

Novel Approach to Tourism Analysis with Multiple Outcome Capability Using Rough Set Theory

Chun-Che Huang¹, Tzu-Liang (Bill) Tseng^{2†}, Kun-Cheng Chen³

¹ *Department of Information Management, National Chi Nan University,
University Road, Pu-Li, Nantau 545, Taiwan
E-mail: cchuang@ncnu.edu.tw*

² *Department of Industrial, Manufacturing and Systems Engineering,
The University of Texas at El Paso
El Paso, Texas 79968
E-mail: btseng@utep.edu*

³ *Department of Information Management, National Chi Nan University,
University Road, Pu-Li, Nantau 545, Taiwan
E-mail: stylishjackal@gmail.com*

Received 11 April 2015

Accepted 19 December 2015

Abstract

To explore the relationship between characteristics and decision-making outcomes of the tourist is critical to keep competitive tourism business. In investigation of tourism development, most of the existing studies lack of a systematic approach to analyze qualitative data. Although the traditional Rough Set (RS) based approach is an excellent classification method in qualitative modeling, but it is can't deal with the case of multiple outcomes, which is a common situation in tourism. Consequently, the Multiple Outcome Reduct Generation (MORG) and Multiple Outcome Rule Extraction (MORE) approaches based on RS to handle multiple outcomes are proposed. This study proposes a ranking based approach to induct meaningful reducts and ensure the strength and robustness of decision rules, which helps decision makers understand tourist's characteristics in a tourism case.

Keywords: Rough Sets, Tourism, Customer Satisfaction, Decision Making, Rule Induction.

1. Introduction

As tourism becomes the largest industry in the world, its impact on the global economy is widely recognized. To keep business competitive, one of the key and preliminary elements in the planning process is to study characteristics of customers in order to understand the relationship between tourist tasks and decision (Zalatan, 1998). That is to customize their product offerings and strengthen marketing policies to an appropriate degree in order to respond to the shifting importance and growing

complexity of customer choice. To date, most of the existing causal studies of tourism research apply quantitative models that use mathematical functions. These models are established on the basis of many statistical assumptions and limitations. However, it is unrealistic to assume that all tourism data are numeric. The tourism industry often possesses numerous qualitative data in flight scheduling databases, retail store employee files, and airline reservation repositories (Law & Au, 2000). The qualitative nature of the data makes them difficult to analyze by standard statistical techniques.

The rough set approach has been developed as an interesting technique for discovering knowledge from data (Pawlak, 1991a). It is suitable for processing qualitative information that is difficult to be analyzed by standard statistical techniques (Heckerman, Mannila, Pregibon, & Uthurusamy, 1997). Moreover, it provides an effective means for analysis of data by synthesizing or constructing approximations (upper and lower) of set concepts from the acquired data (Pal & Mitra, 2003). One of the reasons for using the rough set as an approach to mine data is due to the fact that the qualitative nature of the data (e.g., gender and vegetarian) being analyzed makes it difficult for applying standard statistical techniques (Heckerman *et al.*, 1997; Simoudis, Han & Fayyad, 1996).

Traditional rough set theory has no ability to consider multiple outcomes. It's hard to analyze a data set with more than two outcomes, only then do segments and reduct and provide one single consideration for each outcome. Therefore, this study proposes an approach that applies the rough sets theory for solving multiple outcome situations. Moreover, the direct use of the result, provided by the reduct generation procedure, may lead to many reducts that contain features that are not meaningful. In order to identify meaningful reducts and ensure the strength and stability of rules, this study proposes a ranking based approach to induct and apply each reduct to the test data set corresponding to the training set and collect information regarding the support and accuracy of the reduct in new data for rule extraction.

The objective of this study is to propose a novel reduct extraction approach for multiple outcomes in a tourism case. The aim is to understand the preferences of the customers there for exploring relationship between tourist tasks and decision. The remaining part of this study is organized as follows: Section 2 presents literature review related to tourism and several rough set based approaches. Section 3 develops the solution approach to operate multi-outcomes, while the tourism related case study and computation results based on the proposed approach are depicted in Section 4. Finally, Section 5 concludes the findings and provides suggestions for future research.

2. Literature Review

In this section, literature review related to tourism and rough set based approaches is presented.

2.1 Tourism

Nowadays tremendous growth of tourism as a field of study, coupled with increasing demand for tourism education has led to a heightened focus on research and publications (Jogaratnam, Chon, McCleary, Mena, & Yoo, 2005). This field has seen the development of numerous associations, journals and research consortia, and is linked to groups of researchers in other domains such as business and information technology (Racherla & Hu, 2010). Research topics in the tourism domain are various, for example, sustainability tourism (Divino & McAleer, 2009; Miller & Twining-Ward, 2006), marketing (Butcher, 2010; Frochot, 2010), data mining and knowledge discovering (Buultjens, Brereton, Memmott, Reser, Thomson & O'Rourke, 2010; Liao, Chen & Deng, 2010), quantitative approach (Enright & Newton, 2004; Jang, 2004) such as rules induction (Au & Law, 2002; Cardoso, Themido & Pires, 1999).

Walle (1997) argued that formal scientific methods are needed in tourism research. This would ensure generality of the empirical outcomes as well as maintain the quality and credibility of research findings. Ever since the introduction of formal statistical techniques to the tourism field, researchers have been focus on the application of mathematical functions to model the quantitative relationships between some phenomena in tourism data. At present, the quantity of tourism demanded modeling techniques consist primarily of time series model (Nelson, Dickey & Smith, 2009), needs-functions model to consider more fully the activities that occur in different settings from the market through to the destination and from one sector to another. (Pearce, 2008), dynamic model (Garín-Muñoz & Montero-Martín, 2007), Bayesian model (Wang, Hsiehb, Yehc & Tsai, 2004), and the use of statistical techniques to model the phenomena of numeric tourism oriented data in mathematical functions has achieved certain degree of success in relationship modeling and forecasting (Witt & Witt, 1995). Moreover, most of the existing studies on tourism demand forecasting apply economic models that use mathematical functions, which require many statistical assumptions and limitations (Law, Goh, & Pine, 2004).

However, it is unrealistic to assume that all tourism data are numeric. The tourism industry often possesses numerous qualitative data in flight scheduling databases, retail store employee files, and airline reservation

repositories (Law & Au, 2000). The qualitative nature of the data makes them difficult to analyze by standard statistical techniques. In addition, due to quick changing of information technology (IT) today, some techniques (i.e. rough sets and data mining tools) have become important research trends to both practitioners and academicians (Chen & Cheng, 2009). One of the solutions to resolve the issues is the rough set approach. A rough set approach can capture useful information from a set of mixed data and output this information in the form of decision rules (Pawlak, 1984). This makes the rough set approach an excellent rule induction method in qualitative modeling. Therefore, this study proposed a new approach that applies the rough sets theory. Having discussed Information System (IS), information reduction and decision rules induction in the next section.

2.2 Rough Set Approach

The Rough Set Model (RSM) was proposed by Pawlak in the early 1980s (Pawlak, 1982) as a powerful mathematical tool to deal with imprecise or vague concepts. It allows users to classify objects into sets of equivalent members based on their attributes (Dasgupta, 1988). In recent years, a rapid growth of interest in rough set theory and its applications are witnessed (Swiniarski & Skowron, 2003) through demonstrating that rough set model is applicable to a wide range of practical problems pertaining to supplier selection (Bai & Sarkis, 2010), investment portfolios (Shyng, Wang, Tzeng & Wu, 2010), marketing (Cheng, Chen, & Lin, 2010), EEG analysis (Ningler, Stockmanns, Schneider, Kochs & Kochs 2009), multiple criteria classification (Zhang, Shi & Gao, 2009; Dembczyński, Greco & Słowiński, 2009), expert system (Shao, Chu, Qiu, Gao & Yan., 2009), time series analysis (Yao & Herbert, 2009; Teoh, Cheng, Chu & Chen, 2008; Sarkar, 2006), image compression & segmentation (Mushrif & Ray, 2008; Petrosino & Ferone, 2008), boolean reasoning (Pawlak & Skowron, 2007).

The rough set theory offers a large collection of tools for knowledge discovery from data. Many of these methods, like decision rule induction, classifier construction, discretization, decision tree construction, and representative association rule induction, are based on computing the most relevant sets of attributes called reducts (Nguyen, 2003). A reduct is defined as a

minimal sufficient subset of a set of attributes, which has the same ability to discern concepts as when the full set of attributes is used (Pawlak, 1991a; Ziarko, 1994). Basically, the reducts represent necessary condition attributes in decision making. Furthermore, a subset of the attributes has more than one reduct; hence the simplification of reducts (called reduct rules) is required to extract significant rules (called decision rules) does not yield unique results. Therefore, a rough set based approach often contains two parts: Reduct generation and decision rule extraction.

2.2.1 Reduct Generation

According to the rough set theory, $I = \{U, A\}$ is an information system, where U is a finite set of objects and A is a finite set of attributes. With every attribute $a \in A$, a set of its values V_a is associated. Assume $A = C \cup D$, $B \subset C$, where B is a subset of C ; the positive region $POS_B(D) = \{x \in U: [x]_B \subset D\}$ can be defined. The positive region $POS_B(D)$ includes all objects in U which can be classified into classes of D , in the knowledge B . The degree of dependency between B and D can be defined as:

$$K(B, D) = \frac{\text{card}(POS_B(D))}{\text{card}(POS_C(D))}, \quad (1)$$

where card yields the set cardinality. In general, if $K(B, D) = K(C, D)$, and $K(B, D) \neq K(B - \{a\}, D)$, for any $a \in B$ are held; then B is a reduct of C . Since a reduct (B) preserves the degree of dependency with respect to D and a reduct (B) is a minimal subset, any further removal of condition attributes will change the degree of dependency. The most useful procedure for determining the reducts can be referred in the (Pawlak, 1991b).

In order to find dispensable attributes, an examination of each attribute of the object is required. It might also have to drop one attribute at a time and check whether the intersection of the remaining attributes is still included in the decision attribute. In most case, if one or two features in a reduct formed, the procedure stops because too many features to form a reduct rule will reduce advantage of the rough set theory. The reduct rule would be useless without finding critical features. However, the case of multiple outcomes fairly common in the real world hasn't been seriously considered in traditional rough set theory. Bai and Sarkis (2010) define the data set with ten conditional attributes and two decision attributes for the illustrative application.

The reduct generation procedure is required for improvement and evolving into a powerful tool for decision makers and researchers, especially in complex decision environments associated with multiple outcomes. For example, the multiple outcomes table in Table 1 represents the final classifications. Objects 1 and 2 have the same value of features but difference outcomes ($\{X_1\} = [R, 3, 0]_{o1,o2,o3} \neq \{X_2\} = [A, 3, 0]_{o1,o2,o3}$, and so do objects 3 and 4 in Table 1. After performing the reduct search procedure, no other reduct was generated.

Table 1. The multiple outcomes.

U	F ₁	F ₂	F ₃	F ₄	O ₁	O ₂	O ₃
X ₁	B	L	Y	H	R	3	0
X ₂	B	L	Y	H	A	3	0
X ₃	O	M	N	L	A	2	0
X ₄	O	M	N	L	R	2	1

2.2.2 Rule Induction

The direct use of the results provided by the reduct generation procedure may lead to many reducts that contain features that are not meaningful. And another challenge of rough set based rule induction is the fact that finding all reducts is NP-hard (Ziarko, 1990). Law and Au (2000) present a decision rules induction approach that incorporates rough set theory to model the relations that exist among a set of mixed numeric and non-numeric tourism shopping data. The decision rules induction approach was formed a decision rule is in the form of IF_condition(s)_THEN_decision(s). But without consideration of threshold (e.g., a strength index), it may generate a few redundant rules. Furthermore, the relationship among the reducts is not explored. In order to identify meaningful reducts and ensure the strength and stability of rules, a comparison between the rules and how they react with data must be carried out. In previous studies, a rule extraction algorithm (Tseng and Huang 2007) has been developed for discovering preference-based rules according to the reducts which contain the maximum strength index, and relatively small number of induced rules with reasonable domain interpretation. However, the desired

reducts may not be unique. Extension of the strength index is required for improvement.

In addition, the RS approach has to encode outcomes of decision table to distinguish the individual outcome features, for example, the outcome (1, 2) is different from (1, 3), but both are included in (1, *). It requires a coding tool to represent the hierarchy of the outcomes. From those points of view, this study proposed a multiple outcome reduct generation algorithm and a multiple outcome rule extract algorithm based on rough set theory to generated multiple outcome rules.

Through literature review, the desires to developing the multiple outcome reduct generation algorithm and multiple outcome rule extraction algorithm based rough set are as follows:

1. The tourism industry often possesses numerous qualitative data in flight scheduling databases, retail store employee files, and airline reservation repositories (Law & Au, 2000). The qualitative nature of the data makes them difficult to analyze by standard statistical techniques. One of the solutions to resolve the issues is the rough set approach, which can capture useful information from a set of mixed data and output this information in the form of decision rules (Pawlak, 1984).
2. Although numerous machine learning approaches have considered multiple outcomes (Richard et al., 2013; Thomas et al., 2008), unfortunately most rough set approaches do not consider the issue of multiple outcomes. For example, if there is a dataset with three features and two outcomes, the traditional reduct generation procedure has to be performed twice for each outcome. In addition, a decision rule always consists of a single outcome. It is not easy to understand the commonality of rules in those two outcomes;

Therefore, next, the multiple outcomes problem is defined.

3. Multiple Outcome Problem

In this section, a multiple outcome table for the representation of the relationship between condition features and outcome features is used. Alternatively, a feature and an attribute are exchanable due to their similarity. In Table 2, the element (e_j) denotes the value of feature (F_j) ($j = 1$ to m) that an object (tuple) (X_i) ($i = 1$ to n) contains, while O_j ($j = 1$ to p) depicts the

different outcome set of the corresponding tuple. In addition, the number of values is incorporated in this table.

In this study, the RS multiple outcomes problem is defined as:

Given

- An information set $I = (U, A)$, where U is a finite set of objects and A is a finite set of features. The elements of A are called condition features.
- A decision table $I = (U, A \cup \{d_1, d_2, \dots, d_n\})$, where any d_i in $\{d_1, d_2, \dots, d_n\} \notin A$ is a distinguished outcome feature.

Objective

Minimize the subset of U , in which the specific outcome feature of decision rules (r_i) has the highest Strong Index (SI) and the maximal level number.

Subject to the following constraints:

1. $A = C \cup D$,
2. $B \subset C$
3. $POS_B(D) = \{x \in U: [x]B \subset D\}$
4. For any $a \in B$ of C , $K(B, D) = K(C, D)$, and $K(B, D) \neq K(B - \{a\}, D)$

where, C is the condition domain and D is the outcome domain.

$POS_B(D)$ is the positive region and includes all objects in U which can be classified into classes of D , in the knowledge B .

$$K(B, D) = \frac{card(POS_B(D))}{card(POS_C(D))}$$

is the degree of

dependency between B and D

Attribute $a \in A$, set of its values

4. Solution Approach

The solution approach includes two algorithms: The Multiple Outcome Reduct Generation (MORG) algorithm and Multiple Outcome Rule Extraction (MORE) algorithm.

4.1 The MORG Algorithm

The MORG algorithm generates reducts with multiple outcomes. All steps are described as follows: (1) Input training data set and initialize each object in the beginning; (2) Generate interaction set and combination set of outcomes; (3) Check all subsets of value sets of outcomes, if the interaction set is a subset of value set of outcomes, and then generate combination set of features; (4) Check all subsets of value sets of outcomes, if the value set of features is a subset of value set of outcomes, and then output the result into reduct table; (5) Stop the multiple outcome reduct generation procedure and repeat the above steps until running out all objects. Figure 1 illustrates the step by step instructions to describe the MORG algorithm.

Table 2. Fundamental structure of a multiple outcome table.

Object (X_i)	Condition features (F_j)				Outcome(O_j)	Outcomes features (O_p)			
	F_1	F_2	...	F_m		O_1	O_2	...	O_p
1	e_{11}	e_{12}	...	e_{1m}	1	o_{11}	o_{12}	...	o_{1p}
2	e_{21}	e_{22}	...	e_{2m}	2	o_{21}	o_{22}	...	o_{2p}
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
n	e_{n1}	e_{n2}	...	e_{nm}	n	o_{n1}	o_{n2}	...	o_{np}

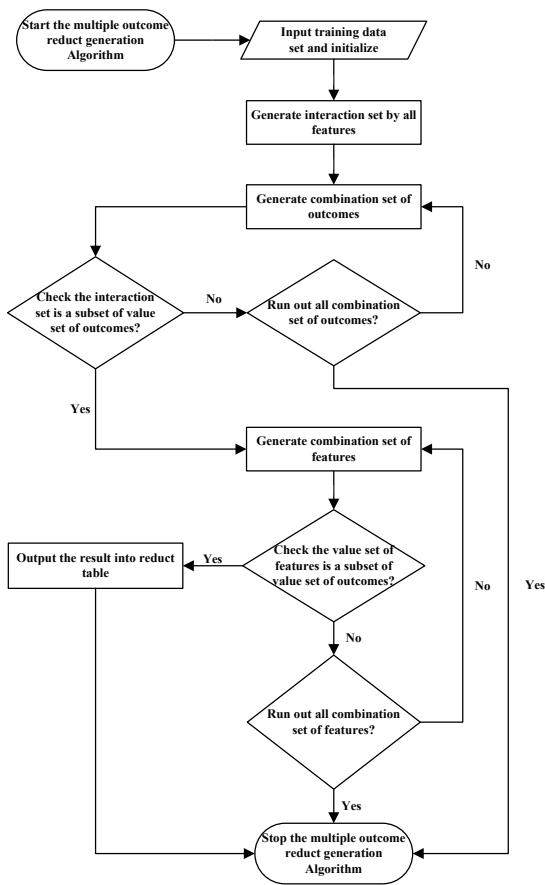


Fig. 1. The flow chart to describe the MORG algorithm.

Notations:

- T_a : The training data set used for training rules.
- T_{reduct} : The ad hoc data set used for store the results.
- Out_N : The total number of outcomes.
- Fea_N : The total number of features.
- $F_{i,j}$: The objects where each F_j feature contains $F_{i,j}$.
- $O_{i,p}$: The objects where each O_p outcome contains $O_{i,p}$.
- $F_Interaction()$: A generation features interaction set function. E.g., $F_Interaction(i) = \{[F_{i,1}]\} \cap \{[F_{i,2}]\} \cap \{[F_{i,3}]\} \cap \dots \cap \{[F_{i,final}]\} = [F_i] \cap T_a$.
- $F_Combination()$: A generation combination set function. E.g., $F_Combination(3, 2)$ can generate a set of two possible from a set $\{1, 2, 3\}$, and the result is $\{12, 13, 23\}$, i.e., $F_Combination(3, 2)$ is equal to C_2^3 .

The Multiple Outcomes Reduct Generation (MORG)

Procedure

Input: T_a .

Output: T_{reduct} .

- Step 1. Initialization: List all objects in T_a .
- Step 2. Generate the reducts for each object.
- Step 2.1. for $i = 1$ to n do
 - Loop:
 - Step 2.2. Set $new_Out_Set = \emptyset$
 - Step 2.3. Set $mini_Set = F_Interaction(i)$
 - Step 2.4. for $out = Out_N$ to 1
 - Step 2.5. Set $comb_Out_Set = F_Combination(Out_N, out)$
 - Step 2.6. for each p in $comb_Out_Set$
 - Step 2.7. if $mini_Set \subset \{[O_p] \cap [O_{i,p}]\}$ then Add p into new_Out_Set
 - endifor
 - Step 2.8. if $new_Out_Set = \emptyset$ then Go to Step2.4
 - Step 2.9. for $fea = 1$ to $Fea_N - 1$
 - Step 2.10. Set $comb_Fea_Set = F_Combination(Fea_N, fea)$
 - Step 2.11. for each p in new_Out_Set
 - Step 2.12. for each j in $comb_Fea_Set$
 - Step 2.13. if $\{[F_j] \cap [F_{i,j}]\} \subset \{[O_p] \cap [O_{i,p}]\}$
 - Step 2.14. then the reducts for X_i is formed
 - Step 2.15. else the reducts for X_i is not formed
 - endifor
 - endifor
 - if any reducts for X_i is formed
 - then break Loop
 - else the reducts for X_i is not formed
 - endifor
 - endifor
 - endifor
- Step 3. Termination: Stop and output the results into T_{reduct} .

4.2 The MORE Algorithm

The steps in the MORE algorithm include: (1) Input training data set, (2) Test data set, (3) Reduct table, (4) Rule table, and initialization in the beginning. Moreover, calculation of four indexes (support and accuracy of the training data set, support and accuracy of the test data set) into the information table for each reduct is required. For each reduct, it is required to calculate four indexes. If any values of the index is lower than the threshold value determined by the expert, then remove the reduct from information table. The decision maker requires to sort the rank of four indexes

and the obtained ranking table id. Figure 2 presents the step by step procedure to describe the MORE algorithm. In order to determine which rules are the strongest ones, comparison between the rules and how they react with data must be made (Yao & Herbert, 2009). Individual rules are ranked according to their strengthes (support and accuracy) and stability, which refers to the changes to support and accuracy from the training data set to testing data set. An overall rank can be determined based on the support and accuracy rankings.

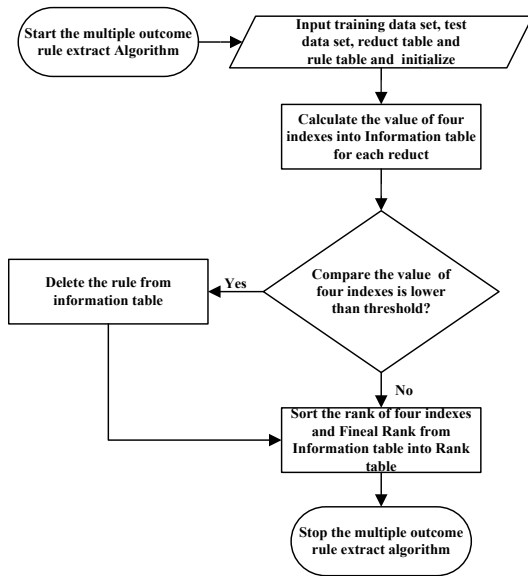


Fig. 2. The flow chart of the MORE algorithm.

Notations:

- T_v : The test data set used for validate rules.
- T_{info} : Information table (There are six columns in this table: support and accuracy of training data set, support and accuracy of test set, change in support (%) and change in accuracy (%)).
- T_{Rank} : Ranking table (There are five columns in this table: total support, average accuracy, change in support, change in accuracy and final rank for each rule.).
- $F_CompareF()$: A function is used for comparing the value of features of a particular rule with value of features of T_a and T_v .
- $F_CompareO()$: A function is used for comparing the value of outcomes of a particular rule with value of outcomes of T_a and T_v .
- $F_Threshold()$: A function is used for extracting the

- desired rule.
- Sup_N : The total mach number of specific rules in the data set.
- Acc_N : The total mach number of specific rules in the data set divided by the total mach number of features of specific rule.

The Multiple Outcome Rule Extract (MORE) Procedure

- Input: T_a, T_v, T_{reduct} and T_{rule} .
- Output: T_{info} and T_{Rank} .
- Step 1. Initialization: List all reduct from T_{reduct} and combine duplicate rule into T_{rule} .
- Step 2. Calculate four indices value into T_{info} for each reduct.
 - Step 2.1. for each row i in T_{rule}
 - Step 2.2. Set $Sup_N = 0$ and $Acc_N = 0$
 - Step 2.3. for each row j in $(T_a) \& (T_v)$
 - Step 2.4. if $F_CompareF(i, j) = 1$
 - Step 2.5. Set $Sup_N +=$ number of objects of row j
 - Step 2.6. if $F_CompareO(i, j) = 1$
 - Step 2.7. Set $Acc_N +=$ number of objects of row j
 - end for
 - Step 2.8. Set $Support = Sup_N /$ total no. of objects. & $Accuracy = Acc_N / Acc_N$
 - Step 2.9. Assign the values of support and accuracy into T_{info} end for
 - Step 2.10. Calculator change in support (%) and change in accuracy (%) for each rule.
- Step 3. Extract the rule by $F_Threshold()$.
- Step 4. Sort the rank of four indices and Finalize the ranking from T_{info} into T_{Rank} .
- Step 5. Output T_{info} and T_{Rank} .

5. Case study

The objective of this case study is to extract the rules to reveal the tourist’s characteristics. The experiment results focus on the use of rules to classify the different features of the tourists and find out what they really care about; to help the tourism service providers satisfy the tourists with the desired travel package.

Middelkoop (2004) argues that tourists do not necessarily maximize their utility in selecting a travel mode; rather, their choice behavior is context dependent. Given particular conditions related to their family, the environment, and other aspects of the tourist experience, they are assumed to use different heuristics. Taking a vacation involves a decision-making process which may be simple or complex. Therefore, to study characteristics

of tourist and to understand the relationship between tourism and decision, it is required to induce decision rules through historical data sets. In short, different characteristics of a tourist might influence the decision-making. The characteristics of a tourist are such as household income, number of trips yearly, presence of children, tourist's age, vegetarian, pet ownership, etc., all have a certain influence on the decision-making process (Harcar, Spillan & Kucukemiroglu, 2005; Zalatan, 1998). Moreover, Zalatan (1998) segments tourism decision-making tasks into 17 categories (see Table 3). And in general,

arranging transportation, selecting restaurants and accommodation, purchase insurance, choosing the destination, purchasing tickets, etc., are included in the tourist package deal. So six characteristics are used in features and six tasks are used in outcomes in this study. Tables 4 and 5 present the description and the domain range of each input feature and the final outcomes, respectively. Two data sets are used for validating the solution approach, a training data set (20 objects) and a testing data set (10 objects).

Table 3. The tourism decision-making tasks list.

1. Initial Trip Tasks	I.	Determining Vacation Date
	II.	Selecting Destination
	III.	Collecting Information
2. Financing Tasks	I.	Arranging Financing
	II.	Tourists Cheques
	III.	Purchasing Travel Package
	IV.	Purchasing Tickets
3. Pre-departure Tasks	I.	Transportation Arrangements
	II.	Accommodation Arrangements
	III.	Preparing Luggage
	IV.	Other Tasks (e.g., medical, insurance)
4. Destination Tasks	I.	Selecting Tourist Sites
	II.	Length of Time at each Site
	III.	Selecting Restaurants
	IV.	Selecting Accommodation
	V.	Shopping Expenditures
	VI.	Other Decisions (e.g., special events)

Table 4. Categorization of the features.

No.	Features name and domain					
	F ₁ ^a	F ₂ ^b	F ₃	F ₄	F ₅	F ₆
	Household income	No. of trips	Number of kids	Age	Vegetarian	Pet ownership
1	< 100 (L)	< 1 (N)	< 1 (N)	< 25 (YO)	Yes (Y)	Yes (Y)
2	100 ~ 200 (M)	1 ~ 2 (L)	1 ~ 2 (L)	25 ~ 50 (MA)	No (N)	No (N)
3	> 200 (H)	3 ~ 4(M)	3 ~ 4(M)	> 50 (OA)		
4		> 4 (H)	> 4 (H)			

Note: a: F₁ unit is US dollars. b: F₂* unit is a year.

Table 5. Categorization of the outcomes.

No.	Outcomes name and domain					
	O ₁	O ₂	O ₃	O ₄	O ₅	O ₆
	Selecting Destination	Purchasing Tickets	Arrangements Transportation	Arrangements Accommodation	Insurance	Selecting Restaurants
1	Short (S)	Yes (Y)	KTX (K)	Yes (Y)	Yes (Y)	Yes (Y)
2	Medium (M)	No (N)	Bus (B)	No (N)	No (N)	No (N)
3	Long (L)		Train (T)			
4			Car (C)			
5			Other (O)			

After performing the reduct generation procedure in using training data set, the list of resulting reducts in Table 6 is obtained. The heuristic procedure is applied to determine the reducts.

Next the MORE algorithm is applied. The information in Table 7 is computed and the ranking of the rules in Table 8 are obtained. Table 7 clearly shows the strength or weakness of the rule through its relative strength and stability measure. Therefore, through the threshold function, weak and unstable rules are not extracted and incorporated into the ranking table. Note that the unselected rules are marked in gray in Table 7. Moreover, Table 8 shows the ranking of the rules in detail. Basically, the heuristic procedure is applied to extract the rules. For example, there is significant discrepancy (e.g., 20% difference) in the change of support and accuracy from the training data to the test data. These rules, for example, #2, #3, #6, etc. (in grey color) are concluded as an unstable rule and excluded in decision rules, since the margin of error of 20% is considerably high from professional perspective in

the field. The rules, for example #1, #9, #12, etc in Table 7 show the low percentage of change in support and no change in accuracy. Basically, they are highly stable and will be included in final decision rules. Furthermore, a threshold function, F_Threshold(), to extract the rules is described as follows: If the candidate rule and its accuracy is higher than 0.7 and then the rule is extract as final decision rules.

According to the final rankings, Table 9 shows the first five rules with their rankings, where the relationships between the characteristics of the tourist's and tourism decision-making tasks. Rules could be interpreted as if the tourist conforms to those features, and then infers he/she would be more concerned about what tasks are on travel arrangements. For example, the fourteenth and the fifteenth rules are the most accurate and stable. The fourteenth rule shows if a tourist has a household income of \$100~200K and has more than four kids; then he/she might care about some tourism tasks like take a short destination trip by train, purchase pre-arranged accommodation and insurance.

Table 6. The list of resulting reducts (T_{reduct}).

Object No.	Reduct No.	Rule No.	F ₁	F ₂	F ₃	F ₄	F ₅	F ₆	O ₁	O ₂	O ₃	O ₄	O ₅	O ₆
X ₁ & X ₂ & X ₃	1	1	*	*	*	*	*	Y	*	*	C	Y	*	Y
	1	2	M	*	M	*	*	*	S	N	C	Y	*	Y
X ₄ & X ₅	2	3	*	M	*	*	*	Y	S	N	C	Y	*	Y
	3	4	*	*	M	MA	*	*	S	N	C	Y	*	Y
	4	5	*	*	M	*	*	Y	S	N	C	Y	*	Y
	1	6	L	H	*	*	*	*	M	Y	C	Y	Y	Y
X ₆	2	7	L	*	H	*	*	*	M	Y	C	Y	Y	Y
	3	8	L	*	*	*	*	Y	M	Y	C	Y	Y	Y
	4	9	*	H	H	*	*	*	M	Y	C	Y	Y	Y
	5	10	*	H	*	*	*	Y	M	Y	C	Y	Y	Y
	6	11	*	*	H	YO	*	*	M	Y	C	Y	Y	Y
	7	12	*	*	H	*	*	Y	M	Y	C	Y	Y	Y
	8	13	*	*	*	YO	*	Y	M	Y	C	Y	Y	Y
	X ₇ & X ₈	1	14	M	*	H	*	*	*	S	*	T	Y	Y
2		15	*	L	H	*	*	*	S	*	T	Y	Y	*
3		16	*	L	*	*	*	N	S	*	T	Y	Y	*
4		17	*	*	H	MA	*	*	S	*	T	Y	Y	*
5		18	*	*	H	*	*	N	S	*	T	Y	Y	*
X ₉ & X ₁₆	1	19	L	N	*	*	*	*	*	N	*	*	Y	*
	2	20	L	*	L	*	*	*	*	N	*	*	Y	*
	3	21	L	*	*	*	*	N	*	N	*	*	Y	*
	4	22	*	N	L	*	*	*	*	N	*	*	Y	*
	5	23	*	*	L	YO	*	*	*	N	*	*	Y	*
X ₁₀ & X ₁₁ & X ₁₄	1	24	*	*	N	*	*	*	L	*	O	*	*	*
X ₁₂	1	25	H	*	L	*	*	*	M	Y	B	N	N	Y
	2	26	*	H	L	*	*	*	M	Y	B	N	N	Y
	3	27	*	H	*	OA	*	*	M	Y	B	N	N	Y
	4	28	*	*	L	OA	*	*	M	Y	B	N	N	Y
	5	29	*	*	*	OA	N	*	M	Y	B	N	N	Y
X ₁₃	1	30	H	N	*	*	*	*	S	N	T	N	N	N
	2	31	*	N	M	*	*	*	S	N	T	N	N	N
	3	32	*	*	M	YO	*	*	S	N	T	N	N	N
X ₁₅	1	33	H	*	N	*	*	*	L	N	O	N	Y	Y
	2	34	*	H	N	*	*	*	L	N	O	N	Y	Y
	3	35	*	*	N	YO	*	*	L	N	O	N	Y	Y
X ₁₇ & X ₁₈ & X ₁₉ & X ₂₀	1	36	*	*	*	*	Y	*	*	*	Y	*	Y	

Table 7. The information table (T_{info}).

Rule no.	Training data set		Test date set		Change in support (%)	Change in accuracy (%)
	Support	Accuracy	Support	Accuracy		
1	0.3140	1	0.25	1	-6.4050	0
2	0.0992	1	0	0	-9.9174	-100
3	0.0992	1	0	0	-9.9174	-100
4	0.0992	1	0.1154	0.75	1.6211	-14.2857
5	0.0992	1	0.1154	0.75	1.6211	-14.2857
6	0.0248	1	0	0	-2.4793	-100
7	0.0248	1	0	0	-2.4793	-100
8	0.0248	1	0	0	-2.4793	-100
9	0.0248	1	0.0962	1	7.1360	0
10	0.0248	1	0.0962	0.3846	7.1360	-44.4444
11	0.0248	1	0	0	-2.4793	-100
12	0.0248	1	0.0962	1	7.1360	0
13	0.0248	1	0	0	-2.4793	-100
14	0.1240	1	0.1346	1	1.0648	0
15	0.1240	1	0.1346	1	1.0648	0
16	0.1240	1	0.1346	0.5384	1.0648	-30
17	0.1240	1	0.1346	0.5833	1.0648	-26.3158
18	0.1240	1	0.1346	0.5833	1.0648	-26.3158
19	0.0579	1	0	0	-5.7851	-100
20	0.0579	1	0	0	-5.7851	-100
21	0.0579	1	0.0962	1	3.8303	0
22	0.0579	1	0	0	-5.7851	-100
23	0.0579	1	0	0	-5.7851	-100
24	0.2231	1	0.2115	1	-1.1602	0
25	0.0496	1	0.1346	0.7	8.5029	-17.6471
26	0.0496	1	0	0	-4.9587	-100
27	0.0496	1	0	0	-4.9587	-100
28	0.0496	1	0	0	-4.9587	-100
29	0.0496	1	0	0	-4.9587	-100
30	0.0413	1	0.0577	0.3	1.6370	-53.8462
31	0.0413	1	0.0577	0.2	1.6370	-66.6667
32	0.0413	1	0.1154	1	7.4062	0
33	0.0165	1	0.1154	1	9.8856	0
34	0.0165	1	0	0	-1.6529	-100
35	0.0165	1	0	0	-1.6529	-100
36	0.1901	1	0.3462	0.72	15.6071	-16.2791

Table 8. The ranking of the rules (T_{Rank}).

Rule no.	Total support	Total accuracy	Change in support	Change in accuracy	Final rank
1	1	1	7	1	4
4	6	10	4	10	10
5	6	10	4	10	10
9	11	1	8	1	6
12	11	1	8	1	6
14	4	1	1	1	1
15	4	1	1	1	1
21	9	1	6	1	5
24	3	1	3	1	3
25	8	13	11	13	13
32	10	1	10	1	8
33	11	1	12	1	9
36	2	12	13	12	12
Weight	1	0.8	0.9	0.7	

Table 9. Induced decision rules.

Rule No.	Rank	Contents of rule
14	#1	IF (Household income = 100 ~ 200) AND (No. of kids = > 4) THEN (Selecting Destination = Short) AND (Arrangements Transportation = Train) AND (Arrangements Accommodation = Yes) AND (Insurance = Yes)
15	#1	IF (No. of trips = 1 ~ 2) AND (No. of kids = > 4) THEN (Selecting Destination = Short) AND (Arrangements Transportation = Train) AND (Arrangements Accommodation = Yes) AND (Insurance = Yes)
24	#3	IF (Number of kids = < 1) THEN (Selecting Destination = Long) AND (Arrangements Transportation = Other)
1	#4	IF (Pet ownership = Yes) THEN (Arrangements Transportation = Car) AND (Arrangements Accommodation = Yes) AND (Selecting Restaurants = Yes)
21	#5	IF (Household income = < 100) AND (Pet ownership = No) THEN (Purchasing Tickets = No) AND (Insurance = Yes)

Table 10. Comparison of coverage between the MORE algorithm and the traditional rough set reduct procedure in this case study.

Algorithm	Number of reducts	Number of objects can generate reduct	Total number of Objects	Coverage
The MORE	36	20	121	100%
The reduct generation procedure through RST	19	4	16	13%

Table 11. Comparison of coverage between the MORE algorithm and the traditional rough set reduct procedure in the random dat set

Algorithm	10 ³ objects		10 ⁶ objects		10 ⁹ objects	
	4 outcome	10 outcomes	4 outcome	10 outcomes	4 outcome	10 outcomes
The MORE	83.3%	81.5%	78.5%	76.0%	72.3%	71.5%
The traditional reduct generation procedure	15%	14.7%	14.3%	13.1%	13.8%	13.7%

Table 10 depicts comparison of coverage between the MORE algorithm and the traditional rough set reduct procedure in this case.

Using the traditional RS approach, the results show there are only four objects can generate reducts. Although the MORE algorithm generate nineteen reducts and each reudct has 100% accuracy, but the total number of four objects is 13% (16 divided by 121) of the total number of training data. That means these reducts can make some special cases in the training data set only.

In addition, to validate practicability of the proposed approach, an experiment with 23 conditional features in different data was made. The benefits of the proposed approach are to extract more rules such that the coverage is higher than the traditional RS approach that distinguishes the individual outcome features. For example, the outcome (1, 2) is different from (1, 3), but it can not both be included in (1,*) with one unique entry for the outcome feature. Comparison of coverage between the MORE algorithm and the traditional rough set reduct procedure is presented in Tab 11.

6. Conclusion

In tourism, most of the existing studies lack a method to analyze the qualitative nature of the data. Although the traditional rough set approach is an excellent classification method in qualitative modeling, but the approach has no consideration for multiple outcome situations. This study presents a novel solution approach, namely the MORG and MORE algorithms, to generate reduct and extraction rule. Moreover, a case study of tourism with multiple outcomes is used as an example and is compared the results with a traditional rough set and rule extraction algorithm. The contributions of this paper are summarized as follows: (1) The MORG algorithm overcomes traditional rough set's drawbacks. The algorithm is not limited to analyze data set with more than a single outcome; (2) The MORE algorithm has stronger rule extraction capability. The algorithm is able to extract decision rules based both on strength and stability. And determines which rules are the strongest by rank.

To date, there are vey few work on multiple outcome rule induction, future research could focus on the following: (1) Using sensitivity methodology to assess

the effects of features in tourism decision-making tasks; (2) Defining threshold values and weight by experts that are used to extract and rank rules; (3) In addition, unknown or incorrect features and outcome values often occur in real world applications. How to pre-process the data set is important; (4) Another consideration in reducing the set of rules is to determine 'infeasible' rules. Even the initial rules generated may miss some lower level performance results that cause infeasibility to occur (Bai & Sarkis, 2010); (5) Performance measurement of the proposed approach could include the precision recall, Receiver Operating Characteristic (ROC) curve; (6) More data sets and larger sample size will be used to validate the model.

Acknowledgments

This work was partially supported by the Ministry of Science and Technology in Taiwan (NSC 102-2410-H-260-045-MY3, MOST 105-2410-H-260-019 -MY2) and the National Science Foundation (DUE-TUES-1246050). The authors wish to express sincere gratitude for their financial support.

References

1. Au, N., & Law, R., (2002). "Categorical classification of tourism dining," *Annals of Tourism Research*, 29(3), 819-833.
2. Bai, C., & Sarkis, J., (2010). "Green supplier development: analytical evaluation using rough set theory," *Journal of Cleaner Production*, 18(12), 1200-1210.
3. Butcher, K., (2010). "Marketing in Travel and Tourism," *Annals of Tourism Research*, 37(1), 280-281.
4. Buultjens, J., Brereton, D., Memmott, P., Reser, J., Thomson, L., & O'Rourke, T., (2010). "The mining sector and indigenous tourism development in Weipa, Queensland," *Tourism Management*, 31(5), 597-606.
5. Cardoso, M. G. M. S., Themido, I. H., & Pires, F. M., (1999). "Evaluating a clustering solution: An application in the tourism market," *Intelligent Data Analysis*, 3(6), 491-510.
6. Chen, Y. S., & Cheng, C. H., (2009). "Evaluating industry performance using extracted RGR rules based on feature selection and rough sets classifier," *Expert Systems with Applications*, 36(5), 9448-9456.
7. Cheng, J. H., Chen, H. P., & Lin, Y. M., (2010). "A hybrid forecast marketing timing model based on probabilistic neural network," *rough set and C4.5, Expert Systems with Applications*, 37(3), pp. 1814-1820.
8. Dasgupta, P., (1988). "Rough sets and information retrieval," *Proceedings of the 11th annual international ACM SIGIR conference on Research and development in information retrieval*, 567-581.
9. Dembczyński, K., Greco, S., & Słowiński, R., (2009). "Rough set approach to multiple criteria classification with imprecise evaluations and assignments," *European Journal of Operational Research*, 198(2), 626-636.
10. Divino, J. A., & McAleer, M., (2009). "Modelling sustainable international tourism demand to the Brazilian Amazon," *Environmental Modelling & Software*, 24(12), 1411-1419.
11. Enright, M. J., & Newton, J., (2004). "Tourism destination competitiveness: a quantitative approach," *Tourism Management*, 25(6), 777-788.
12. Frochot, I., (2010). "Tourism and hospitality marketing – A global perspective," *Tourism Management*, 31(5), 692-693.
13. Garín-Muñoz, T., & Montero-Martín, L. F., (2007). "Tourism in the Balearic Islands: A dynamic model for international demand using panel data," *Tourism Management*, 28(5), 1224-1235.
14. Harcar, T., Spillan, J. E., & Kucukemiroglu, O., (2005). "A multi-national study of family decision-making," *Multinational Business Review*, 1(2), 3-21.
15. Heckerman, D., Mannila, H., Pregibon, D., & Uthurusamy, R., (1997), *Proceedings of the Third International Conference on Knowledge Discovery and Data Mining*, Menlo Park, AAAI Press, CA.
16. Jang, S. C. S., (2004). "Mitigating tourism seasonality: A quantitative approach," *Annals of Tourism Research*, 31(4), 819-836.
17. Jogaratnam, G., Chon, K., McCleary, K., Mena, M., & Yoo, J., (2005). "An analysis of institutional contributors to three major academic tourism journals: 1992–2001," *Tourism Management*, Vol.26, No.5, pp. 641-648.
18. Law, R., & Au, N., (2000). "Relationship modeling in tourism shopping: a decision rules induction approach," *Tourism Management*, 21(3), 241-249.
19. Law, R., Goh, C., & Pine, R., (2004). "Modeling Tourism Demand: A Decision Rules Based Approach," *Journal of Travel & Tourism Marketing*, 16(2), 61-69.
20. Liao, S. H., Chen, Y. J., & Deng, M. Y., (2010). "Mining customer knowledge for tourism new product development and customer relationship management," *Expert Systems with Applications*, 37(6), 4212-4223.
21. Middelkoop, M. V., Borgers, A., & Timmermans, H., (2004). "Inducing Heuristic Principles of Tourist Choice of Travel Mode: A Rule-Based Approach," *Journal of Travel Research*, 42(1), 75-83.
22. Miller, Graham., & Twining-Ward, Louise., (2006). *Monitoring for a Sustainable Tourism Transition: The Challenge of Developing and Using Indicators*, CABI Publishing, Wallingford, Oxfordshire, UK; Cambridge, MA.
23. Mushrif, M. M., & Ray, A. K., (2008). "Color image segmentation: Rough-set theoretic approach," *Pattern Recognition Letters*, 29(4), 483-493.

24. Nelson, L. A., Dickey, D. A., & Smith, J. M., (2009). "Estimating time series and cross section tourism demand models: Mainland United States to Hawaii data," *Tourism Management*, 1–11.
25. Nguyen, H. S., (2003). "On the Decision Table with Maximal Number of Reducts," *Electronic Notes in Theoretical Computer Science*, 82(4), 1-8.
26. Ningler, M., Stockmanns, G., Schneider, G., Kochs, H. D., & Kochs, E., (2009). "Adapted variable precision rough set approach for EEG analysis," *Artificial Intelligence in Medicine*, 47(3), 239-261.
27. Pal, S., & Mitra, P. (2003). Rough sets, EM algorithm, MST and multispectral image segmentation. *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, 589-589.
28. Pawlak, Z., & Skowron, A., (2007). "Rough sets and Boolean reasoning," *Information Sciences*, 177(1), 41-73.
29. Pawlak, Z., (1982). "Rough Sets," *International journal of Computer and Information Sciences*, 11(5), 341-356.
30. Pawlak, Z., (1984). "Rough classification," *International Journal of Man-Machine Studies*, 20(5), 469-483.
31. Pawlak, Z., 1991a, *Rough Sets*, Dordrecht, Kluwer Academic Publishers, the Netherlands.
32. Pawlak, Z., 1991b, *Rough sets: Theoretical aspects of reasoning about data*, Kluwer Academic Publishers, Dordrecht, Boston, and London.
33. Pearce, D. G., (2008). "A needs-functions model of tourism distribution," *Annals of Tourism Research*, 35(1), 148-168.
34. Petrosino, A., & Ferone, A. (2008). "Rough fuzzy set-based image compression," *Fuzzy Sets and Systems*, 160(10), 1485-1506.
35. Racherla, P., & Hu, C., (2010). "A social network perspective of tourism research collaborations," *Annals of Tourism Research*, Article in Press, Corrected Proof.
36. Richard W., Cynthia J. G., Robert J. L., M. Alan B. and Til S., (2013). "Variable selection for propensity score models when estimating treatment effects on multiple outcomes: a simulation study," *Pharmacoepidemiology and Drug Safety*, 22(1), 77-85.
37. Sarkar, M., (2006). "Ruggedness measures of medical time series using fuzzy-rough sets and fractals," *Pattern Recognition Letters*, 27(5), 447-454.
38. Shao, X. Y., Chu, X. Z., Qiu, H.B., Gao, L., & Yan, J., (2009). "An expert system using rough sets theory for aided conceptual design of ship's engine room automation," *Expert Systems with Applications*, 36(2), 3223-3233.
39. Shyng, J. Y., Shieh, H. M., Tzeng, G. H., & Hsieh, S.H., (2010). "Using FSBT technique with Rough Set Theory for personal investment portfolio analysis," *European Journal of Operational Research*, 201(2), 601-607.
40. Simoudis, E., Han, J., & Fayyad, U., (1996), *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, Menlo Park, AAAI Press, CA.
41. Swiniarski, R., & Skowron, A., (2003). "Rough set methods in feature selection and recognition," *Pattern Recognition Letters*, 24(6), 833-849.
42. Teoh, H. J., Cheng, C.H., Chu, H.H., & Chen, J.S., (2008). "Fuzzy time series model based on probabilistic approach and rough set rule induction for empirical research in stock markets," *Data & Knowledge Engineering*, 67(1), 103-117.
43. Thomas, M. B., Diane M., Solange N., Ellen W., and Bill R., (2008). "Multiple Capacities, Multiple Outcomes: Delving Deeper into the Meaning of Community Capacity," *Journal of Rural and Community Development*, 3, 56–75.
44. Tseng, Tzu-Liang (Bill) and Huang, Chun-Che, (2007), "Rough set based approach to feature selection in customer relationship management," *The International Journal of Management Science, OMEGA journal*, 35(4), 365-383.
45. Walle, A.H. (1997). "Quantitative versus qualitative tourism research," *Annals of Tourism Research*, 24(3), 524-536.
46. Wang, K. C., Hsieh, A. T., Yeh, Y. C., & Tsai, C. W., (2004). "Who is the decision-maker: the parents or the child in group package tours?," *Tourism Management*, 25(2), 183-194.
47. Wang, X., Yang, J., Jensen, R., & Liu, X., (2006). "Rough set feature selection and rule induction for prediction of malignancy degree in brain glioma," *Computer Methods and Programs in Biomedicine*, 83(2), 147-156.
48. Witt, S. F., & Witt, C. A., (1995). "Forecasting tourism demand: A review of empirical research," *International Journal of Forecasting*, 11(3), 447–475.
49. Yao, J. T., & Herbert, J. P., (2009). "Financial time-series analysis with rough sets," *Applied Soft Computing*, 9(3), 1000-1007.
50. Zalatan, A., (1998). "Wives' involvement in tourism decision processes," *Annals of Tourism Research*, 25(4), 890-903.
51. Zhang, Z., Shi, Y., & Gao, G., (2009). "A rough set-based multiple criteria linear programming approach for the medical diagnosis and prognosis," *Expert Systems with Applications*, 36(5), 8932-8937.
52. Ziarko, W., (1994). *Rough Sets, Fuzzy Sets and Knowledge Discovery*, Workshop in Computing, Springer, London.
53. Ziarko, W., (1990). "The Discovery, Analysis, and Representation of Data Dependencies in Databases," *Knowledge Discovery in Database*, G. Piatetsky-Shapiro and W.J. Frawlwy, (eds.), Menlo Park, CA: AAAI/MIT, 213-228.