*2009 International Symposium on Computing, Communication, and Control (ISCCC 2009)*
*Proc .of CSIT vol.1 (2011) © (2011) IACSIT Press, Singapore*

# Ascertaining Data Mining Rules Using Statistical Approaches

Izwan Nizal Mohd Shaharanee [1], Tharam S. Dillon [2], and Fedja Hadzic [3 +]

[1] Digital Ecosystem and Business Intelligence Institute Curtin University of Technology
Perth 6102, Australia

**Abstract.** Knowledge acquisition techniques have been well researched in the data mining community. Such techniques, especially when used for unsupervised learning, often generate a large quantity of rules and patterns. While many rules generated are useful and interesting, some information is not captured by those rules, such as already known patterns, coincidental patterns and patterns with no significant value for the real world applications. Sustaining the interestingness of rules generated by data mining algorithm is an active and important area of data mining research. Different methods have been proposed and have been well examined for discovering interestingness in rules. These measures often only reflect the interestingness with respect to the database being observed, and as such the rules will satisfy the constrains with respect to the sample data only, but not with respect to the whole data distribution. Therefore, one can still argue the usefulness of the rules and patterns with respect to their use in practical problems. As the data mining techniques are naturally data driven, it would benefit to affirm the generated hypothesis with a statistical methodology. In our research, we investigate how to combine data mining and statistical measurement techniques to arrive at more reliable and interesting set of rules. Such a combination is greatly essential to conquer the data overload in practical problems. A real world data set is used to explore the ways in which one can measure and verify the usefulness of rules from data mining techniques using statistical analysis.

**Keywords:** data mining, significant rules, statistical analysis.

## 1. Introduction

Data mining is the process of discovering useful information, hidden patterns or rules in large quantities of data. Data mining techniques such as characterization, discrimination, association rules, classification/prediction, cluster analysis, outlier analysis, and data evolution analysis play important roles in acquiring useful information for data analysis. The hidden patterns or rules that are obtained from these data mining techniques, are considered interesting and useful if they are comprehensible, valid on tests and new data with some degree of certainty, potentially useful, actionable, and novel [1].

Whilst there are many data mining techniques available in acquiring hidden patters and rules, each of the techniques differ in terms of objectives, outcomes, and representation techniques. In response to this, [2] claims that the majority of data mining/machine learning type patterns are rule based in nature with a well defined structure, such as rules derived from decision trees and association rules. [3] agrees with this claim by indicating that the most common patterns that can be evaluated by interestingness measures include association rules, classification rules, and summaries. The techniques such as association rule mining are used for discovering interesting associations and correlations between data elements in a diverse range of applications [4]. As such they have captured a lot of interest in the data mining community.

[+] Corresponding author.
*E-mail ad*dress: (izwan.mohdshaharanee@postgrad.curtin.edu.au, {tharam.dillon, f.hadzic} @ cbs.curtin.edu.au.).

The main problems in association rule mining are discovering frequent patterns and rules construction. Frequent pattern extraction plays an important part in generating good and interesting rules, and is considered the most difficult task. Approaches such as candidate generation, frequent pattern grown (FP-grown) have been proposed to discover frequent patterns [1]. While there have been many association rule mining algorithms proposed in data mining literature and successfully used in many applications, there are still aspects of domain knowledge that these rules failed to capture [4]. Another problem is the large amount and complexity of extracted rules [5] which makes it impractical or impossible for a domain expert to analyze in an efficient manner [6]. This is why the problem of sustaining the interestingness of rules generated by data mining techniques is an active and important area of data mining research. Different methods have been proposed for discovering interesting rules from data such as support and confidence [7], collective strength [8], lift/interest [9], chi-squared test [10], correlation coefficient [11], three alternative interest measure that is; any-confidence, all confidence, and bond [12], log linear analysis [11], leverage [13, 14], and empirical bayes correlation [11].

Despite the fact that interesting association rules may be found from a database, by satisfying the various interestingness measures used, a problem that remains is that they may only reflect aspects of the database being observed. [13] emphasize that each assessment of whether a given rule satisfies certain constraint is accompanied by a risk that the rules will satisfy the constrains with respect to the sample data but not with respect to the whole data distribution. They do not reflect the "real" association rules between the underlying attributes and even if the rules are found and pass appropriated statistical test, these may be caused by purely a statistical coincidence [15]. In addition, [5, 16] agrees that the real requirement is to consider how many of the discovered rules are real rather than chance of fluctuations in the database. Therefore, we can still argue the validity of the rules and patterns to be used in practical problems. Since the nature of data mining techniques is data driven, the hypotheses generated by these algorithms must be validated by statistical methodology for them to be useful in practice [17]. On the other hand, data mining techniques are automated and are scalable and effective for finding associations between large numbers of variables, while statistical techniques can only address small number of variables. It is therefore of great importance to combine the benefits of both approaches.

In this paper, we concentrate on developing systematic ways to verify the usefulness of rules obtained from association rules mining using statistical analysis. We provide a framework that can determine the properties of rules and offer a rigorous statistical test to access the quality of rules.

In the next section, we review the basic concept of the association rules problem. The interestingness of rules and some existing interestingness measures are discussed in Section 3. In Section 4, we provide more detail about three rule interestingness measures, namely *support and confidence*, *Goodman and Kruskal's Asymmetrical Tau* and *Zhou and Dillon's Symmetrical Tau*. In this work, these will act as a basis for comparison of rule interestingness according to support and confidence framework or a statistical technique. Some statistical approaches for ascertaining association rules are presented in Section 5. In Section 6, we describe our proposed unification approach and some preliminary experimental findings are given in Section 7. Section 8 concludes the paper and describes our future work in this study.

## 2. Association Rule Mining

Association rule mining searches for interesting relationships among items in a given dataset under minimum support and confidence conditions. The problem of finding association rules $X \Rightarrow Y$ was first introduced in [7] as a data mining task of finding frequently co-occurring items in a large Boolean transaction database. An association rule $X \Rightarrow Y$ means if consumer buys the set of items X, then he/she probably also buys items Y. These items are typically called as itemsets [6].

### 2.1. Basic Concepts

Let $I = \{i_1, i_2, ..., i_m\}$ be a set of items. Each transaction $T$ is a set of items, such that $T \subseteq I$. For example, this may correspond to a set of items which a consumer may buy in a basket transaction. An association rule is a condition of the form of $X \Rightarrow Y$ where $X \subseteq I$ and $Y \subseteq I$ are two sets of items. [8] assert that the idea of the association rule is to develop a systematic method by which user can figure out how to infer the presence of some sets of items, given the presence of other items in a transaction. Such information is useful in making decisions such as shopper targeting, shelf spacing, and sales promotions.

[7] developed a two-phase large itemset approach in association rule mining problem. The first step is to find all frequent itemsets. Each of the itemsets will occur at least as frequently as predetermined minimum support threshold. The second step is to generate strong association rules from the frequent itemsets. These strong rules must satisfy the minimum support and confidence constraints. The support of a rule $X \Rightarrow Y$ is the number of transactions that contain both $X$ and $Y$, while the confidence of a rule $X \Rightarrow Y$ is the number of transactions containing $X$, that also contain $Y$.

# 3. Interestingness Of Rules

Measuring the interestingness of discovered patterns is an active and important area of data mining research. Although much work has been done in this area, so far there is no well-known agreement on a formal definition of interestingness in this context [3]. Yet, several researchers [1,3,18] agree that *conciseness, generality, reliability, peculiarity, diversity, novelty, surprisingness, utility, actionability, coverage* and *accuracy* are the eleven criteria used to determine whether or not a pattern is interesting. [1, 2] boils down interestingness into two classes: *objective* and *subjective*. An objective measurement is based on the structure of the discovered pattern and the statistics underlying them. On the contrary subjective measurement is based on user beliefs in the data. While [3, 19], add another class of interestingness which is semantic, and is a measure of the explanatory power of the pattern.

## 3.1. Interestingness Measures

Association mining is a useful technique for discovering interesting rules and patterns from large quantities of data. However, the association mining algorithms tend to generate a large volume of rules. This makes it impossible for a domain expert to easily find useful insights from the discovered rules [6]. Determining which of these patterns are useful can be very challenging.

Although there are various criteria in determining the usefulness of rules [1, 3, 18], the measures usually just reflect the usefulness of rules with respect to the specific database being observed [13]. It is hard to determine whether the rules produced are useful in practice or are valid in real world problems. Applying a data mining algorithm to practical problems may not be sufficient because we need to ensure that the results have a sound statistical basis. Even data mining algorithms founded on a sound statistical basis are not sufficient, if they cannot solve a practical problem [17]. Therefore, in our research, we aim to investigate how to combine data mining and statistical measurement techniques to arrive at more reliable and interesting set of rules. Such unification is sorely needed to conquer the data overload in practical problems [17].

# 4. Measuring Association Rules

## 4.1. Support and Confidence

As mentioned earlier, in a classical association rule mining framework, rules are considered interesting when their *support* and *confidence* exceed some user defined thresholds. [20] develop the APRIORI algorithm in two steps within the support and confidence framework in which the user sets the minimum support and confidence threshold. Once the frequent itemsets from transaction database have been found, rules are generated that are considered as interesting because they satisfy the minimum support and confidence thresholds. Yet, these rules may not be essentially interesting either from a statistical or expert's point of view [6].

## 4.2. Goodman and Kruskal's Asymmetrical Tau

Goodman and Kruskal proposed their measure of association, namely the Asymmetrical Tau, for cross-classification task in the statistical area [21]. The Asymmetrical Tau is a measure of the relative usefulness of one variable in improving the ability to predict the classifications of members of the population with respect to a second variable. The measure is calculated using a contingency table, which is the table that classifies a member of sample according two criteria. If one criterion has *m* values and the other has *n*, then an *m*n* contingency table is created. The contingency table can be used to test whether two criteria are independent.

We can predict the category of variable *B* from the category variable *A* by assuming *B* is statistically independent of *A* or assuming *B* is a function *A*. Thus the degree of association is defined as the relative improvement in predicting the *B* category obtained when the *A* category is known, as opposed to when the *A* category is not known [21]. Let;

- There be *I* rows and *J* columns in a contingency table;
- *P (ij)* denotes probability that an individual belongs to both row category *I* and column category *j*;

*P (i+)* and *P (j+)* the marginal probability in row category *I* and column category *j*, respectively.

$$Tau \quad A|B = \frac{\sum_{j=1}^{J} \sum_{i=1}^{I} \frac{P(ij)^2}{P(+j)} - \sum_{i=1}^{I} P(i+)^2}{1 - \sum_{i=1}^{I} P(i+)^2} \quad \text{and} \quad Tau \quad B|A = \frac{\sum_{i=1}^{I} \sum_{j=1}^{J} \frac{P(ij)^2}{P(i+)} - \sum_{j=1}^{J} P(+j)^2}{1 - \sum_{j=1}^{J} P(+j)^2}$$

### 4.3. Zhou and Dillon's Symmetrical Tau [21]

[21] asserted that the Asymmetrical Tau tend to favour features with more values when used as a feature selection criterion during decision tree building. They noted that when the classes of attribute *B* are increased through subdividing of existing classes, more is known about attribute *B* and the probability error in predicting the class of attribute *A* may decrease. On the other hand, attribute *B* becomes more complex which may cause the probability error in predicting its category according to the *A* category to increase. This trade-off effect has inspired them to combine the two asymmetrical measures in order to obtain a balanced feature selection criterion which is in turn symmetrical. The Symmetrical Tau [21] is defined as:

$$Tau = \frac{\sum_{j=1}^{J} \sum_{i=1}^{I} \frac{P(ij)^2}{P(+j)} + \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{P(ij)^2}{P(i+)} - \sum_{i=1}^{I} P(i+)^2 - \sum_{j=1}^{J} P(+j)^2}{2 - \sum_{i=1}^{I} P(i+)^2 - \sum_{j=1}^{J} P(+j)^2}$$

Since it has been demonstrated in [21] that using Asymmetrical Tau measure the variables with more values are favoured, in our study we will only use the Symmetrical Tau measure as it is expected to be more reliable.

## 5. Statistical Approaches

### 5.1. Statistical Significance of Association Rules

Several researchers have anticipated an assessment on pattern discovery by applying a statistical significance test before accepting the patterns. For example, in [22] correlation of rules is used, [23] proposed a pruning and summarizing approach, [24] apply a statistical test with a correction for multiple comparison by using a Bonferroni adjustment, [25] proposed an alternative approaches of encountering a rules by change and applying hypothesis testing, [4] contribute to a significant statistical quantitative rules and recently [13] summarize holdout evaluation techniques. Of all the approaches listed, [13] indicates that these approaches are only applicable to a limited range of hypothesis tests and furthermore still have a high risk of erroneously accepting spurious patterns, i.e. patterns valid with respect to sample data but not with respect to the whole data distribution. While [13] has overviewed the latest development in significance of rules discovery, some areas worth further exploration involve: issues concerning the optimal split between training and testing data, selecting a suitable statistical test and accessing quality of rules with more than one itemset in the consequent.

### 5.2. Chi-Squared Analysis

A common multivariate statistical analysis is analyzing the association problem [26]. For associations between categorical variables there are several inferential methods involved. For a two categorical variables the contingency table can be used to test the whether two criteria are independent. Chi-squared analysis is then often used to measure the difference between observed and expected frequencies. Smaller values mean that the observed value and expected frequencies are in close agreement, while larger value means that the observed and expected frequencies are somewhat apart. The significance used of the chi-squared statistics is for hypothesis testing in tests of independence. The chi-squared value is calculated as:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

where $E_i$ is the expected and $O_i$ is the observed frequency. It was proposed in 1900 by Karl Pearson. While this analysis is effective and efficient in verifying association rules, however, [22] claims that this type of association analysis has flaws when large datasets are used.

### 5.3. Logistic Regression

The logistic regression methods have become an integral component of any data analysis concerned with describing the relationship between a response variable and one or more explanatory variables. It is often the case that the outcome variable is discrete, taking on two or more possible values. The logistic regression model has become one of the standard methods of classification problems [26].

Logistics regression is used to estimate the probability that a particular outcome will occur. The dependent variable in logistic regression is the odd ratios and the outcome variable is binary or dichotomous. The coefficients are estimated using a statistical technique called maximum likelihood estimation. The interpretation of regression coefficient in terms of odd ratios is a familiar concept in analysis of categorical data [27]. Thus, it creates more flexibility for the logistic regression analyses compare to the *2 x 2* contingency table in chi-squared analysis.

### 5.4. Model Building and Fitting

The selection of logistic regression model involves two competing goals: the model should be complex enough to fit the data well, while at the same time simpler models are preferred since they are easier to interpret [26]. Model building principally involves seeking and determining parsimonious (simple) model that explains the data. The rationale for preferring simple models is that they tend to be numerically more stable, and they are easier to interpret and generalize [27, 28]. There are different strategies for model selections. The first approach is to consider all possible regressions. In this approach, only certain key summary statistics such as residual sum of squares are used to narrow the list of models. Second approach is to use certain model selection algorithm that can be employed by computer. There are three automatic selection procedures such as *forward selections*, *backward selection*, and *stepwise selections*.

One should always be cautious in using the "automatic" algorithms for model selection. In cases, when there is a high degree of multi co-linearity among the independent variable, the three methods may lead to a different final model. It is more preferable to examine the all possible regression because such analysis can produce several models perform quite similarly [27]. The selection of the "best" model from all possible models depends on the criterion that is chosen. In the experiment of this paper, we use the lowest values of *deviance* and *Pearson chi-square* as the selection criteria.

## 6. Proposed Unification Approach

### 6.1. A Conceptual Framework

Based on Figure 1, a set of data mining rules obtained from the association rule mining will reflect the database and the real world data (implicitly). Such rules are considered interesting if, the rules satisfy some user specific threshold namely support and confidence. However, we never know whether the rules truly represent/reflect the real data, since the mining process occurs at the database level.

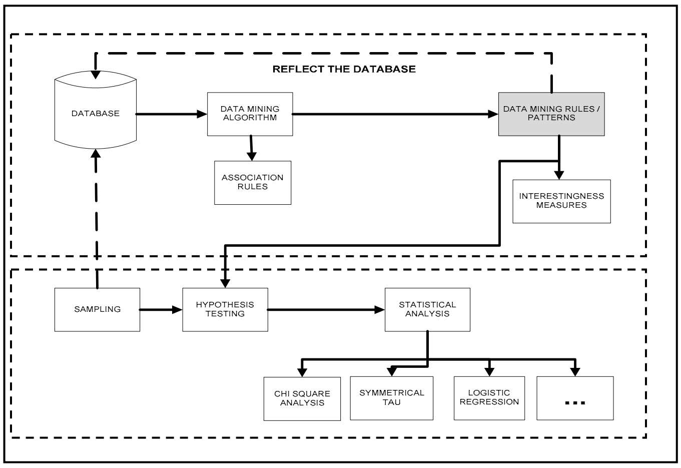

Fig. 1: Framework for analyzing rule interestingness in Association Rules for large database.

The similar data that is used for the rules generation in data mining process, further needs to be verified by the statistical analysis approaches. Such approach, requires sampling process, hypothesis development, model building and finally a measurement using statistical analysis techniques to verify and prove the usefulness and quality of rules discovered using association rules mining.

Hypothesis development will be based on the interesting rules obtained from association rules mining. An appropriated sample set will be drawn from the database or real world data. A specific statistical analysis will then be utilized. Firstly, the chi-squared analysis is used to discover the properties of data attributes; principally on the data dependency. Secondly, the Symmetrical Tau is utilized to provide the relative usefulness of attributes. The logistic regression analysis is then employed to provide the classification power of the data. The development of logistic regression modelling involves the model building strategies. The purpose of model building strategies is to select variables that results in a “best” and most parsimonious model that still explains the data [28]. These statistical analysis results are used to determine and verify the applicability of the association rules to the real world data. This information will facilitate the association rule mining framework to determine the right and high quality rules that represent the real data.

# 7. Preliminary Results and Discussions

Some preliminary evaluation of the framework towards unification of data mining and statistical analysis has been performed using a real world dataset. The datasets used correspond to the hand washing pattern in bathrooms at the University Of IOWA [27]. The data consists of whether users of the bathroom washed their hands (0, no; 1, yes), gender (0, male; 1, female), whether users carried a backpack (0, no; 1, yes), and whether others were present (0, no; 1, yes).

## 7.1. Association Rules Results (Min Support = 10% and Min Confidence = 60%).

There are 50 rules discovered using the association rule mining method with the minimum support of 10% and confidence value of 60%. Next, the usefulness of these rules needs to be verified by statistical analysis approaches.

## 7.2. Statistical Analysis Approaches

Statistical analysis approaches, namely chi-squared, Symmetrical Tau and logistic regression were employed in order to determine the usefulness of rules obtained from the association rule mining method. The results from chi-squared analysis are discussed first.

Of all two categorical variables combination, we found that there are three significance associations between variables, namely “wash” and “gender” $x^2(1)=12.79, p<0.05$, “wash” and “backpack” $x^2(1)=4.44, p<0.05$, and “backpack” and “other” $x^2(1)=1.67, p<0.05$. For example the rule *Female=>WashYes* can be verified by examine the significance value of chi-squared analysis between variable “wash” and “gender”. This is a positive correlation, which suggests that, women are washing hand more frequently.

Results in Table 1 show that, “gender” and “backpack” variables have better predictive capability for the values of the wash variable. In predicting the values of the other variable, we found only “backpack” as the useful variable. The “other” and “wash” variables are relatively useful variables in improving the ability to predict the “backpack” variable, and finally “wash” is the significant predictor for “gender”.

The results show that the significance of certain rules can be ascertained by checking that the Symmetrical Tau value is sufficiently high between the attributes contained in the precedent of the rule for predicting the values of the attribute contained in the consequent. For example the significance of the rule *Male&NoBackpack=>WashYes* can be ascertained by checking that the Symmetrical Tau value for the attributes “gender” and “backpack” in predicting the value of attribute “wash” is sufficiently high, i.e. 0.085 and 0.03, respectively (Table 1).

A logistic regression approach with proper model building techniques produced a good model in classification problems [28]. For each model developed, we first, test the variable homogeneity, and then we determine the odd ratio to explore the form of the relationship. We also determine the estimation and goodness-of-fit results for the selected model. The goodness-of-fit statistics in this model are based on the value of deviance residual and Pearson residuals. For each target variables, the results are shown in Table 2. For example, the rule *Alone&NoBackpack&Female => WashYes* is equivalent to a model with target variable “wash”; we start our model with all three covariates. We find that “other” is not statistically significant contributor to the model, which leads us to discard the above rule. The estimation and goodness-of-fit results for this model are satisfactory, i.e. The Pearson and deviance residual do not exceed the 95th percentile.

Table 1: Symmetrical Tau Measure For Variables

| **Predicted Variables** | **Predictors** | | |
|---|---|---|---|
| Wash | Gender | Backpack | Other |
| | 0.085 | 0.030 | 0.0004 |
| Other | Backpack | Gender | Wash |
| | 0.0313 | 0.009 | 0.0004 |
| Backpack | Other | Wash | Gender |
| | 0.0313 | 0.030 | 0.011 |
| Gender | Wash | Backpack | Other |
| | 0.085 | 0.011 | 0.009 |

Table 2: Logistic Regression Approach

| **Target Variables** | **Covariates** | | |
|---|---|---|---|
| Wash | Gender | Backpack | Other |
| | Significant | Significant | Not Significant |
| Other | Backpack | Gender | Wash |
| | Significant | Not Significant | Not Significant |
| Backpack | Wash | Other | Gender |
| | Significant | Significant | Not Significant |
| Gender | Wash | Backpack | Other |
| | Significant | Not Significant | Not Significant |

## 8. Conclusion and Future Works

This was our preliminary work towards the combination of data mining and statistical techniques in ascertaining the extracted rules/patterns. The combination of the approaches used in this method showed a number of ways for ascertaining the significant patterns obtained using association rule mining approaches. In this paper we employed hypothesis testing using statistical analysis that provide some control in lowering the risk of discovering a pattern that is false and spurious.

In the currently proposed framework, we focused mainly on association rules with only one variable in the consequent and a small sized dataset was used for preliminary evaluation purposes. However, to fully address the problem of discovering significant/interesting rules, we must considered several factors such as the multiple items in the consequent part of the rules, types of variable involved, and selection of appropriated statistical test.

Thus, for future work we aim to test the approach using much larger datasets to demonstrate the benefits of sampling cases from independent datasets to ascertain the validity of the discovered rules. We will further verify the pattern by applying a log linear analysis; as such technique provides facilities in determining the association between multiple itemsets in a rule.

## 9. References

[1] J. Han, and M. Kamber, Data mining: concepts and techniques, Morgan Kaufmann Publishers, 2001.

[2] K. McGarry, “A survey of interestingness measures for knowledge discovery,” Knowl. Eng. Rev., vol. 20, no. 1,

2005, pp. 39-61.

[3] L. Geng, and H.J. Hamilton, “Interestingness measures for data mining: A survey,” ACM Comput. Surv., vol. 38, no. 3, 2006, pp. 9.

[4] H. Zhang, B. Padmanabhan, and A. Tuzhilin,, “On the discovery of significant statistical quantitative rules,” In Proceeding of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle,WA,USA,2004, pp. 374-383.

[5] D.J. Hand, “Data Mining: Statistics and More?,” The American Statistician, vol. 52, no. No. 2, 1998.

[6] L. Philippe, M. Patrick, V. Benoît, and L. Stéphane, “On selecting interestingness measures for association rules: User oriented description and multiple criteria decision aid,” European Journal of Operational Research, vol. 184, no. 2, 2008, pp. 610-626.

[7] R. Agrawal, T. Imieliski, and A. Swami, “Mining association rules between sets of items in large databases,” SIGMOD Rec., vol. 22, no. 2, 1993, pp. 207-216.

[8] C.C. Aggarwal, and P.S. Yu, “A new framework for itemset generation,” Book A new framework for itemset generation, Series A new framework for itemset generation, ed., Editor ed.^eds., ACM, 1998.

[9] S. Brin, R. Motwani, J.D. Ullman, and S. Tsur, “Dynamic itemset counting and implication rules for market basket data,” ACM SIGMOD Record, Volume 26, Issues 2, 1997, pp. 255-264.

[10] C. Silverstein, S. Brin, and R. Motwani, “Beyond Market Baskets: Generalizing Association Rules to Dependence Rules,” Data Min. Knowl. Discov., vol. 2, no. 1, 1998, pp. 39-68.

[11] T. Brijs, K. Vanhoof, and G. Wets, “Defining interestingness for association rules,” International journal of information theories and applications, vol. 10(4), 2003, pp. 370-376.

[12] E.R. Omiecinski, “Alternative interest measures for mining associations in databases,” Knowledge and Data Engineering, IEEE Transactions on, vol. 15, no. 1, 2003, pp. 57-69.

[13] G.I. Webb, “Discovering Significant Patterns,” Machine Learning, Springer, no. 68, 2007, pp. 1-33.

[14] G. Piatetsky-Shapiro, “Discovery, analysis; and presentation of strong rules,” Knowledge discovery in database, 1991, pp. 229.

[15] Y. Aumann, and Y. Lindell, “A Statistical Theory for Quantitative Association Rules,” J. Intell. Inf. Syst., vol. 20, no. 3, 2003, pp. 255-283.

[16] D.J. Hand, “Statistics and data mining: intersecting disciplines,” SIGKDD Explor. Newsl., vol. 1, no. 1, 1999, pp. 16-19.

[17] A. Goodman, C. Kamath, and V. Kumar, “Data Analysis in the 21st Century,” Stat. Anal. Data Min., vol. 1, no. 1, 2008, pp. 1-3.

[18] L. Nada, F. Peter, and Z. Blaz, “Rule Evaluation Measures: A Unifying View,” Inductive Logic Programming, 1999, pp. 174-185.

[19] Q. Mei, D. Xin, H. Cheng, J. Han, and C. Zhai, “Semantic annotation of frequent patterns,” ACM Trans. Knowl. Discov. Data, vol. 1, no. 3, 2007, pp. 11.

[20] R. Agrawal and R. Srikant,, “Fast Algorithms for Mining Association Rules in Large Databases,” Proceeding of the 20th International Conference on Very Large Data Bases, San Francisco, CA, USA, 1994, pp. 487-499.

[21] X. Zhou and T.S. Dillon, 1991. “A statistical-heuristic feature selection criterion for decision tree induction”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no.8, August, pp 834-841.

[22] S. Brin, R. Motwani, and S. Craig, “Beyond market baskets: generalizing association rules to correlations,” Proceeding of the 1997 ACM SIGMOD International Conference on Management of Data, Tucson, Arizona, USA, 1997, pp. 265 - 276.

[23] B. Liu, W. Hsu, and Y. Ma, “Pruning and summarizing the discovered associations,” In Proceeding of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, California, USA, 1999, pp. 125 - 134.

[24] S.D. Bay, and M.J. Pazzani, “Detecting Group Differences: Mining Contrast Sets,” Data Mining and Knowledge Discovery, vol. 5, no. 3, 2001, pp. 213-246.

[25] N. Meggido, and R. Srikant, “Discovering Predictive Assocaition Rules,” 4th International Conference on Knowledge iscovery in Databases and Data Mining, pp. 274-278.

[26] A. Agresti, An introduction to categorical data analysis, 2nd edition, Wiley-Interscience, 2007.

[27] A. Bovas, and L. Johannes, Introduction to regression modeling, Brooks/Cole, 2006.

[28] D.W. Hosmer, and S. Lemeshow, Applied logistic regression, Wiley, 1989.