Semantic Analysis of Association Rules via Item Response Theory

Shinichi Hamano and Masako Sato

Department of Mathematics and Information Sciences
College of Integrated Arts and Sciences
Osaka Prefecture University, Sakai, Osaka 599-8531, Japan
e-mail: shinchan0112@aol.com, sato@mi.cias.osakafu-u.ac.jp

Abstract. This paper aims to install Latent trait on Association Rule Mining for the semantic analysis of consumer behavior patterns. We adapt Item Response Theory, a famous educational testing model, in order to derive interesting insights from rules by Latent trait. The primary contributions of this paper are fourfold. (1) Latent trait as an unified measure can measure interestingness of derived rules and specify the features of derived rules. Although the interestingness of rules is swayed by which measure could be applied, Latent trait that combines descriptive and predictive property can represent the unified interestingness of the rules. (2) Negative Association rules can be derived without domain knowledge. (3) Causal rules can be derived and analyzed by the Graded Response Theory which is extended model of Item Response Theory. (4) The features of consumer choice that is based on the concept of multinomial logit mode in Marketing Science could be extracted. Especially the effect of promotions and product prices based on Causal rules can be generated. Our framework has many important advances for accomplishing in mining and analyzing consumer behavior patterns with diversity.

1 Introduction

In the past decade, Association Rule Mining[3] has become the focus of attention. An association rule is an implication of the form $X \Rightarrow Y$, where X and Y are item-sets satisfying $X \cap Y = \phi$ and represents the consumer purchasing patterns in a transaction. In order to measure interestingness of Association rules, many useful measures have been proposed. However respective measures have their own interestingness so that we often encounter the interpretational problems. Because there is no measure to describe the unified interestingness of Association rules, each interestingness differs much from the others. Even support and confidence, the most fundamental measures, are not efficient enough to represent the unified interestingness. Additionally there exists no measure that is not only descriptive but also predictive. The prediction and description tasks have been treated as the distinct problems for mining Association rules. As just stated, measures with well-matched properties are demanded for these purposes. We address these issues by introducing Latent trait to Association rule analysis.

An implication of the form $X \Rightarrow \neg Y$ is called a Negative Association rule[2], and represents a rule that customers who buy an item-set X are not likely to buy at least one item in an item-set Y. Negative Association rules are significant for understanding consumer behavior patterns. Although Negative Association rule mining is quite useful, too many worthless Negative Association rules could be derived. To settle this issue, Indirect Association rule mining was proposed in [3] for deriving the interesting Negative Association rules effectively.

An Indirect Association rule is an implication of the form $X \Rightarrow Y$ with $Y = \{y_1, y_2\}$ where there is a negative correlation between y_1 and y_2. In the frameworks in [3, 4] an item-set X called a mediator represents a common item-set for the consumers, i.e., X illustrates the similarity of consumer behavior patterns. The relations of X and each of items in Y illustrate that the consumers who bought all items of a mediator X take different actions of buying an item y_1 or y_2. For example, if we discover an Indirect Association rule $(X \Rightarrow y_1, X \Rightarrow y_2)$, consumers that buy all items in X tend to buy either y_1 or y_2 but not both. It enables us to figure out that the items y_1 and y_2 are in a choice item-set, i.e., they could be competitive products or alternative products.

We have previously extended the model of an Indirect Association rule in order to represent interesting consumer behavior patterns. An Indirect Association rule due to S. Hamano and M. Sato[4] is an implication of the form $X \Rightarrow \beta_{y_j}$ with $\beta_{y_j} \in \{y_j, \neg y_j\}$ for $j = 1, 2$ and illustrates behaviors of consumer choice with two alternatives. For example, if we discover an Indirect Association rule $(X \Rightarrow y_1, X \Rightarrow \neg y_2)$, consumers that buy all items in X tend to buy together with y_1 but not with y_2. Even though an Indirect Association rule is quite valuable to recognize consumer behavior patterns, it can describe only consumer choice between two alternatives. In order to illustrate consumer choice behavior with multiple alternatives, we are going to introduce a new framework of Association rule analysis via Item Response Theory[8].

IRT is a mathematical model in educational testing for studying individual responses. The aim of IRT is to obtain fine estimates of Latent trait called ability. This is an attractive model to predict the probability of each correct response as a function of the Latent trait and some parameters. Latent trait is a powerful unobserved factor of measuring interestingness of derived rules as an unified measurement and specify the features of derived rules. Latent trait is estimated by the EM algorithm[5] with some parameters. These parameters are also significant as well as Latent trait in terms of capturing the features of rules. One of the parameter a called discrimination parameter can classify Association rules and Negative Association rules without domain knowledge and remove uninteresting rules effectively. The parameter c called guessing parameter can remove trivial rules without asking experts.

Graded Response Theory which is the extended model of IRT can be applied to Causal rule mining[7]. From this application, the effect of promotions and product prices based on Association rules can be generated like multinomial logit model[6]. As you have seen, our framework has many important advances for accomplishing in mining and analyzing consumer behavior patterns with diversity.

This paper is organized as follows. In the next section, Item Response Theory for analysis of Association rules is introduced. In Section 3, we present Graded Response Theory for analysis of Causal rules. Experimental results are presented in Section 4.

2 Semantic Analysis of Association Rules via IRT

2.1 Preliminary

Let I be a finite set of items, and D be a set of customer transactions, called a *database*, where each transaction T is a set of items such that $T \subseteq I$. Each transaction has a unique identifier i, called Transaction ID (TID), is denoted by T_i. We denote subsets of I, called *item-sets*, by X, Y and items of X and Y by $x_1, x_2, x_3, \cdots$ and by $y_1, y_2, y_3, \cdots$ respectively. In this paper, an item-set X means not only the subset of I but also the event that a transaction contains all items in the set X, and $Pr(X)$ denotes the probability that a transaction contains the set X. Moreover, $\neg X$ denotes the negation of the event X, i.e., the event that a transaction does not contain at least one item in X, and thus $\neg\neg X = X$ and $Pr(\neg X) = 1 - Pr(X)$.

We consider an item-set Y such that $Y = \{y_1, y_2, ..., y_n\}$. Let y_{ij} represent a presence or an absence of an item y_j in a transaction with $TID = i$, that is, $y_{ij} = 1$ if $y_j \in T_i$ and $y_{ij} = 0$ if $y_j \notin T_i$. Item response data with $TID = i$ for given an item-set Y, denoted by $T_{i,Y}$, are represented as a binary vector, i.e., $T_{i,Y} = (y_{i1}, y_{i2}, ..., y_{in})$. Remark that $T_i \cap Y = \{y_j | y_{ij} = 1, j = 1, 2, ..., n\}$. For example, consider an item-set $Y = \{y_1, y_2, y_3, y_4, y_5\}$. The item response data $T_{4,Y} = (1, 0, 0, 0, 1)$ indicate that the transaction with $TID = 4$ contains items y_1 and y_5, but does not contain items y_2, y_3, and y_4.

2.2 Semantic Analysis of Association Rules via IRT

In this section, we introduce Item Response Theory for an Association rule with Multiple Alternatives. Let X and $Y = \{y_1, y_2, ..., y_n\}$ be item-sets such that X, $Y \subseteq I$ and $X \cap Y = \phi$. Let D_X be the set of transactions which contains all items in X. We assume the condition that all items in Y are statistically independent in the database D_X, called the local independence condition. We denote an Association rule with Multiple Alternatives $(X \Rightarrow \beta_{y_1}, X \Rightarrow \beta_{y_2}, \cdots, X \Rightarrow \beta_{y_n})$ with $\beta_{y_j} \in \{y_j, \neg y_j\}$ by $(X; \beta_{y_1}, \beta_{y_2}, \cdots, \beta_{y_n})$.

We presume that there is a Latent trait between item-sets X and Y. Let θ be a Latent trait that is an unobserved factor of measuring the underlying ability and θ_X represent the ability of an transaction that contains all items in X. That is, D_x has an inherent unobserved variable, Latent trait θ_X, and this θ_X dominates the probability of occurring an item y_j. Let $p_j(\theta_X)$ and $q_j(\theta_X)$ represent the probabilities of the presence and the absence of an item $y_j \in Y$ in D_X respectively as follows:

$$p_j(\theta_X) = Pr(y_j|X), \qquad q_j(\theta_X) = 1 - p_j(\theta_X) = Pr(\neg y_j|X). \tag{1}$$

The interestingness of an Association rule $X \Rightarrow y_j$ is measured by the conditional probability of y_j given X while the interestingness of a Negative Association rule $X \Rightarrow \neg y_j$ is measured by the conditional probability of $\neg y_j$ given X. The measures λ for evaluating interestingness of an Association rule and a Negative Association rule are defined as follows:

$$\lambda(X \Rightarrow y_j) = Pr(y_j|X), \qquad \lambda(X \Rightarrow \neg y_j) = Pr(\neg y_j|X). \tag{2}$$

Because there is a Latent trait between item-sets X and Y and the Latent trait θ_X of D_X dominates the probability of occurring y_j, the conditional probability can be represented as a function of θ_X. Therefore $p_j(\theta_X)$ and $q_j(\theta_X)$ represent the interestingness of an Association rule and a Negative Association rule respectively. The probability of that a transaction with $TID = i$ whether contains an item y_j can be represented briefly as follows:

$$f(y_{ij}|\theta_X) = p_j(\theta_X)^{y_{ij}} q_j(\theta_X)^{1-y_{ij}}. \tag{3}$$

For item response data $T_{i,Y} = (y_{i1}, y_{i2}, \cdots, y_{in})$, translating process into a joint probability model based on the local independence assumption results in the following:

$$f(T_{i,Y}|\theta_X) = \prod_{j=1}^{n} p_j(\theta_X)^{y_{ij}} q_j(\theta_X)^{1-y_{ij}}. \tag{4}$$

Logarithmic likelihood function of the above probability is illustrated as follows:

$$\log L(T_{i,Y}|\theta_X) = \sum_{j=1}^{n} \left[y_{ij} \log p_j(\theta_X) + (1 - y_{ij}) \log q_j(\theta_X) \right]. \tag{5}$$

The parameters θ_X is estimated by maximization of the above Logarithmic likelihood function.

IRT is a prominent mathematical model in educational testing for studying individual responses. The aim of IRT is to obtain fine estimates of Latent trait called ability. This is an attractive model to predict the probability of each correct response as a function of the Latent trait and some parameters. The function called an Item Response Function (IRF) of two parameter logistic model (2PL) is defined as follows:

$$p_j(\theta_X) = \frac{1}{1 + \exp(-1.7 a_j (\theta_X - b_j))} \qquad (2PL) \tag{6}$$

where a_j is the discrimination parameter, b_j is the difficulty parameter and θ_X is the ability level. The parameters a_j, b_j and θ_X are estimated by maximization of the above Logarithmic likelihood function. We apply the IRF defined above to Association Rule Mining for measuring interestingness of the rules.

The IRF for Association Rule Mining represents the conditional probability of y_j given X where θ_X is the Latent trait of causing the event y_j. These parameters a_j, b_j and θ_X could be generated from the database by the EM algorithm[5].

In our new perspective, Association and Negative Association rules are discriminated by the parameter a_j. If $a_j > 0$, the Association rule $X \Rightarrow y_j$ is derived because the conditional probability of y_j given X monotonically increases with θ_X. On the other hand, if $a_j < 0$, the Negative association rule $X \Rightarrow \neg y_j$ is derived because the conditional probability of y_j given X monotonically decreases with θ_X. Hence, the parameter a_j is the key factor of distinguishing Association and Negative Association rules. Moreover the higher the absolute value of the discrimination parameter is, the more interesting the rule is. The Fisher information of an item-set Y is defined as follows:

$$I(\theta_X) = E\left[\frac{\partial}{\partial\theta} logL(T_{i,Y}|\theta_X)^2\right] = 1.7^2 \sum_{j=1}^{n} a_j^2 p_j(\theta_X) q_j(\theta_X) \tag{7}$$

Moreover the Fisher information of an item y_j is defined as follows:

$$I_{y_j}(\theta_X) = 1.7^2 a_j^2 p_j(\theta_X) q_j(\theta_X). \tag{8}$$

The discrimination parameter a_j for an item y_j is noticeably significant in terms of augment the information. Therefore the parameter a_j is applied for measuring interestingness of the rules and classifying Association rules and Negative Association rules.

The parameter b_j represents how much the occurrence of X relates the occurrence of each of items in a set of alternatives, i.e., causal relationships between X and each of items could be measured.

The most important and fascinating factor is the parameter θ_X that represents the Latent trait. By estimating the parameter θ_X, the responses of each rule could be predicted. When the three parameters are reasonably accurate, the predictive property will be assured. The parameter θ_X represents the interestingness of the rules. The higher the ability level is, the more interesting the rule is. We can say that these parameters are descriptive measures for derived rules. Item Characteristic Curve (ICC) is a graph of Item Response Function and a visual tool of illustrating the choice probability of each items.

According to the Latent trait model, the local independence condition is necessary for estimation of Latent trait and some parameters. That is, the condition of that correlations of any items are all statistically independent is absolutely necessary in order to generate fine estimates. However it is expensive to generate an item-set that satisfy the local independence condition. Therefore we introduce choice item-set for relaxing the condition of local independence in the next section.

2.3 Common Item-sets and Choice Item-sets

In order to apply IRT to analysis of an Association rule with Multiple Alternatives, we introduce two item-sets called common item-set and choice item-set. The common item-sets help to reduce search space and the number of uninteresting and trivial rules. Common item-sets illustrate the consumer behavior patterns as ordinal patterns. The common item-set O is defined as follows:

$$O = \{X \subset I \mid Pr(X) \geq \eta_f\},$$

where η_f is a *common item-set threshold* predefined by the user. We should note that it is desirable that there are at least 500 transactions containing all items in a common item-set for stable parameter estimations of the Latent trait. Even though common item-sets can describe the similar consumer behaviors, it is not sufficient enough to recognize the characteristics of consumers. In order to illustrate characteristics, we are going to define choice item-set as a set of Multiple Alternatives. The items in choice item-set could give us great insights by deriving with a common item-set because we could see many kinds of different selecting actions of consumers as characteristics. Let $\Im$ be a family of sets as follows:

$$\Im = \{Y \subset I \mid \forall i_j, \forall i_k \in Y,\ |\rho(i_j, i_k)| \leq \eta_\rho, j \neq k, |Y| \geq 2\},$$

where ρ is the coefficient of correlation between the pair of items and η_ρ is a correlation threshold predefined by the user. Let us define an choice item-set S that is a maximal set in $\Im$, i.e., there is no sets $S^{'}$ in $\Im$ satisfies $S \subsetneq S^{'}$. We should also note that it is desirable that the size of a choice item-set has to be secured to a certain degree for stable parameter estimations of the Latent trait.

Definition 1. *Let X be a common item-set and Y be a choice item-set where X,$Y \subseteq I$ and $X \cap Y = \phi$ and a_j be a discrimination parameter of an item y_j. An Association rule with Multiple Alternatives can be extracted as an interesting rule for 2PL IRT model if*
(1) $|a_j| \geq \eta_d$ (Discrimination parameter Condition).

In the next section, we introduce Graded Response Theory for analysis of Causal rule which is extension of Association rule.

3 Causal Rule Analysis via Graded Response Theory

A Causal rule proposed in [7] is an implication of the form $X \Rightarrow Y$ where X and Y are sets of categorical variables with $X \cap Y = \phi$. For each categorical variable X_i, $R(X_i)$ called range consists of finite order categorical items. We assume that X is a conjunction of explanatory categorical variables X_i and y_j is a target item. By adapting Graded Response Theory (GRT) which is extended model of IRT, interesting insights from Causal rules could be derived such as influence and interestingness of categorical variables. Let X_i be categorical variable that has K ordered value as follows:

$$X_i = 0, 1, 2, \cdots, k, \cdots, K-1 \tag{9}$$

The probability of $X_i = k$ is defined as follows:

$$p(X_i = k|\theta_{y_j}) = p_{ik} = p^*_{ik}(\theta_{y_j}) - p^*_{ik+1}(\theta_{y_j}), \tag{10}$$

where $p^*_{ik}(\theta_{y_j})$ represents the probability of that $X_i \geq k$. Note that $p^*_{i0}(\theta_{y_j}) = 1$ and $p^*_{iK}(\theta_{y_j}) = 0$. The probability $p^*_{ik}(\theta_{y_j})$ for 2PL logistic model is defined as follows:

$$p^*_{ik}(\theta_{y_j}) = \frac{1}{1 + \exp(-1.7a_i(\theta_{y_j} - b^*_{ik}))}. \tag{11}$$

Let M_l be the categorical response data matrix. By the local independence condition, the probability of an observation M_l is illustrated as follows:

$$p(M_l|\theta_{y_j}) = \prod_{i=1}^{n} \prod_{k=0}^{K-1} p_{ik}(\theta_{y_j})^{X_{ik}}. \tag{12}$$

Logarithmic likelihood function of the above probability is defined as follows:

$$\log L(M_l|\theta_{y_j}) = \sum_{i=1}^{n} \sum_{k=0}^{K-1} X_{ik} \log p_{ik}(\theta_{y_j}). \tag{13}$$

The parameters a_i, b^*_{ik} are estimated by the EM algorithm as maximization of the above Logarithmic likelihood function. The Fisher information for a target item y_j is defined as follows:

$$I_{X_i}(\theta_{y_j}) = 1.7^2 a_i^2 \sum_{k=0}^{K-1} \frac{(p^*_{ik}(\theta_{y_j})q^*_{ik}(\theta_{y_j}) - p^*_{ik+1}(\theta_{y_j})q^*_{ik+1}(\theta_{y_j}))^2}{p_{ik}(\theta_{y_j})} \tag{14}$$

Let $\hat{\theta}_X$ and $\hat{\theta}_{X_i}$ be an estimated Latent trait of X and X_i respectively for a target item y_j. The estimated Latent trait $\hat{\theta}$ is regarded significant as much as a discrimination parameter. Hence an estimated Latent trait is one of criteria for measuring the interestingness of derived Causal rules.

Definition 2. *Let $X = \{X_1, X_2, \cdots, X_n\}$ be a set of explanatory categorical variables and y_j be a target item. Let $\hat{\theta}_X$ and $\hat{\theta}_{X_i}$ be an estimated Latent trait of a set X of explanatory categorical variables and each explanatory categorical variables X_i respectively for a target item y_j. Causal rule $X \Rightarrow y_j$ can be extracted as an interesting causal rule, if*
(1) $\hat{\theta}_X \geq \frac{1}{n}\sum_{i=1}^{n} \hat{\theta}_{X_i}$, (Latent Trait Condition)
(2) $|a_i| \geq \eta_d$, (Discrimination Condition).

3.1 Analyzing Effect of Promotion and Price

Multinomial logit model[6] have been contributed Marketing Science for identifying the variables that affect consumer choices from among a set of alternatives such as price, promotion, and so on. Even though data of price and promotion are significant factors for marketing, there have been no effective method to analyze the effect of promotions and product prices based on Association rules for market basket data. Our framework can analyze them by introducing variables for price and promotion data as explanatory categorical variables for a target item. Let $Promo_{y_j}$ be a promotion variable and $Price_{y_j}$ be a price variable for a target item y_j. A Latent trait and a discrimination parameter for each explanatory variables can be estimated by the EM algorithm. The higher discrimination parameter is, the more influential and significant variable is. For example, suppose there are three explanatory variables X_1, $Promo_{y_1}$ and $Price_{y_1}$ for a target item y_1, i.e., $X = \{X_1, Promo_{y_1}, Price_{y_1}\}$ and the estimated discrimination parameters for each variables are 1.0, 1.5 and 0.5 respectively. As easily seen, the Promotion variable is more significant than the Price variable. Hence the promotion is the most influential factor for consumers who bought a target item y_1. Moreover the explanatory variable X_1 is more significant than the Price variable. This could be an interesting insight because there is the factor that is more influential than the price of an item y_1. As a result, we can recommend a marketing manager that the pricing for an item y_1 should be reconsidered. Moreover the higher Latent trait is, the more significant categorical variable is. For example, three kinds of promotions have been done for a target item y_1. The promotion categorical variable is presented as 1, 2, and 3 for *"only special display promotion"*, *"only advertisement promotion"*, and *"both special display and advertisement promotion"* respectively. Note that the promotion variable 0 means *"no promotion"*. If estimated Latent traits are -1.5, 0.5, 1.0, 1.5 for each promotion categorical variable 0, 1, 2, and 3 respectively, then the ICC as in figure 1 is represented. The ICC indicates that most effective promotion was *"both special display and advertisement promotion"* and the promotion *"only special display"* was not effective.

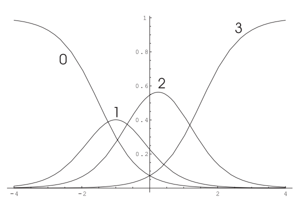


Fig. 1. Item Characteristic Curve

4 Experiments

We have performed analysis of Direct Marketing data distributed for the KDD CUP in 1998. These experiments are for deriving interesting Causal rules based on the graded response model and analyzing categorical variables. Suppose a Causal rule has three categorical variables for a set X of categorical variable, i.e., $X = \{X_1, X_2, X_3\}$. The domains for categorical variables X_1, X_2, X_3, and

```
Input :  D– Database
         y_i– Target item
         t_d– minimum discrimination parameter threshold
Output : A list of interesting Causal rules

1 ) for each explanatory categorical variable
2 )     Estimate (a_{X_i}, θ_{X_i}) by the EM algorithm
3 )         if | a_{X_i} |≥ t_d then
4 )         X = X ∪ {X_i}
5 ) for X = {X_1, X_2,..., X_n}
6 )     Estimate θ_X by the EM algorithm
7 )         if θ_X ≥ (1/n) Σ_{k=1}^{n} θ_k then
8 )         Output (X, y_i, θ_X)
```

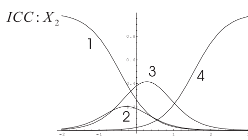
Algorithm of mining Causal rules via Graded Response Theory

	RFA_2	RFA_3
1	L4D	S4D
2	L4E	S4E
3	L3D	A4D
4	L3E	N4D
5	L4E	S4F

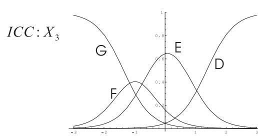
Top 5 interesting Causal rules

Fig. 2. Algorithm and top 5 interesting Causal rules

Y are $X_1 = \{F, N, A, L, S\}$, $X_2 = \{1, 2, 3, 4\}$, $X_3 = \{A, B, C, D, E, F, G\}$, and $Y = \{Donor\}$. Due to the space limitation, we present only the results of RFA_2 and RFA_3.

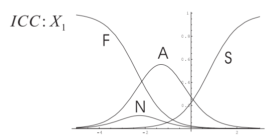


	X_2
a_{X_2}	1.3879
b_1	-0.4418
b_2	-0.08925
b_3	0.65885
b_4	1.0544

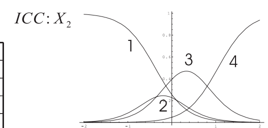


	X_3
a_{X_3}	1.329
b_G	-1.3588
b_F	-0.6011
b_E	0.75925
b_D	1.3619

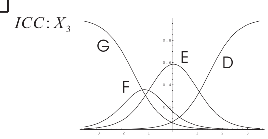
Item Characteristic Curve for RFA_2



	X_1
a_{X_1}	0.9284
b_F	-2.3822
b_N	-2.09545
b_A	-0.5079
b_S	0.7929



	X_2
a_{X_2}	1.6797
b_1	-0.3766
b_2	-0.0256
b_3	0.68875
b_4	1.0521



	X_3
a_{X_3}	1.0550
b_G	-1.5237
b_F	-0.691
b_E	0.80405
b_D	1.4664

Item Characteristic Curve for RFA_3

Fig. 3. Experimental results

Figure 3 depicts the ICC of estimated parameters of each categorical variable from the data of RFA_2 and RFA_3. The domain of the first categorical variable X_1 in RFA_2 is {L}, that is, all donors belong to the category $\{L\}$. Although the higher discrimination parameter is, the more significant the categorical variable is, there is hardly difference between the item discrimination parameters for X_2 and X_3 so that categories X_2 and X_3 are regarded as much the same categorical variables. The fact that the highest item difficulty parameter of X_2 is a category 4 indicates people who belong to category 4 are likely to be Donor. Low item difficulty parameters for categories 1 and 2 indicate that people who belong to these categories are not likely to be Donor. Although category 3 has much high difficulty parameter, it could not be an attractive category because its highest probability is not high enough. The highest item difficulty parameter of X_3 is D so that the people are potentially attractive Donor. Additionally, the second highest parameter E is attractive enough because the highest probability is high enough. The domain of the first categorical variable X_1 in RFA_3 is $\{F, N, A, S\}$, of the second variable X_2 is $\{1, 2, 3, 4\}$ and of the third variable X_3 is $\{D, E, F, G\}$. Because categories A, B and C in the third category are too low to generate secure estimates. I and L in the first category are too low as well as A, B and C. The ICC and parameter estimations indicate S, 3, 4 and G are attractive categories for Donor. The most interesting rules are $L4D$ and $S4D$ for RFA_2 and RFA_3 respectively. Therefore the people who belong to these categories have propensity for being Donor.

5 Conclusion and Future Work

We are convinced that our framework produce good results from semantic analysis of Association rules and Causal rules. As a future work, we are going to extend the current model to relax local independence condition.

References

1. R. Agrawal, T. Imielinski and A. Swami: *Mining Association Rules between Sets of Items in Large Databases*, in Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, Washington, D.C., USA, May 26–28, 1993, pp. 207–216, 1993.
2. A. Savasere, E. Omiecinski and S. Navathe: *Mining for Strong Negative Associations in a Large Database of Customer Transactions*, in Proceedings of the Fourteenth International Conference on Data Engineering, Orlando, Florida, USA, February 23–27, 1998, pp. 494–502, 1998.
3. P.N. Tan: *Discovery of Indirect Association and Its Applications*, Dr. Thesis, UNIVERSITY OF MINNESOTA, July, 2002.
4. Shinichi Hamano and Masako Sato: *Mining Indirect Association Rules*, in Proceedings of the 4th Industrial Conference on Data Mining, ICDM2004, Leipzig, Germany, July 4-7, 2004, Lecture Notes in Computer Science 3275, pp. 106–116, 2004.
5. A. Dempster, N. Laird, and D. Rubin: *Maximum likelihood from incomplete data via the em algorithm*, Journal of Royal Statistical Society, B(39):1–38, 1977.
6. D.McFadden: *Conditional logit analysis of qualitative choice behavior*, Frontiers in Economics, pp. 105–142, 1974.
7. Shichao Zhang and Chengqi Zhang: *Discovering Causality in Large Databases*, Applied Artificial Intelligence, Vol. 16, 5(2002): 333-358.
8. Anne Boomsma, Marijtje A. J. van Duijn and Tom A. B. Snijders: *Essays on Item Response Theory* Lecture Notes in Statistics **157**, 2001.