 0.8, which indicates that the rule is not interesting at all. The explanation is that its lift is the same as the lift of $a \rightarrow Y$ (see Table 3), which means that

F contributes nothing in the rule. Therefore, our new measures are more useful than the traditional confidence and lift.

It is more interesting to organize the rules into contrasting pairs shown in Table 5, where I_{pair} is the interestingness of the rule pair. $\mathcal{P}_1$ is a rule pair for male group, and it shows that a is associated with churn but b with stay. $\mathcal{P}_1$ is actionable in that it suggests b is a preferred action/policy to keep male customers from churning. Moreover, male customers should be excluded when initiating campaign a. $\mathcal{P}_2$ is a rule pair with the same campaign but different demographics. With the same action b, male customers tend to stay, but female tend to churn. It suggests that b is a preferable action for male customers but an undesirable action for female customers.

From the previous example, we can see that rule pairs like $\mathcal{P}_1$ and $\mathcal{P}_2$ provide more information and are more useful and actionable than traditional simple rules shown in Table 3 and in this paper, they are referred to as *combined patterns*. A straightforward way to find the rules in Table 4 is to join Tables 1 and 2 in a pre-processing stage and then apply traditional association rule mining to the derived table. Unfortunately, it is often infeasible to do so in many applications where a dataset contains hundreds of thousands of records or more. Moreover, the rule clusters which organize related rules together are more useful and actionable than individual rules. To find the above useful knowledge like $\mathcal{P}_1$ and $\mathcal{P}_2$, a novel idea of combined patterns will be proposed in the next section.

4 Combined Pattern Mining

In this section we provide definitions of combined association rules and combined rule pairs/clusters, and then presents their interestingness and redundancy.

4.1 Definitions of Combined Patterns

Combined patterns take forms of *combined association rules*, *combined rule pairs* and *combined rule clusters*, which are defined as follows.

Definition 1 (Combined Association Rule). *Assume that there are k datasets $\mathcal{D}_i$ $(i = 1..k)$. Assume I_i to be the set of all items in datasets $\mathcal{D}_i$ and $\forall i \neq j,\ I_i \cap I_j = \emptyset$. A combined association rule R is in the form of*

$$A_1 \wedge A_2 \wedge ... \wedge A_k \rightarrow T, \tag{1}$$

where $A_i \subseteq I_i$ $(i = 1...k)$ is an itemset in dataset $\mathcal{D}_i$, $T \neq \emptyset$ is a target item or class and $\exists i, j, i \neq j, A_i \neq \emptyset, A_j \neq \emptyset$.

For example, A_1 can be a demographic itemset, A_2 can be a transactional itemset on marketing campaign, A_3 can be an itemset from a third-party dataset, and T can be the loyalty level of a customer. The combined association rules are then further organized into rule pairs by putting similar but contrasting rules together as follows.

Definition 2 (Combined Rule Pair). *Assume that R_1 and R_2 are two combined rules and that their left sides can be split into two parts, U and V, where U and V are respectively itemsets from $\mathcal{I}_U$ and $\mathcal{I}_V$ ($\mathcal{I} = \{I_i\}$, $\mathcal{I}_U \subset \mathcal{I}, \mathcal{I}_V \subset \mathcal{I}, \mathcal{I}_U \neq \emptyset, \mathcal{I}_V \neq \emptyset$ and $\mathcal{I}_U \cap \mathcal{I}_V = \emptyset$). If R_1 and R_2 share a same U but have different V and different right sides, then they build a combined rule pair $\mathcal{P}$ as*

$$\mathcal{P} : \begin{cases} R_1 : U \wedge V_1 \rightarrow T_1 \\ R_2 : U \wedge V_2 \rightarrow T_2 \end{cases}, \tag{2}$$

where $U \neq \emptyset$, $V_1 \neq \emptyset, V_2 \neq \emptyset, T_1 \neq \emptyset, T_2 \neq \emptyset, U \cap V_1 = \emptyset$, $U \cap V_2 = \emptyset$, $V_1 \cap V_2 = \emptyset$ and $T_1 \cap T_2 = \emptyset$.

A combined rule pair is composed of two contrasting rules, which suggests that for customers with same characteristics U, different policies/campaigns, V_1 and V_2, can result in different outcomes, T_1 and T_2. Based on a combined rule pair, related combined rules can be organized into a cluster to supplement more information to the rule pair.

Definition 3 (Combined Rule Cluster). *A combined rule cluster $\mathcal{C}$ is a set of combined association rules based on a combined rule pair $\mathcal{P}$, where the rules in $\mathcal{C}$ share a same U but have different V in the left side.*

$$\mathcal{C} : \begin{cases} U \wedge V_1 \rightarrow T_1 \\ U \wedge V_2 \rightarrow T_2 \\ \cdots \\ U \wedge V_n \rightarrow T_n \end{cases}, \tag{3}$$

where $U \neq \emptyset$; $\forall i, V_i \neq \emptyset, T_i \neq \emptyset, U \cap V_i = \emptyset$; and $\forall i \neq j, V_i \cap V_j = \emptyset$.

The rules in cluster $\mathcal{C}$ have the same U but different V, which makes them associated with various results T. Note that two rules in a cluster may have a same T. For example, assume that there is a rule pair $\mathcal{P}$ and a rule cluster $\mathcal{C}$ is built based on $\mathcal{P}$ by simply adding a third rule as follows.

$$\mathcal{P} : \begin{cases} U \wedge V_1 \rightarrow stay \\ U \wedge V_2 \rightarrow churn \end{cases}, \qquad \mathcal{C} : \begin{cases} U \wedge V_1 \rightarrow stay \\ U \wedge V_2 \rightarrow churn \\ U \wedge V_3 \rightarrow stay \end{cases}. \tag{4}$$

From $\mathcal{P}$, we can see that V_1 is a preferable policy for customers with characteristics U. However, if for some reason, policy V_1 is inapplicable to the specific customer group, $\mathcal{P}$ is no longer actionable in that it provides little knowledge on how to prevent the customers from switching to another company. Fortunately, rule cluster $\mathcal{C}$ suggests that another policy V_3 can be employed to retain those customers.

4.2 Interestingness Measures for Combined Patterns

Interestingness of Combined Association Rules. Traditional interestingness measures contribute little to selecting actionable combined patterns, because they are limited to the traditional simple association rules. Based on traditional supports, confidences and lifts, two new lifts are designed as follows for measuring the interestingness of combined association rules.

$$Lift_U(U \wedge V \to T) = \frac{Conf(U \wedge V \to T)}{Conf(V \to T)} = \frac{Lift(U \wedge V \to T)}{Lift(V \to T)} \tag{5}$$

$$Lift_V(U \wedge V \to T) = \frac{Conf(U \wedge V \to T)}{Conf(U \to T)} = \frac{Lift(U \wedge V \to T)}{Lift(U \to T)} \tag{6}$$

$Lift_U(U \wedge V \to T)$ is the lift of U with V as a precondition, which shows how much U contributes to the rule. Similarly, $Lift_V(U \wedge V \to T)$ gives the contribution of V in the rule. Based on the above two new lifts, the interestingness of combined association rules is defined as

$$I_{\text{rule}}(U \wedge V \to T) = \frac{Lift_U(U \wedge V \to T)}{Lift(U \to T)}. \tag{7}$$

It's easy to get

$$I_{\text{rule}}(U \wedge V \to T) = \frac{Lift(U \wedge V \to T)}{Lift(U \to T)\, Lift(V \to T)} \tag{8}$$

$$= \frac{Lift_V(U \wedge V \to T)}{Lift(V \to T)}. \tag{9}$$

I_{rule} indicates whether the contribution of U (or V) to the occurrence of T increases with V (or U) as a precondition. Therefore, "$I_{\text{rule}} < 1$" suggests that $U \wedge V \to T$ is less interesting than $U \to T$ and $V \to T$. The value of I_{rule} falls in $[0,+\infty)$. When $I_{\text{rule}} > 1$, the higher I_{rule} is, the more interesting the rule is.

I_{rule} works similarly as direction setting (DS) rules proposed by Liu et al. [6]. The difference is that their method gives an qualitative judgement on a rule whether it is a DS rule or not, while I_{rule} is a quantitative measure of the interestingness of a rule. I_{rule} measures how much is the unexpectedness of a combined rule against traditional simple association rules.

Interestingness of Combined Rule Pairs and Clusters. Suppose that $\mathcal{P}$ is a combined rule pair composed of R_1 and R_2 (See Formula 2), the interestingness of the rule pair $\mathcal{P}$ is defined as

$$I_{\text{pair}}(\mathcal{P}) = Lift_V(R_1)\, Lift_V(R_2)\, dist(T_1, T_2), \tag{10}$$

where $dist(\cdot)$ denotes the dissimilarity between two descendants. It is sometimes written as $I_{\text{pair}}(R_1, R_2)$ in this paper. For class with nominal values, such as "Pass" and "Fail", the dissimilarity can be defined as zero for two same descendants and as 1 for two different descendants. For ordinal class levels, such as

"Outstanding, Excellent, Good, Satisfactory, Fail", the similarity between "Outstanding" and "Fail" can be set to 1 and that between "Excellent" and "Good" can be set to 0.25. I_{pair} measures the contribution of the two different parts in antecedents to the occurrence of different classes in a group of customers with the same demographics or the same transaction patterns. Such knowledge can help to design business campaigns and improve business process. The value of I_{pair} falls in $[0,+\infty)$. The larger I_{pair} is, the more interesting a rule pair is.

For a rule cluster $\mathcal{C}$ composed of n combined association rules $R_1, R_2, \ldots, R_n$, its interestingness is defined as

$$I_{\text{cluster}}(\mathcal{C}) = \max_{i \neq j, R_i, R_j \in \mathcal{C}, T_i \neq T_j} I_{\text{pair}}(R_i, R_j). \tag{11}$$

The definition of I_{cluster} that we have provided indicates that interesting clusters are the rule clusters with interesting rule pairs, and the other rules in the cluster provide additional information. Same as I_{pair}, the value of I_{cluster} also falls in $[0,+\infty)$.

The interestingness of combined rule pair and cluster is decided by both the interestingness of rules and the most contrasting rules within the pair/cluster. A cluster made of contrasting confident rules is interesting, because it explains why different results occur and what to do to produce an expected result or avoid an undesirable consequence.

Selecting Combined Patterns. With the above interestingness measures, actionable combined patterns will be selected. First, the interesting combined rules are selected from the learned rules with support, confidence, lift, $Lift_U$, $Lift_V$ and I_{rule}. Second, the rules with high support and confidence are organized into pairs and then the pairs are ranked with I_{pair} to find contrasting rule pairs. Finally, related rules are added to selected rule pairs to build rule clusters.

Combined patterns are "actionable" in that: 1) for a single rule, $Lift_v$ measures the contribution of V to the result, which may suggest that V can be used to produce an expected outcome; and 2) the difference in the left hand of contrasting rules within a cluster explains why different results occur and how to get an expected result or avoid an undesirable consequence.

4.3 Redundancy in Combined Patterns

There are two kinds of redundancy in combined patterns: 1) the redundancy of combined rules within a rule cluster, and 2) the redundancy of combined rule pairs, which are defined as follows.

Definition 4 (Redundant Combined Association Rule). *Let $\mathcal{C}$ be a combined association rule cluster, and $R : U \wedge V \rightarrow T$ and $R' : U \wedge V' \rightarrow T'$ be two combined rules in $\mathcal{C}$, $R \in \mathcal{C}$, $R' \in \mathcal{C}$. R is redundant if $V' \subseteq V$, $T' = T$, $Lift(R') \geq Lift(R)$, $Lift_U(R') \geq Lift_U(R)$, $Lift_V(R') \geq Lift_V(R)$ and $I_{\text{rule}}(R') \geq I_{\text{rule}}(R)$.*

Definition 5 (Redundant Combined Rule Pair). *A combined rule pair $\mathcal{P}$ is redundant if: 1) there exists a rule pair $\mathcal{P}'$ with $I_{\text{pair}}(\mathcal{P}') \geq I_{\text{pair}}(\mathcal{P})$; and 2)* *for each* $R : U \wedge V \rightarrow T \in \mathcal{P}$*, there exists a rule* $R' : U' \wedge V' \rightarrow T' \in \mathcal{P}'$ *with* $U' \subseteq U$*,* $V' \subseteq V$*,* $T' = T$*,* $Lift(R') \geq Lift(R)$*,* $Lift_U(R') \geq Lift_U(R)$*,* $Lift_V(R') \geq Lift_V(R)$ *and* $I_{\mathrm{rule}}(R') \geq I_{\mathrm{rule}}(R)$*.*

Our method for removing the two kinds of redundancy of combined patterns is composed of the following two steps.

1. Removing redundant rules in each rule cluster. This step is similar to the traditional way of removing redundant association rules, but only the redundancy within each rule cluster is removed here. Within each rule cluster $\mathcal{C}$ with the same U, each rule $R \in \mathcal{C}$ is checked to see whether there exist a rule R' in the same cluster with the same T and greater confidence, Lift, $Lift_U$, $Lift_V$ and I_{rule} and $V' \subseteq V$. If yes, then R is removed from $\mathcal{C}$ as a redundant rule.
2. Pruning redundant rule pairs. This step reduces the number of rule pairs. For two rule pairs $\mathcal{P}$ and $\mathcal{P}'$, if, for each rule $R \in \mathcal{P}$, there exists a rule $R' \in \mathcal{P}'$ with the same T and greater confidence, Lift, $Lift_U$, $Lift_V$ and I_{rule}, where U' and V' in R' is are respectively subsets of U and V in R, then all the rules in $\mathcal{P}$ are redundant with respect to $\mathcal{P}'$, and $\mathcal{P}$ is a redundant rule pair in terms of $\mathcal{P}'$. So $\mathcal{P}$ will be removed to reduce the number of rule pairs.

5 A Case Study

The technique we propose was tested with real-life data in Centrelink, a Commonwealth Government agency delivering a range of services to the Australian community. The data used was customer debts raised in calendar year 2006 and corresponding customer circumstances data and transactional arrangement / repayment data in the same year. The cleaned sample data contained 355,800 customers and their demographic attributes, as well as individual debt repayment arrangements. The aim was to find the association between demographics, arrangement/repayment methods and the class of customers, which could be used to recover debts as early as possible.

We discovered combined patterns in four steps. Firstly, the transactional data (with arrangements and repayments) was mined for frequent patterns. Secondly, the whole population was partitioned into groups by frequent transactional patterns. Thirdly, the demographic data of each customer group was mined for association rules. And lastly, combined patterns were generated by combining the above results. The minimum support was set to 20 (in the count of customers instead of percentage) and the minimum confidence was set to 60%. To discover interesting combined rules, we set $Lift > 1$, $Lift_U > 1$, $Lift_V > 1$, $I_{\mathrm{pair}} > 1$ and $I_{\mathrm{rule}} > 1$, and to discover interesting combined rule clusters, the selected rules were organized into clusters, with the rule clusters then ranked by I_{cluster}.

Generally speaking, to prune redundancy in association rules, when two rules have the same confidence and one rule is more general than the other, preference was given to the shorter one. Nevertheless, when analyzing the rules discovered in this exercise, we found that because some rules were on almost the same

Table 6. Traditional Association Rules

V		*T*	*Conf* (%)	*Count*	*Lift*
Arrangement	Repayment	Class			
irregular	cash or post office	A	82.4	4088	1.8
withholding	cash or post office	A	87.6	13354	1.9
withholding & irregular	cash or post office	A	72.4	894	1.6
withholding & irregular	cash or post office & withholding	B	60.4	1422	1.7

Table 7. Selected Combined Rules

Rules	*U*	*V*		*T*	*Cnt*	*Conf* (%)	I_r	*Lift*	L_U	L_V	*Lift* of $U \to T$	*Lift* of $V \to T$
	Demographics	Arrangement	Repayment	Class								
r_1	age:65+	withholding & irregular	withholding	C	50	63.3	2.91	3.40	2.47	4.01	0.85	1.38
r_2	income:0 & remote:Y & marrital:sep & gender:F	withholding	cash or post & withholding	B	20	69.0	1.47	1.95	1.34	2.15	0.91	1.46
r_3	income:0 & age:65+	withholding	cash or post & withholding	A	1123	62.3	1.38	1.35	1.72	1.09	1.24	0.79
r_4	income:0 & gender:F & benefit:P	withholding	cash or post	A	469	93.8	1.36	2.04	1.07	2.59	0.79	1.90

group of customers, business experts tended to prefer longer rules which provided more detailed information concerning the overall characteristics of the group. Therefore, in this case study, those rules with confidence less than 1.05 times that of more specific rules were removed as redundant rules, and the same was done to remove redundant rule clusters.

There were 7,711 association rules before removing redundancy of combined rules. After removing redundancy of combined rules, 2,601 rules were left, which built up 734 combined rule clusters. After removing redundancy of combined rule clusters, 98 rule clusters with 235 rules remained, which was within the capability of human beings to read. The traditional association rules we discovered from transactional data are given in Table 6. Some selected combined patterns are shown respectively in Tables 7 and 8. In the two tables, columns L_U and L_V stand for $Lift_U$ and $Lift_V$, respectively.

In Table 7, r_1: "*Age:65+, arrangement=withholding and irregular, repayment=withholding* $\to$ *C*" has a high I_{rule} of 2.91. "*Lift of U* $\to$ *C*" indicate that the lifts of "*Age:65+* $\to$ *C*" is 0.85, which suggests that "*Age:65+*" is negatively associated with "C". However, $Lift_U = 2.47$ suggests that, under "*arrangement=withholding and irregular, repayment=withholding*", "*Age:65+*" becomes positively associated with "C". Moreover, $Lift_V$ is greater than "*Lift of V* $\to$ *C*", which suggests that the contribution of the specific arrangement and repayment to the occurrence of "C" also increases in customer group "*Age:65+*". What's more, $Lift = 3.40$ also suggests that the combination of "*Age:65+*" and "*arrangement=withholding and irregular, repayment=withholding*" more than triples the probability of the occurrence of "C". Therefore, r_1 is a very interesting rule, which explains why it has a high value of I_{rule}. In contrast, r_5 in Table 8 has an I_{rule} of 0.86 (shown as I_r), which indicates that it is not interesting as a single

Table 8. Selected Combined Rule Clusters

Clusters	Rules	U	V		T	Cnt	$Conf$	I_r	I_c	$Lift$	L_U	L_V	$Lift$ of	$Lift$ of
		demographic	arrangement	repayment			(%)						$U \to T$	$V \to T$
R_1	r_5	age:65+	withhold	cash or post	A	1980	93.3	0.86	6.5	2.02	1.06	1.63	1.24	1.90
	r_6		irregular	cash or post	A	462	88.7	0.87		1.92	1.08	1.55	1.24	1.79
	r_7		withhold & irregular	cash or post	A	132	85.7	0.96		1.86	1.18	1.50	1.24	1.57
	r_8		withhold & irregular	withhold	C	50	63.3	2.91		3.40	2.47	4.01	0.85	1.38
R_2	r_9	marital:sin &gender:F &benefit:N	irregular	cash or post	A	400	83.0	1.12	6.3	1.80	1.01	2.00	0.90	1.79
	r_{10}		withhold	cash or post	A	520	78.4	1.00		1.70	0.89	1.89	0.90	1.90
	r_{11}		withhold & irregular	cash or post & withhold	B	119	80.4	1.21		2.28	1.33	2.06	1.10	1.71
	r_{12}		withhold	cash or post & withhold	B	643	61.2	1.07		1.73	1.19	1.57	1.10	1.46
	r_{13}		withhold & vol. deduct	withhold & direct debit	B	237	60.6	0.97		1.72	1.07	1.55	1.10	1.60
	r_{14}		cash	agent	C	33	60.0	1.12		3.23	1.18	3.07	1.05	2.74
R_3	r_{15}	income:0 &age:22-25	irregular	cash or post	A	191	76.7	1.03	5.1	1.66	0.93	1.85	0.90	1.79
	r_{16}		cash	cash or post	C	440	62.1	1.08		3.34	1.31	2.76	1.21	2.56
R_4	r_{17}	benefit:Y &age:22-25	irregular	cash or post	A	218	79.6	1.15	4.1	1.73	0.97	2.06	0.84	1.79
	r_{18}		cash	cash or post	C	483	65.6	0.78		3.53	1.38	1.99	1.78	2.56

rule. Although r_5 has a high lift of 2.02, its $Lift_U$ and $Lift_V$ are respectively less than "*Lift of* $U \to C$" and "*Lift of* $V \to C$", which suggests that the contribution of U and V to the occurrence of C becomes less when they are combined together. That is, for r_5, $U \wedge V \to C$ is actually less interesting or useful than $U \to C$ and $V \to C$. Nevertheless, it does not necessarily mean that r_5 is not interesting as a part of a rule cluster, since I_{rule} measures the interestingness of a single rule, not that of a rule cluster.

Some selected rule clusters are shown in Table 8. The clusters are ordered descendingly by I_{cluster} (shown as I_c). Within each cluster, the rules are ordered first ascendingly by class and then descendingly by $Lift_V$ (shown as L_V). For customers with "*maritle:single, gender:F, benefit:N*" (see R_2), "*Arrangement=irregular or withholding, Repayment=cash or post office*" is associated with *class A* (see r_9 and r_{10}), while "*Arrangement=cash, Repayment=agent recovery*" is associated with *class C* (see r_{14}). Here, *Class A* is preferable than *Class B*, and *Class B* is preferable than *Class C*. Therefore, for a single female customer with a new debt, if her benefit type is N, she may be encouraged to repay under "*Arrangement=irregular or withholding, Repayment=cash or post office*", and be persuaded not to repay under "*Arrangement=cash, Repayment=agent recovery*". In such a way, her debt will probably be repaid more quickly. For the above customer group of single female on benefit N, the priority of arrangement-repayment methods is given by the rules from r_9 to r_{14}. Such kind of knowledge is actionable in that it can help to improve policy or design campaigns to recover debts as soon as possible.

6 Conclusions

This paper presents a new idea of combined patterns. The concepts of combined association rules, combined rule pairs and combined rule clusters are defined; the interestingness of each is designed; and two kinds of redundancy are analyzed. The proposed combined patterns are more useful and actionable than traditional simple association rules. And our technique, which has been tested with real-world data, has provided some interesting and helpful results.

Acknowledgments

Our thanks go to Mr. Fernando Figueiredo, Mr. Peter Newbigin and Mr. Brett Clark from Business Integrity Programs Branch, Centrelink, Australia, for their domain knowledge and helpful suggestions.

References

1. Agrawal, R., Imielinski, T., Swami, A.: Mining association rules between sets of items in large databases. In: Proc. of the ACM SIGMOD International Conference on Management of Data, Washington, D.C., USA, May 1993, pp. 207–216 (1993)
2. Hilderman, R.J., Carter, C.L., Hamilton, H.J., Cercone, N.: Mining Market Basket Data Using Share Measures and Characterized Itemsets. In: Wu, X., Kotagiri, R., Korb, K.B. (eds.) PAKDD 1998. LNCS, vol. 1394, pp. 159–170. Springer, Heidelberg (1998)
3. Lent, B., Swami, A.N., Widom, J.: Clustering association rules. In: Proceedings of the 13th International Conference on Data Engineering, Birmingham, U.K, April 7–11, pp. 220–231. IEEE Computer Society, Los Alamitos (1997)
4. Liu, B., Hsu, W.: Post-analysis of learned rules. In: Proceedings of the 13th National Conference on Artificial Intelligence (AAAI 1996), Portland, Oregon, USA, pp. 828–834 (August 1996)
5. Liu, B., Hsu, W., Ma, Y.: Integrating classification and association rule mining. In: Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining (KDD 1998), pp. 80–86. AAAI Press, Menlo Park (1998)
6. Liu, B., Hsu, W., Ma, Y.: Pruning and summarizing the discovered associations. In: Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 1999), pp. 125–134. ACM Press, New York (1999)
7. Zaïane, O.R., Antonie, M.-L.: On pruning and tuning rules for associative classifiers. In: Khosla, R., Howlett, R.J., Jain, L.C. (eds.) KES 2005. LNCS (LNAI), vol. 3683, pp. 966–973. Springer, Heidelberg (2005)