Journal of Theoretical and Applied Information Technology
10th January 2014. Vol. 59 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

SELECTION AND AGGREGATION OF INTERESTINGNESS MEASURES: A REVIEW

[1,2]KOK KEONG BONG, [2]MATTHIAS JOEST, [3]CHRISTOPH QUIX
[1,4]TONI ANWAR, [5]SELVAKUMAR MANICKAM
[1]Thai German Graduate School of Engineering, Thailand
[2]Heidelberg Mobil International GmbH, Germany
[3]RWTH Aachen and Fraunhofer FIT, St. Augustin, Germany
[4]Universiti Teknologi Malaysia (UTM), Malaysia
[5]National Advanced IPv6 Center of Excellence (NAv6), Malaysia
E-mail: [1]bong.kokkeong@heidelberg-mobil.com

ABSTRACT

Association Rule Mining is the process of retrieving frequent patterns that occur in a transaction database. Initially used as a market basket analysis solution for retail businesses, it has grown to cover many other fields such as medicine [1, 2], traffic estimation [3] and anomaly detection [4, 5]. An association rule has two components (antecedent and consequent) which is derived from a pattern (a set of items). However, it is not clear when investigating a frequent item set, which items imply the others (i.e., which is antecedent, and which is consequent). Therefore, several combinations of items as antecedent and consequent are generated. This leads to a huge amount of association rules being output by an algorithm for Association Rule Mining. Thus, it is imperative that data miners require some type of measures to evaluate the "interestingness" of these rules. There exist in excess of 70 well-known measures and countless other manually crafted measures in the literature. In this survey, we systematically discuss the methods which users could use to select or aggregate the interestingness measures, applicability of such methods and evaluation of the usage of such methods.

Keywords: *Association Rule Mining, Objective Interestingness Measures, Data Mining, Clustering, Information Retrieval.*

1. INTRODUCTION

One of the many pillars of data mining is Association Rule Mining, which is the problem where given a database of items and transactions (that grouped different items together), the goal is to find association rules that associate the items in the database. The common example of such database is the supermarket database and the goal is to find items that are associated among each other. Typical association rule mining algorithms produce an exhaustive list of rules in the form of "Antecedent → Consequent" that exceed a predefined popularity threshold. For example, the association rule "Egg ^ Milk → Bread" consists of the antecedent of "Egg ^ Milk" and the consequent of "Bread". The general interpretation of this rule is "customers who had bought Egg and Milk also bought Bread". The aforementioned rule comes from the itemset of {Egg, Milk, Bread} that occurs frequently in the transaction database. The same itemset could evaluate to other rules like "Milk ^ Bread → Egg" and "Egg ^ Bread → Milk". If the threshold is set too high, then there would be very little to none rules that are useful to the data miner. Worse still, popular rules can also be a common-sense and nothing about it is interesting or helpful. Setting the threshold too low would yield too many noise in the result (when every single association is uncovered by the algorithm). Data miners often set these thresholds at the lower end to avoid missing out on important rules and utilize measures to further reduce the amount of resulting rules. Approximately 70 measures have been formulated and these measures target different types of rules (e.g., popular or surprising-type). With the appropriate interestingness measure (IM), users can efficiently remove rules that are not worthy for them (i.e., not interesting).

1.1 Interestingness as a Measure to Reduce Rule Overload

The inherent problem of association rule mining is that the algorithms generate huge amounts of association rules as output. This is because each rule is a permutation of the items in a frequent itemset and there might be a huge number

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

of frequent itemsets existing in a transaction table (depending on the user-defined support threshold).

In order to further filter the amount of rules that are produced by the algorithm, users could opt to use more complex formulas or heuristics to determine whether a rule is interesting or not. More interesting rules can be ranked on top of other not-so-interesting rules before the association rules are presented to the users. Filtering and ranking of association rules depend largely on the type of association rules a user would like to find. The rules can be grouped into several types according to the concept of interestingness [6]: concise, general, reliable, peculiar, novel, surprising, diverse, useful and actionable. Any given association rule can belong to several types. For example, a general rule is often also a concise rule. A peculiar association rule can also be a novel rule.

1.2 Objective vs. Subjective IMs

There are two general categories of interestingness measures: Objective Interestingness Measures (OIM) and Subjective Interestingness Measures (SIM). In [6], the authors defined an additional type of interestingness measure: Semantic Interestingness Measures. The survey [6, 7] defined OIMs as the measures that denote the statistical strengths or certain properties of the discovered patterns and it is based solely on the raw data. For example, the measure Support is used to denote the popularity of the association rule and this definition remains the same across different domains. Furthermore, it is a form of statistical strength of the association rule and so, it is grouped under OIMs.

Some users prefer to find association rules that are surprising or actionable to them. Surprising association rules are those that deviate from the user's belief. This sort of deviation calculation would yield a measure that is categorized as a SIM because it is based on the concept of comparing association rules with a user expectation [7]. Furthermore, the users belief vary from one user to another. However, there has been some research that attempts to find surprising rules through the use of OIMs. One research relies on the assumption that the user's belief is the association rules with high support and confidence values (called common-sense rules) [8]. Thus, the separation between OIMs and SIMs are not always as clear-cut as by determining the type of rules that a user would like to find (e.g., the view that discovering surprising rules can be done only through SIMs).

Our survey focuses on the state of art of practical utilization of OIMs. Our discussion on latest methods in the selection of OIMs is presented in Section 3.1. Meanwhile, the Section 3.2 can be loosely categorized as Semantic Interestingness Measures. Although methods for aggregation include the optimization of the OIM such that it increases classification accuracy [9] or closeness to user manual ranking [10, 11] (categorized as Semantic IMs), there are also aggregation methods that use input from the OIM properties [12] which do not require input from users (hence, categorized as Objective IMs).

1.3 Contributions and Outline of the Survey

This survey is about the recent development in utilizing OIMs available in literature and directions that a user can follow to maximize the usage of these predefined OIMs. Thus, we would not be explaining one by one the many measures that are readily available. Readers could refer to [6, 7, 13, 14] for such listing.

Our work differs from existing survey [7] in the sense that we focused heavily on the utilization of existing OIMs (selection and aggregation). Although the survey from Geng and Hamilton [6] included a small section on OIM selection, we extend here their discussion on the selection by highlighting improvements over the selection methods. In addition, we have included a new discussion of the state of art in aggregating OIMs which is not found in existing surveys. Thus, we discuss critically the progress in the selection of the appropriate measures and the recent area of combining interestingness measures.

We will first give a basic definition of association rules and some basic measures (support and confidence) in Section 2. Section 3 highlights the methods used by data miners to apply the OIMs on their dataset while Section 4 discusses the methods to quantify the benefits of the OIM usage and potential challenges. This is an important discussion as a user moves away from derivation of custom measures towards selection and aggregation of measures. Finally, we conclude the survey in Section 5.

Journal of Theoretical and Applied Information Technology
10th January 2014. Vol. 59 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

2. DEFINITION

The transaction table is a table with features (as columns) and observations (as rows). The values in the transaction table are constrained to have only boolean values, e.g., 0 (if the feature is not applicable to the observation) or 1 otherwise. In the context of the previous example, the observations are the customers' purchases while the features are the items purchased by the customer in the supermarket (e.g., milk, bread) within one transaction. The goal of association rule mining is to mine for rules that associate the items in the transaction database. For a transaction table with size k X m (k is the number of rows, m is the number of columns), we will use the notation Fm to identify each feature (column) and T_k to denote each observation. Conveniently, a value of 1 at T_{km} would denote that a customer bought the item Fm in the transaction T_k.

Each association rule is made of a set of items (hereon referred to as an itemset) and to get to the rules, we need to collect the itemsets. One of the basic algorithm for association rule mining is the Apriori algorithm [15]. It starts by collecting single-item itemsets that exceed a certain frequency threshold of transactions. In the supermarket example, these are the items that were bought frequently. This frequency threshold is called the Support value. It is the ratio between the number of observations that contain the itemset (|A|) and the total observations (N) in the database.

$$Support = \frac{|A|}{N} \quad (1)$$

From the single-item itemset, the algorithm expands to two-item itemsets and so forth until the current length itemset cannot be expanded without having the support value fall below the threshold. This search and expand is feasible because of the anti-monotonicity property of the support value of the itemsets. This property dictates that the support of an itemset is less than or equal to the support values of any of its subsets. Thus, an itemset may be frequent only if its subsets are frequent. The set of itemsets that exceeded the frequency threshold is called the frequent itemsets. In reverse, the transactions that contain an itemset are called transactions that "cover" the itemset.

Upon retrieving the frequent itemsets from the transactional database, association rules are generated by permutating the items in the itemsets. From each itemset, a subset will go into the antecedent set while the others, into the consequent set. We refer to those sets as sub-itemsets of an itemset. In this paper, the association rules will be represensented by the form "A → B", with the sub-itemset A denoting the antecedent set and the sub-itemset B denoting the consequent set. This would generate a huge combination of rules and to control the outcome size of the algorithm, another measure is used. This measure is called the Confidence (Equation 2). The confidence is interpreted as the probability that "A → B" occurs if A has already occurred.

$$Confidence = \frac{Support(AB)}{Support(A)} \quad (2)$$

From a mathematical perspective, each association rule is represented as a contingency table (cf. Table 1). The contingency table contains the cardinalities of each sub-itemset and combination sub-itemsets in the association rule. In other words, it shows how many transactions contain both the antecedent and consequent component (cell N_{ab}), how many transactions contain the antecedent (cell N_a), how many transactions contain the antecedent but not the consequent (cell $N_{a\bar{b}}$) and how many transactions that do not contain both antecedent and consequent (cell $N_{\bar{a}\bar{b}}$).

Table 1:The format of a contingency table (subscript "a" being the antecedent while "b" is the consequent)

	Cons, b	Cons, $\bar{b}$	
Ante, a	N_{ab}	$N_{a\bar{b}}$	N_a
Ante, $\bar{a}$	$N_{\bar{a}b}$	$N_{\bar{a}\bar{b}}$	$N_{\bar{a}}$
	N_b	$N_{\bar{b}}$	Total, N

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 2: Methods toward selecting the appropriate OIM.

Selection Style	Definition and Qualifying Step	Analysis Step	Implementation
Custom	Dataset properties	If the OIM produces values consistent with the defined properties then it is selected.	[2, 31]
	OIM Mathematical Properties	User chooses which to use based on context.	[22, 28]
Decision aids	OIM Mathematical Properties	Decision aids	[21, 32]
	User labeled rules	Learning algorithms to predict selection	[10]
Empirical	OIM output values	Clustering	[26, 34, 35, 36 , 37, 38, 39]
	OIM output values	Classification accuracy	[3]
	User ranked sampled rules	Ranking correlations	[17]

Various formulas as surveyed and studied in [2, 7, 6, 16, 17, 18, 19, 20] combine the contingency table cell values to produce a final value that quantifies the interestingness. Two basic interestingness formulas are the Support and Confidence (as introduced above). In the context of the contingency table, the formula for Support is the ratio between the contingency table cell N_{ab} and the cell Grand-Total. The value of Confidence is obtained by dividing the contingency table cell N_{ab} by cell N_a.

3. UTILIZING OIM

Mechanisms are required to filter out possibly irrelevant rules in the face of huge amounts of data and huge amounts of association rules mined. The survey in [6] noted that the roles of OIMs in association rule mining consists of:

1. pruning search space so as to enable efficient rule mining [28, 29, 30]
2. rule ranking, and
3. post-processing to uncover only interesting rules.

Despite the roles, the primary methods employed for using OIMs are the selection of OIM and the aggregation of OIMs. They will be discussed in turn in the subsequent sub sections.

3.1 Selection of OIMs

Users are left to decide on which OIMs to use based on their experience and assumption about the association rules that they would like to find. The general process for selecting an appropriate interesting measure is broken down into three steps:

1. Defining required properties and its weight (importance). The properties could either be the properties of the dataset, properties of the interestingness measure itself (cf. Table 3, 4, 5) or even the values produced by the OIM in test dataset.
2. Quantifying the similarity and differences between properties and OIMs. OIMs that have the same properties are considered redundant (using one is approximately the same as using all) and can be removed.
3. Analyze/Tradeoff OIMs with those properties. OIM that fits the tradeoff or preference is selected.

Based on the three-step process above, we can group the OIM selection methods into three groups: custom, decision aids, and empirical methods based on the data they required and automation level. Custom and decision aid approaches require data that is not found in the database. This information includes user's goal and mathematical properties of the OIM. These approaches also require significant investment in terms of user efforts. On the other hand, empirical methods are defined as requiring only the database and the interestingness values produced by the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

OIM. It also requires the least amount of user intervention among the three approaches to OIM selection. Table 2 summarizes the state-of-the-art according to the three selection styles and three-step process. The three methods will be discussed in the next subsections and analyzed at the end of this section.

3.1.1 Custom inspection approach to selection

The most straightforward but resource intensive method is by manually inspecting the OIM properties. In selecting the best measure to evaluate correlations between Chinese medicine syndrome elements and symptoms, Zhang [2] defined a set of three custom properties (based on dataset) that they desire the OIM to account for when calculating the interestingness of the association rules. The properties are based on their application domain and are listed below:

- If two syndrome elements are completely different, then the difference between them should be large.
- If two syndrome elements are the same then the distance between them are small.
- If two syndrome elements are the same then the distance between them across different (but the same application domain) datasets should be similar.

In other words, the author in [2] tried to choose an interestingness measure that fits the property in his dataset. He ran calculations using 60 OIMs to see which OIM satisfy their desired properties. This arrangement enabled them to trim down the choices from 60 OIMs to only three OIMs. Then the final OIM chosen was the one with the lowest computational complexity.

In a separate research, Wu et al. [14] used the manual selection method to reduce their option of 35 OIMs to 13 OIMs. Their defined properties are:

- Uniqueness: An interestingness measure assigns one score to an itemset.
- Extensibility: An interestingness measure must be able to provide scores for multi-itemsets. Multi-itemsets are itemsets that have more than one component in the antecedent and the consequent.
- Antimonotonicity: An interestingness measure must be able to provide scores in such a way that a pattern is considered interesting only if all its sub-patterns are interesting.

As in the work of Vaillant et al. [31], the authors proposed choosing an OIM based on the property of how an OIM value changes with additional counter-examples in the association rule (sensitivity against new counter examples). In a market basket analysis example, the counter example for the association rule "Cereals → Milk" would be rules that has the same antecedent but differing consequent (like "Cereals → Bread"). Recall that in a contingency table context, the amount of counter example that an association rule has is shown in the cell $N_{a\bar{b}}$. In this case, the authors defined interestingness as a decreasing function of counter-example additions and used first and second derivatives to model the sensitivity (rate of change against addition of counter examples) of ten OIMs. They selected an OIM that assigned less interestingness to association rules that have a large number of counter examples.

Wu, Chen and Han [28] suggested that the null invariance property is of high importance for large datasets. The null invariance property of an association rule means that its interestingness score is not affected by association rules that do not share any components at all. For example, the interestingness of the association rule "Beer → Diapers" should not be affected by association rules like "Apple → Orange" (which shares no similar components). In relation to the contingency table sense, the interestingness value should not be affected by the values in the cell N_{ab}. This property is important in their application context because they were looking for rare association rules (interesting rules but occurring rarely in the dataset).

As a summary of this section, the custom inspection methodology mentioned suffers from the lack of formalism and this deters it from being used widely across different application areas. All properties need to be defined, evaluated and compared for tradeoffs by the user.

3.1.2 Decision aids approach to selection

Since there exists several defined properties and users are required to find one OIM that satisfies the most optimal amount of properties (with respect to certain weights of importance for each property), the problem of selecting an OIM can be framed as an optimization problem. In the field of optimization, users can look towards decision aids

Journal of Theoretical and Applied Information Technology
10th January 2014. Vol. 59 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

as a method to formalize the properties of OIMs and systematically choose the appropriate/optimal OIM.

The difference between decision aids and manual inspection is the assistance provided by the system to make it easier to select an OIM that is optimal according to the user-required properties. The first discussion in this section deals with the visualizations while the next discusses research focused on providing checks on user requirements (making sure that it is consistent). User requirement consistency checking is important because some properties contradict with each other. Assigning the same importance weight to two contradicting properties would yield a sub-optimal selection result.

The first work discussed here is by Lenca et al. [21], which defined eight properties (refer Table 3) for assessing and selecting the appropriate OIM using Multi Criteria Decision Aid (MCDA). The properties are then populated with values for each OIM on whether it satisfies the properties. Users can also assign weights to the properties to indicate the importance of such properties to be fulfilled by OIMs. Upon collection of such information, the decision aid then performs the following steps:

1. Pair wise comparison of all OIMs according to each defined properties. Each comparison results in a preference of one OIM (with respect to one property) over the other OIM. The preference has values between 0 (not-preferred) and 1 (preferred).
2. Aggregate the preference value of each property for each pair of OIM into a single preference value for each pair of OIM.
3. The aggregated values are used to build outranking flows for each OIM.
4. The outranking flows are then used to build a partial and complete ranking of the OIMs.
5. The ranking is visualized on a GAIA plane.

Figure 1 shows a GAIA plane. It is used to determine how one OIM is related to another OIM with regard to the properties defined. In that figure, the square-plots denote the properties while the triangular-plots denote the OIMs being investigated. The lines from the origin to the square-plots indicate directionality of the properties. The length of those lines indicates whether the property is a discriminating factor (it separates clearly between those OIMs which fulfill and do not fulfill the property).

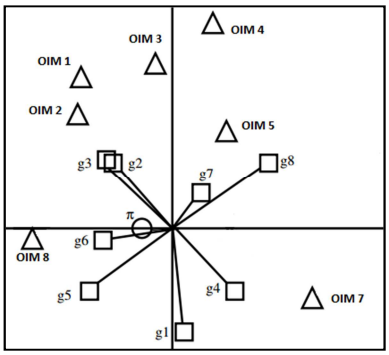

Figure 1: Decision aid visualization using the GAIA plane [21].

If two square-plots/properties are plotted in the same direction from the origin, then the properties are considered similar. This can be seen from the g2 and g3 properties in the top-left quadrant of the figure. Similarly, if two square-plots/properties are plotted such that they have opposite directions, then the properties are negatively correlated (example, g5 - bottom left quadrant and g8 - upper right quadrant). However, if two squareplots properties are plotted orthogonally, they are considered as independant to each other. An example of independant properties would be between g2 (in top left quadrant) and g8 (top right quadrant).

In relation to the triangular-plots/OIMs, OIMs that are plotted in the same direction as the square-plot are considered to fulfill that property. From the figure 1, we can observe that OIM 1 and OIM 2 have the properties g2 and g3. This is the opposite of the OIM 7 which has property g4. After the properties and OIMs relationship are visualized to show similarity, independence or complementary, the user can choose the appropriate OIM. This is done by observing the π-plot. The longer the π-plot (in this case it is on the x-axis) indicates a strong decision power and the OIMs lying close to direction indicated by π-plot would provide the best optimal OIM based on the properties. If the π-plot is short, then the most optimal OIM would lie very close to the origin of the chart. In Figure 1, the most optimal OIMs can be found along the negative x-axis, which is OIM 8.

The main drawback of such decision aid is the required resources to operate them. Despite the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

visualization capability, this method requires an MCDA expert to explain how the method works and the meaning of the results from the decision aid, plus a domain expert capable of knowing which criteria to put more importance on. Without a careful expert, users may assign equally important weights to conflicting properties. This could result in suboptimal selections of the OIM.

A more recent development is the work of [32] which addressed the danger of conflicting weight assignment by users. This is based on a decision aid method named Analytic Hierarchy Process (AHP). The authors proposed that a good selection process requires:

- accounting for user requirements.
- avoiding inconsistencies in user decision.
- adaptable to changing user requirements.
- invariant to number of rules.

Thus, they focused on eliminating user inconsistencies in weight setting to improve the selection process. In the AHP system, they combined the properties from Table 3 with the properties listed in Table 4 as their criteria for selecting an OIM. The AHP method follows the same process as the decision aid used by [21]. However, there exists several differences between the work of [21] and the AHP:

1. The OIM can only take a boolean value for each property (1 for fulfilling the property, 0 otherwise).
2. For property comparison: AHP only does chain-wise comparison of properties while [21] performs a full pair wise comparison.
3. Consistency of user weights are being checked (and adjusted if inconsistent) in AHP but other decision aid allows arbitrary weights.
4. There are no visualization involved in AHP, the most optimal OIM is chosen with the highest weighted score.

In the AHP, individual weights $(W_1, W_2, \dots W_n)$ are assigned for a property list of $P_1, P_2, \dots P_n$ by the user. A ratio R_i is defined as the ratio between the weight of the current property (W_i) and the weight of the next property (W_{i+1}). The last ratio for the last property (R_n) is taken as the ratio between the last property W_n and the first property, W_1. A perfect consistency is achieved if the product of all ratio (R) equals to 1. The authors noted that a consistency of up to 90% is acceptable. Otherwise, the user is notified of the inconsistency or an adjusted weight is recalculated for the properties.

To bridge the gap between domain expert requirements and OIM selection, researchers tried meta-learning methods to model the domain expert's knowledge and predict whether a rule with its values of OIMs will be deemed interesting by the domain user. Abe et al. [10] proposed a rule evaluation support system that learns to predict the OIM that a domain expert would select. In addition to calculating the OIM values for all mined rules, the system required a domain expert to label the mined association rules with a three-point preference scheme: Interesting, Not Interesting, Not-Understandable. The same study [10] applied five different learning algorithms from the Weka tool: C4.5 decision tree, neural network, SVM, classification through linear regressions and One-R. They reported a learning curve prediction accuracy of over 80% after using more than 20% training samples.

3.1.3 Empirical approaches to selection

The empirical approaches of brute force and clustering eliminate most of the requirement for a domain expert participation in OIM selection. In proposing the selection of an appropriate OIM for mining classification rules in traffic prediction application, the authors in [3] used ten different OIMs to predict their traffic dataset and proposed the best one that had the highest accuracy. So, the one with the highest accuracy on the previous day is then chosen to be the OIM used to predict traffic on the current day.

In selecting the right interestingness measures, Tan et al. [17] proposed a method that selects the interestingness measure based on relative rankings provided by an expert user. The user is presented with a sample of twenty rules to manually rank based on user-perceived interestingness. Upon receiving the feedback, the system then calculates the correlation between the user-rank and the ranking produced by the OIMs. The system then selects the OIM that has the least amount of ranking conflict with the user-rank.

Besides brute force and user feedback based empirical selection, users can also perform clustering as a method to choose the appropriate OIM. The basic idea in using the clustering approach to find the appropriate OIM is that OIMs that are clustered together have the same attributes.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The attributes could be that they fulfill similar type of properties or that they produce similar results for the same association rule. The set of similar OIMs are considered as redundant. On the other hand, dissimilarity between sets of OIMs denote that they are complementary of each other (sets) and using one of it means trading off the properties of other clusters (sets). After redundant OIMs are removed from consideration, users can choose a representative from the cluster of OIM that fits their goals. The clustering-themed OIM selection approach can be split into two different directions [6]: property-based clustering and experimental-based clustering.

As the name implies, property-based clustering is the approach where users defined a set of properties, then fills up the OIM-Property matrix with the appropriate values and finally, clusters the OIMs according to such matrix. Grissa et al. [36] performed studies of grouping OIMs using clustering methods of K-means and AHC. The properties that they used to evaluate 61 OIMs are the combination of those from Table 3, 4 and 5. After the 61 OIMs are clustered, the clusters are verified for their validity. The validity is manually checked using a visualization tool.

A cluster is considered valid if the cluster members fulfilled many of the same properties that OIMs in other clusters do not. So, users would need to navigate around the cluster members to check the closeness of the cluster members in the visualization tool. Although the paper [36] focused more on verifying clusters obtained by clustering 61 OIMs against 19 properties, it also highlights the option of performing K-means and Agglomerative Hierarchical Clustering (AHC) on the OIM-property matrix. Hence it is included in this survey. Users following such approach could adjust clustering parameters to obtain “more valid” clusters and then choose the appropriate OIM from the valid clusters.

A recent development in property-based clustering involves the use of Boolean Factor Analysis (BFA) to decompose the IM-Property matrix into important factors (properties) was proposed in [39]. This method is similar to principle component analysis but is performed on binary (boolean) data instead. 21 properties are reused from Table 3, 4 and 5. The 21 properties are then supplemented with their negations to form a list of 41 properties. All the properties can take a value of 1 (fulfill property) or 0 otherwise. 62 OIMs were evaluated using the 41 properties mentioned in the preceding paragraph. This resulted in an OIM-property matrix. To reduce the amount of properties to be evaluated, the authors performed the boolean factor analysis to reduce the amount of properties that needs to be used. This is akin to feature selection where we select only features (in this case, the properties) that account for the most variance in the OIM-Property matrix. The authors managed to reduce the property space to only 28 factors that covers more than 95% of the variance in the matrix. Note that a factor could contain more than one property.

With such composition, the authors elaborated that each factor can be considered as a cluster. In addition, because an OIM can reside in more than one factor, this means that the factors deduced can form overlapping clusters. This is the advantage of their method because the authors claim that OIMs should not be clustered into mutually exclusive clusters (where each member belongs only to one cluster). Their work is in contrast to empirical approaches towards selection of OIM using clustering methods such as K-means and AHC in [36] which could only extract mutually exclusive clusters (no object/OIM is a member of more than one cluster).

Experimental-based clustering uses the interestingness values produced by respective OIMs as the clustering dimensions. In other words, instead of taking the OIM-property matrix as input, it takes the Association Rule-OIM value matrix as input for clustering. OIMs that produce either the same interestingness values/rankings for same rules are clustered together. Comparing interestingness rankings is a relaxed version from comparing interestingness values for different OIMs because although two OIMs could have different value ranges (e.g., 0 to 1 or 0 to 1), they can still rank two association rules in the same way. Experimental-based clustering include clustering based on Pearson's Correlation Coefficient [34], the ARQAT tool [35] that utilizes AHC and correlations to cluster rules based on their final values and combining property-based with experimental-based clustering [36].

The experimentation method reveals which OIMs are correlated, which OIMs are not and which OIMs negatively correlate with each other. Correlated OIMs are considered as redundant. This is similar to property based selections but without the need to define and

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

maintain a list of properties. For correlations, the options commonly used are the Pearson's Product Moment Correlation Coefficient, Spearman's Rho and Kendall's Tau. The most common correlation measure is the Pearson's Product Moment Correlation Coefficient (PPMCC, denoted as r) [40], which has a range of possible values from -1 (negatively correlated) to 1 (positively correlated). A value of 0 means that the two variables under investigation are independent of each other (not correlated). Correlation between OIMs can also be calculated through the similarity of the association rule rankings produced by any two OIMs. Two of the popular rank-based similarity measures are namely: Spearman's and Kendall's Tau.

3.1.4 Analysis of OIM selection methods

In this section, we have highlighted the methods to choose an appropriate OIM from a pool of approximately 70 OIMs that is available in the literature. The methods differ in the amount of human interaction required and how well the selection fit the situation. At the high end of human interaction, researchers can opt to do manual research on the desired properties and match it to the properties that OIMs have. This is limited by the resource required to undertake such surveys and the users own understanding of the dataset. The benefit obtained from investing such resources is that the user understands why a rule is rank higher and that results are predictable.

While having required high amount of user intervention in the process of selecting a suitable OIM, the manual inspection and decision aid approaches have the advantage of its complexity is independent of the data size. These approaches take almost the same effort because the properties defined are valid for thousand-row database and also large million-row databases. This point is important as some selection approaches are further complicated by the number of rules present (or the size of the database). This is true for experimental clustering based selection methods.

At the medium to low end of human interaction, users can choose to perform selection based on experimentation. Tan et al. [17] proposed to use a rule sampling method to extract only rules with the most ranking conflict for the user to manually rank or annotate. The best OIM that fits the manual ranking is selected as the appropriate OIM. Without such sampling method, experimental clustering require more computational resources as more association rules are considered.

Users must be aware that the success of uncovering valid correlations from the experimental data (association rule-OIM value matrix) depends on the size of the data. It is highly possible that a small dataset would yield false correlations between OIMs. However, a valid (real) correlation might not be useful because in experimental clustering two OIMs that are positively correlated are redundant in the particular dataset and can be reduced to using only one of them. Thus, it is important to note that while experimental clustering eliminates some human intervention in selecting an OIM, it needs to be performed per dataset basis. Property-based clustering requires higher human intervention, but it is more usable (without alteration if the users required-properties remain the same) across different datasets.

The major benefit of experimentation (clustering and correlations) is that it could reduce the amount of OIMs to consider (within each dataset) without human intervention. However, it still depends on human selection or some automated assumption to select which one to use among the few representative OIMs. In addition, correlations coefficient like Kendall's Tau are computationally expensive.

As a last note in this section, the current studies [17, 27, 33] have noted that no OIMs perform better consistently across different datasets. This is possibly due to the changing distribution in the dataset and suggested that the selection should be done at regular intervals. In addition, to better gauge the benefit of OIM utilization, users should provide more information about the distribution of association rules alongside their experimentation. An interestingness value is an interplay of operands and most OIMs proposed used the contingency table cells as operands in their equation. Thus, if the distribution of contingency tables within the examined ruleset is the same, the categorization of OIMs will inevitably be the same. This is more obvious in a temporal dataset because if the selection is done within short periods of time, the distribution of contingency table (and hence operands) will not differ much to make an impact in the OIM selection.

3.2 Aggregation of OIMs

Apart from choosing which OIM to use, some users followed the path of aggregating the values of multiple OIMs into a single interestingness value. Aggregation of multiple OIMs also include mining rules that satisfy multiple

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

OIMs simultaneously [28, 29, 30, 41, 42, 43, 44, 45] and using evolutionary algorithms to derive their own optimal formula [10, 11, 46]. In a comparison study conducted by Abe and Tsumoto [1], the authors used Principle Component Analysis (PCA) to retrieve seven groups of measures accounting for 92% variability in the dataset and showed that by combining their surveyed 39 measures (with the PCA component as the weight) yields comparable results (in prediction accuracy) against choosing the one best OIM by domain expert. Thus, OIM aggregation is also an appealing method for utilizing OIMs.

3.2.1 Combining values from different OIMs

In general, combining values of different OIMs involve the derivation of a proper weightage function to aggregate the values of various OIMs into one single interestingness value. In this section, we will discuss two methods in which a weightage function could be defined: normalizing function and optimization function. The former involves normalizing the interestingness values while the latter explores genetic algorithm for providing a weightage function.

Nguyen Le et al. [12] proposed the usage of Choquet integral to aggregate the values of different OIMs together into single interestingness value. Recall from earlier section that the set of similar OIMs which were redundant can be removed and those which are not correlated (plus negatively correlated) represents a set of alternative OIMs to choose from. Other than choosing OIMs, we can use the all the OIMs. Although we include the OIMs, we use weights to ensure that the effect of the redundant OIMs are not significant (as compared to contributions of other negatively correlated OIMs) to the single interestingness value. This way the final result is as though we selected on OIM to use, but minus the manual labor of selection.

As a formal definition to the Choquet integral, consider every OIM to be used by the user (labeled M_1 to M_n) will produce interestingness values $(\vartheta_1, \vartheta_2, \dots, \vartheta_n)$ for every association rule. The system would sort the OIMs such that the condition $\vartheta_i \leq \vartheta_{i+1} \leq \vartheta_{i+2} \leq \cdots \leq \vartheta_k$ is fulfilled. With this definition, the Choquet Integral is defined as:

$$C_\mu(\vartheta_1, \vartheta_2, \dots, \vartheta_k) = \sum_{i=1}^{k} \vartheta_i[\mu(S_i) - \mu(S_{i+1})] \quad (3)$$

In the Equation 3, S_k is the sorted set of all OIMs used $(M_1, M_2, \dots, M_k)$, Si is the sorted subset S_k that spans $\{M_i, M_{i+1}, \dots, M_k\}$ and $\mu(S_i)$ is the subset-weight for the sorted subset of OIM values S_i. The single value of interestingness (C_μ $(\vartheta_1, \vartheta_2, \dots, \vartheta_k)$) is in fact the summation of all the sorted interestingness values $(\vartheta_i, \vartheta_{i+1}, \dots, \vartheta_k)$ weighted respectively by the difference in subset-weights of two different sorted subsets. The first is the sorted subset starting with $M_i(S_i)$ while the second sorted subset begins with $M_{i+1}(S_{i+1})$. Equation 4 illustrates the subset-weight for a subset containing two OIMs (e.g., M_x and M_y). In this case, S_i is a set containing only two members $\{M_x, M_y\}$.

$$\mu(S_i) = max[\mu(M_x), \mu(M_y)] - (1 - r)min[\mu(M_x), \mu(M_y)] \quad (4)$$

From the Equation 4 above, we can see that with respect to equal singleton weight ($\mu(M_x)$or $\mu(M_y)$ = 1), the subset weight ($\mu(S_i)$ = r) is actually the correlation coefficient between $M_x$and M_y. To quantify the correlation between OIMs, the authors used the Pearson's correlation coefficient. Due to the fact that users can consider more than two OIMs to be used in their system, Equation 4 is expanded to Equation 5 to cover subsets that contain more than two OIMs. In Equation 5, $\bar{\rho}$ is the mean Pearson's correlation coefficient of the measures in set S_i and S is the subset of S_k that spans $\{M_i, M_{i+1}, \dots, M_k\}$.

$$\mu(S_i) = \max_{M \in S_i}[\mu(S_i \backslash M_i)] - (1 - \bar{\rho})\mu(M_i) \quad (5)$$

The Choquet integral's behavior is such that for those subsets of OIMs that are correlated with each other, the weights assigned to them are done in such a way that it reduces their overall effect. For example, if OIM 1 and OIM 2 are correlated, the Choquet integral will assign weights that will make the aggregation of the values (OIM 1 + OIM 2) smaller than the actual summation of OIM 1 and OIM 2. However, for the subset of OIMs that contains non-correlated OIMs, the Choquet Integral would assign the weights such that the aggregation of their values is equal to its maximal member. Likewise, for subsets of OIMs that are negatively correlated, the aggregated values would be more than the mere summation of their values.

Journal of Theoretical and Applied Information Technology
10th January 2014. Vol. 59 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

If all the OIMs are noncorrelated to each other, the Choquet Integral will produce the same value as the weighted mean of all the OIMs. But if all OIMs are negative correlated to each other but still assigns high interestingness to an association rule, the association rule is said to have the highest possible interestingness. In essence, similar OIMs have less effect on the outcome of the aggregated interestingness, non-correlated OIMs have neutral effect and negatively correlated (complementary) OIMs have larger effect on the aggregated interestingness.

In general, all the OIMs are just a combination of variables using mathematical operators. Through experience, researchers formulate such combinations so that the rules that they wanted would have a higher value (and subsequently being ranked higher among other association rules). This is especially true where there are ample evaluation/training association rules like those in predictive or classification application contexts. Those evaluation mechanisms allow an automated optimization approach towards assigning individual weights to each OIM.

Genetic Network Programming (GNP) [9] algorithm can be used to find an optimal equation to aggregate several OIMs [46]. The algorithm would derive an equation (GNP EQ) that combine the values of several OIMs into one interestingness value (GNP EQ(X_i)) for each association rule (X_i). The association rules are then ranked by their GNP EQ(X_i) interestingness and the rules with high-scores are used for classification. The GNP EQ that provided the highest classification accuracy is considered the best.

The GNP algorithm starts with a small population of random solutions (called Genetic Networks, GNs) and evolves them to converge on a final optimal equation. So, each GN in the population is a candidate to be the best GNP EQ. Each GN has a networked structure consisting of three different types of nodes (Start-Node, Judgment-Node and Processing Node). Within each GN, the Start-Node is where equation combination starts and Judgment-Node controls the flow of combination (whether Processing-Node A links to Processing-Node B or to Processing-Node C). The connections between the nodes are also randomly assigned at first but the Start Node is always the entry point.

The Processing Node holds the most data among the three nodes. Its attributes are a weight value (range of 0 to 1), OIM options (either Support or Confidence), Operator options (addition, subtraction, multiplication, division, square, square root, maximum, minimum and absolute). Those attributes will form sub-equations that will be stored in each processing node. So, in general, each GN is a network of sub-equations (Processing Node) and the algorithm evolves these GNs until they are fit above certain threshold. Figure 2 depicts a simple GN with two Judgment-Node (J_1 and J_2) and two Processing-Node (P_1 and P_2). Intuitively,S is the Start-Node. The attributes of a Processing-Node are symbolized with W_p (weights), M_q (OIMs), o_r (operators) and e_s (sub-equations).

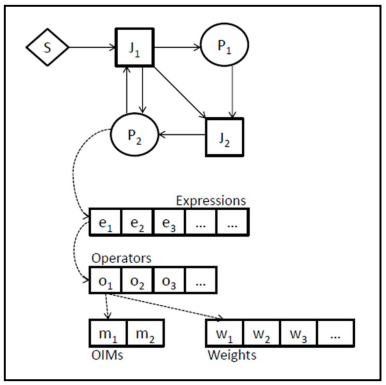

Figure 2: Structure of a Genetic Network and Processing-Node attributes [46].

In the start, several GNs are randomly chosen to form the initial population. Then the algorithm seeks to find out which GN has the best fitness. The fitness is evaluated as a function of the classification accuracy of the top association rules ranked by the GNP EQ. The fit GNs are selected for reproduction through genetic operators. The operators are as follows:

- Cossover: A connection in two preselected GNs are switched.
- Connection-Mutation: A connection within a GN is switched.
- Node-Content-Mutation: The contents of the processing node are changed. This could be that the sub-equation is changed, or just the weight is changed or the options for the sub-equation attributes are changed (e.g., removing the division operator).

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The population gets evolved from generation to generation until an optimal solution with minimum specified fitness is found. The optimal solution is the equation that will be used to aggregate OIM values into one single interestingness value. The initial prototype reported in [46] only allows six possible combinations of Support and Confidence through the addition, subtraction, multiplication and division operators. An example of the resultant fit equation is "$0.11 \cdot Support - 0.6 \cdot Confidence$".

3.2.2 Meta-learning approach to combining multiple OIMs

Delpisheh and Zhang [11] used neural network with back propagation to select a combination of multiple OIMs that would rank the rules as closely as possible to the training set ranked by domain experts. The domain experts would evaluate several association rules (X_i) and give them an interestingness value ($M_{eval}(X_i)$). This will form the training set of which the neural network will use to learn to model the data. Using the same neural network, we would then input the values of unseen observations (in this context, the interestingness values of an association rule, without a predefined ideal value given by the domain expert) and at the end of the network, we will obtain one single interestingness value that is the most optimal (with respect to the training examples).

3.2.3 Finding optimal values for OIMs

In Section 3.1, we have discussed that many OIMs share the same mathematical properties and produce similar interestingness values for the same association rule. Thus, it is also possible that users can mine association rules that satisfy multiple OIMs. This is the area of optimal rule mining and we elaborate these methods as an option towards combining multiple OIMs. However, these methods are delimited by the nature of the OIM. In other words, the OIMs combined usually have their properties formalized and proven to be similar using mathematical process.

In [28, 29, 42], the authors proved that mathematical lower bounds exist for certain group of OIMs and proposed that mining within such lower bounds will satisfy other related OIMs as well. Table 6 lists the null-invariant OIM investigated in [28]. The authors went on to prove that a total-order (Equation 6) exists between the five measures in Table 6.

Table 6: Null-invariant OIMs investigated in [28]

Name	Formula
AllConf (X_i)	$min\big(P(a\|b), P(b\|a)\big)$
Coherence'(X_i)	$\frac{1}{[P(a\|b)^{-1} + P(b\|a)^{-1} - 1]}$
Cos(X_i)	$\sqrt{P(a\|b) \times P(b\|a)}$
Kulc(X_i)	$\frac{P(a\|b) + P(b\|a)}{2}$
MaxConf(X_i)	$max\big(P(a\|b), P(b\|a)\big)$

This implies that given a threshold on a high-order measure (e.g., Kulc), the association rules filtered by the Kulc measure would be a superset of those produced by a lower-order measure (e.g., Cosine or Cos).

$$AllConf \leq Coherence' \leq Cos \leq Kulc \leq MaxConf \quad (6)$$

Partial orders can also be defined to mine association rules that satisfy many interestingness measures [42]. The authors defined two partial orders: $\leq_{sc}$ and $\leq_{s\neg c}$ that are based on the Support ($Supp(X_i)$) and Confidence ($Conf(X_i)$) interestingness values for an association rule (X_i).

For any two association rule X_1 and X_2, $X_1 <_{sc} X_2$ if and only if:

- $Supp(X_1) \leq Supp(X_2) \wedge Conf(X_1) > Conf(X_2)$
- $Supp(X_1) < Supp(X_2) \wedge Conf(X_1) \geq Conf(X_2)$

On the other hand, $X_1 <_{s\neg c} X_2$ if and only if:

- $Supp(X_1) \leq Supp(X_2) \wedge Conf(X_1) < Conf(X_2)$
- $Supp(X_1) < Supp(X_2) \wedge Conf(X_1) \leq Conf(X_2)$

For both partial orders, $X_1 =_{sc} X_2$ (respectively also $X_1 =_{s\neg c} X_2$) is true only if X_1 and X_2 have the same Support and Confidence interestingness value. If one were to plot the Support and Confidence values for all the association rules (X_i) on a Confidence-vs-Support chart, the set of association rules that fulfill the partial order $\leq_{sc}$ will form the upper border while the set of association rules that fulfill the partial order $\leq_{s\neg c}$ will form the lower border. Figure 3 illustrates this concept. With these partial orders,

Journal of Theoretical and Applied Information Technology
10th January 2014. Vol. 59 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

the authors have shown than mining association rules using the upper border (using partial order ≤sc) would yield association rules that satisfy the OIM list of conviction, lift, Laplace, gain and the Piatetsky-Shapiro. Mining the lower border would produce association rules that satisfy the measures highlight in the previous sentence plus the entropy gain, Gini Index and Chi-Square values.

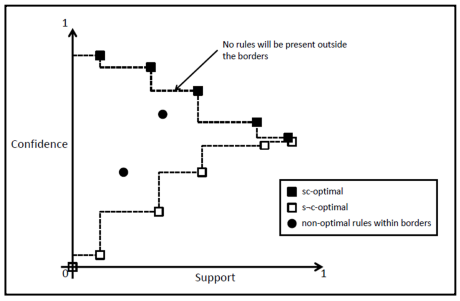

Figure 3: The set of association rules fulfilling the partial orders ($\leq_{sc}$ and $\leq_{s\neg c}$) forming the upper and lower boundaries [42].

Hébert and Cremilleux unified 17 OIMs through the usage of three parameters [41]: minimal antecedent frequency (cell N_a in contingency table, denoted as γ), maximal consequent frequency (cell N_b in contingency table, η) and maximal counter-example frequency (cell N_{ab} in contingency table, δ). The authors have shown that by expressing the 17 OIMs in terms of the three parameters, one can observe that the interaction between the three parameters in each OIM is the same (just the magnitude of interaction different). Users can assign values to the parameters (γ, η, δ) and calculate the corresponding lower bounds for each OIM. The OIM's original formula and its lower bound are illustrated in Table 7. Hence, it is possible to simultaneously optimize all 17 OIMs according to a lower bound. In other words, any association rules will satisfy the minimum interestingness (the lower bound) in all 17 OIMs if it has the following contingency table cell values:

1. $N_a \geq \gamma$
2. $N_b \leq \eta$
3. $N_{ab} \leq \delta$

Evolutionary algorithms are also used to mine rules that satisfy multiple OIMs. They are the class of algorithms modeled over nature. We will elaborate on a variant of evolutionary algorithms as a foundation to this section: Genetic Algorithm (GA). The GA is an algorithm that is modeled using natural evolution of genes. GA solves an optimization problem by having a set of early solution possibilities (encoded as genes) and evolving them into a more optimal population. An optimal solution is the one having the best fitness determined through the fitness function. The evolution is done through processes of mutation, inheritance and crossover. There is also a selection step mimicking the natural selection process to choose the genes qualified for breeding.

Those processes (also known as genetic operators) are performed in cycles and at the end of each cycle, the fitness function is used to check the optimality of the population. If the population is less than optimal, the cycle continues on. Thus the two most important aspects/parameters to consider when comprehending an implementation of GA are:

1. The encoding used for the genes.
2. The genetic operators.

Francisci and Collard [43] employed multi-objective evolutionary approach to mine rules that have the optimal combination of OIMs. In their GA implementation, each association rule X_i is encoded in the Michigan approach. The genetic operators used are crossover and tournament selection. The crossover process grows association rules into larger (or smaller rules) by adjunction (or deletion). In effect, the system starts with single item rules (rules with only antecedent) and grows them until the population achieves a certain fitness level. An optimal population would contain association rules (X_i) that has the most optimal combination of OIM values.

The fitness function is defined as follows: We have a list of OIMs labeled from $M_1, M_2, \dots, M_n$ which provide interestingness values (Equation 7) to association rules labeled from $X_1, X_2, \dots, X_k$.

$$V_i = \left(M_1(X_i), M_2(X_i), \dots, M_N(X_i)\right) \quad (7)$$

The fitness (also known as the Pareto Frontier) is called a non-dominated vector and is defined in Equation 7 where $j, k \in \{1,2, \dots m\}$ are indices to denote each OIM in the set of OIMs. It illustrates that the association rule X_p is more optimal than X_q and hence, X_p is more interesting that association rule X_q. In the context of Equation 8, one can say that the association rule X_p dominates X_q.

$$\forall j: M_j(X_p) \geq M_j(X_q), \exists k: M_k(X_p) > M_k(X_q) \quad (8)$$

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Other researches that used GA in association rule mining includes Anand, Vaid and Singh [45]. Their implementation was set to optimize only four OIMs: comprehensibility, support, confidence and cosine. Martinez-Ballesteros and Riquelme [44] offers enhancement in the GA optimization process by first extracting only a subset of OIMs. Four groups of OIMs are selected through PCA process in which one OIM (with the highest eigenvalue) is selected as the representative of each group. This offers a view into OIMs that have different perspectives (because the selected OIMs represent OIMs with differing properties). As an extension to evolutionary algorithms satisfying multiple OIMs, Nandhini et al. [47] used Particle Swarm Optimization (PSO) to estimate the optimal threshold for Support and Confidence before using it to mine association rules.

Similar to the non-dominated (optimal) set of association rules defined in Equation 7, Bouker et al. [48] proposed an algorithm to find the set of optimal association rules. To find a set of non-dominated association rules, one would need to compare each association rule with every other association rule. This would be time-consuming even if it is possible. Thus, the authors approximated the comparison by designing the algorithm such that, when it checks for each rule for dominance, it only checks the similarity between a candidate rule X_i and a super rule which dominates all other rules. This is in contrast to the taxing process of doing pair-wise comparison for all possible pairs of association rules. They assume the existence of a super association rule and any association rule (X_i) that is similar to the super rule is granted dominance status. This super rule also has the maximal values for all OIMs $(M_1, M_2, \dots, M_m)$. The measure to calculate the similarity between two rules, $Sim(X_p, X_q)$ is illustrated in Equation 9.

$$Sim(X_p, X_q) = \left(\sum_{i=1}^{m} |\widehat{M_i(X_p)} - \widehat{M_i(X_q)}| \right) \cdot m^{-1} \quad (9)$$

Note that since each OIM had its own range of possible values, it is important to normalize the interestingness values before running the Equation 9. Hence normalized OIM values are labeled as $\widehat{M_i(X)}$.

3.2.4 Analysis of OIM aggregation

Aggregation of OIMs allows users to specify weights for a particular OIM. This is an advantage over the OIM selection method because even if there are approximately 70 OIMs available in the literature, it is not possible for them to fit human interest accurately. Aggregation provides a refinement to the existing OIM so that it fits a user interest better. On the contrary, OIM aggregation as can be seen from the literature could result in expressions that cannot be comprehended by the user. For example, in the work of Yang et al. [46], a resultant expression of "$0.11 \cdot Support - 0.6 \cdot Confidence$" will not make such sense to a user especially when the expression changes from time to time to fit the prediction accuracy goal.

For short term exploratory projects, there is no benefit to train a learning algorithm. This highlights the drawback of a learning system: it needs to be trained. Investigations further into the viability of meta-learning as an OIM selection methodology can be done by integrating sampling methods (e.g., in [17]). This way the learning curve could be improved faster and less effort is required by the user to annotate their preference.

The derivation of custom expressions (although done autonomously by optimization algorithms) could be undermined by the fact that the eligible operands also included the approximately 70 measures available in the literature and this amount of options would lead to a combinatorial explosion. However, the derivation can be made tractable if:

- the operand is limited to only contingency table cells.
- feature elimination algorithms be run to retrieve only significant operands, specifically those that contribute highly to the variability across different association rules mined [44].

4. EVALUATION OF OIM USAGE

Selecting an appropriate OIM required the capability to categorize an OIM. In turn, the categorization depends on the method of how similarity between OIMs is calculated. The paper [49] highlights the four potential pitfalls in the effort of categorization of OIMs: rule bias, data bias, expert bias and search bias that users should be concerned about. The first bias (rule bias) refers to the rule-set being mined where different sets of rules would yield different correlations between OIMs. This would not be an issue if the user uses

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

basic form of algorithms like Apriori [15] and FP-Growth [23].

Users should be careful if they use the optimal variant algorithms like [28, 41, 42, 43, 44, 45]. The second bias (data bias) could occur because different datasets from the same domain could yield different correlations between OIMs (even after mitigating the rule-bias). Users investigating or utilizing OIMs should also be aware of the expert and search biases. Suzuki in [49] defined expert bias as the condition where different domain experts will have different opinions. In relation to the selection or aggregation that is based on user inputs, the result would highly dependent on which user was the one who provided the expert opinion. The search bias refers to the methods used in correlating OIMs. As illustrated in preceding sections, researchers could use clustering (property-based or experimental-based), MCDA, PCA or correlation coefficients (Pearson's Coefficient, Spearman's Rho, Kendall's Tau). All the preceding methods would produce differing correlations depending on the input.

The author [49] proposed that varying the datasets and methods of experiment would enable mitigation of those four biases. As an example, Vaillant et al. [33] used experimental clusterings of 20 OIMs across ten datasets to verify the properties they defined for their property-based clustering. Abe and Tsumoto [1] experimented with 32 UCI datasets to verify their grouping of OIMs by underlying theories (e.g., probabilistic, statistical, information and relative-based OIMs). Jalali-Heravi and Zaïane [27] performed their study across 20 UCI datasets. This highlights the importance of setting up the appropriate architecture to support so much validation across different methods and datasets in both selection and aggregation of OIMs.

Similar to the usage of OIMs which should be tailored to each domain, the methods for evaluating the success of using OIMs are also tailored to their application context. The methods to evaluate the effectiveness of the OIM usage used within the literature surveyed includes classification accuracy [3, 24, 27], verified by expert user [1, 50], reduction in rule count [27], against other measures or methods [25, 33, 51] and against content-based similarity [14]. The last method of evaluation is applicable when there is content information available. An example of content information is a database for product information which stores product attributes like the category of the product (food, clothes, stationeries) or the ingredients of the products (or materials of non-food products). From such info, we could deduce similarity between items (apple is more like an orange than a T-shirt) and use this to compare between the associations made by the association rule. In addition, the content similarity can also be used to model user expectations so that the association rule mining can uncover surprising rules.

5. CONCLUSION

Basic algorithms in association rule mining produce an exhaustive list of rules that satisfy the Support-Confidence thresholds. Given this fact, the amount of rules generated are often too large to be processed by a user and too redundant that it becomes noise for automated processing. There is a need to manage this rule overload. One of the methods is through the usage of Objective Interestingness Measures (OIMs) and there exist approximately 70 OIMs available to be used. Thus, the question of reducing rule overload becomes a problem of utilizing the appropriate filters (OIMs). In this paper, we have analyzed recent developments in determining interestingness and methods available to use the OIMs. We split the usage into two: OIM selection and OIM aggregation. After systematic examination of both methods (refer Section 3.1.4 and 3.2.4 respectively), we conclude that OIM selection provides comprehensible results at the expense of human resources while OIM aggregation is better at fulfilling goals set with reduced human intervention albeit the low result-comprehensibility and high computational expenses.

With regard to the fact that each OIM represents a point of view on the data, we noted that these two usages should be used together in an association rule mining process. The selection methods allow users to filter out redundant OIMs to arrive at a smaller set of OIMs and the aggregation method can be used to integrate user-preference weights to combine the values from this smaller set of OIMs. This arrangement will ensure that users do not end up optimizing redundant OIMs that are already correlated. In addition, the selection process or rule sampling can be used to speed up the aggregation process due to the smaller set of OIMs. This is especially true if a computationally expensive correlation coefficient (like the Kendall's Tau) or aggregation method (genetic algorithms) is used.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

At present, OIMs are better used in filtering association rules so that it is more manageable for both users and computers. This is inline with the work of [50] in evaluating the correlations between OIM-measured rules and human interest-measured rules. The study was done using eight datasets and 11 OIMs. Their result was that ranking of rules by OIMs only correlate with the expert user for only 35% of the time. However, with the advent of OIM aggregation, the gap between OIM and human interest could be further reduced. This could be achieved by trading off the comprehensibility of the aggregated expressions and if there is a specific goal to optimize (e.g., in classification problems) that allows the whole process to be automated.

REFERENCES:

[1] Abe, H., Tsumoto, S., "A comparison of composed objective rule evaluation indices using PCA and single indices." *In: Proceedings of the 4th International Conference on Rough Sets and Knowledge Technology. RSKT '09*, Berlin, Heidelberg, Springer-Verlag (2009)

[2] Zhang, L., Yu, D.L., Wang, Y.G., Zhang, Q.M., "Selecting an appropriate interestingness measure to evaluate the correlation between chinese medicine syndrome elements and symptoms." *Chinese Journal of Integrative Medicine (2011)*

[3] Xianneng, L., Mabu, S., Huiyu, Z., Shimada, K., Hirasawa, K., "Analysis of various interestingness measures in classification rule mining for traffic prediction." *In: Proceedings of The Society of Instrument and Control Engineers (SICE) Annual Conference. (Aug 2010)* 1969-1974

[4] Bourassa, M., Fugere, J., Skillicorn, D., "Interestingness - directing analyst focus to significant data." *European Intelligence and Security Informatics Conference 0 (2011)* pp. 300-307

[5] Kao, L.J., Huang, Y.P., "An efficient strategy to detect outlier transactions for knowledge mining." *In: IEEE International Conference on Systems, Man, and Cybernetics (SMC).* (Oct. 2011) pp. 2670-2675

[6] Geng, L., Hamilton, H.J., "Interestingness measures for data mining: A survey." *ACM Computing Survey 38 (September 2006)*

[7] McGarry, K., "A survey of interestingness measures for knowledge discovery." *Knowledge Engineering Review (2005)* pp. 39-61

[8] Hussain, F., Liu, H., Suzuki, E., Lu, H., "Exception rule mining with a relative interestingness measure." *In: Proceedings of the 4th Pacific-Asia Conference on Knowledge Discovery and Data Mining, Current Issues and New Applications. PAKDD '00*, London, UK, Springer-Verlag (2000) pp. 86-97

[9] Yang, G., Shimada, K., Mabu, S., Hirasawa, K., "A nonlinear model to rank association rules based on semantic similarity and genetic network programming." *IEEJ Transactions on Electrical and Electronic Engineering* 4(2) (2009) pp. 248-256

[10] Abe, H., Ohsaki, M., Tsumoto, S., Yamaguchi, T., "Evaluating a rule evaluation support method with learning models based on objective rule evaluation indices - a case study with a meningitis data mining result." *In: Proceedings of the 5th IEEE International Conference on Hybrid Intelligent Systems.* HIS '05, Washington, DC, USA, IEEE Computer Society (2005) pp. 169-174

[11] Delpisheh, E., Zhang, J.Z., "A dynamic composite approach for evaluating association rules." In: The 7th International Conference on Natural Computation (ICNC). (2011) pp. 1893-1898

[12] Nguyen Le, T.T., Huynh, H.X., Guillet, F., "Finding the most interesting association rules by aggregating objective interestingness measures." Springer-Verlag, Berlin, Heidelberg (2009) pp. 40-49

[13] Bonchi, F., Lucchese, C., "Pushing tougher constraints in frequent pattern mining." *In: Proceedings of the 9th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining.* PAKDD'05, Berlin, Heidelberg, Springer-Verlag (2005) pp. 114-124

[14] Wu, J., Zhu, S., Xiong, H., Chen, J., Zhu, J., "Adapting the right measures for pattern discovery: A unified view." *The IEEE Transactions on Systems, Man, and Cybernetics*, Part B: Cybernetics PP(99) (2012) pp. 1-12

[15] Agrawal, R., Srikant, R., "Fast algorithms for mining association rules in large databases." *In Bocca, J.B., Jarke, M., Zaniolo, C., eds., VLDB, Morgan Kaufmann* (1994) pp. 487-499

[16] Brijs, T., Vanhoof, K., Wets, G., "Defining interestingness for association rules" (2005)

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[17] Tan, P.N., Kumar, V., Srivastava, J., "Selecting the right interestingness measure for association patterns." *In: Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '02, New York, NY, USA, ACM (2002) pp. 320-341

[18] Huebner, R.A., "Diversity-based interestingness measures for association rule mining." *American Society of Business and Behavioral Sciences* 16(1) (2009)

[19] Carvalho, D., Freitas, A., Ebecken, N., "A critical review of rule surprisingness measures." *In Ebecken, N., Brebbia, C., Zanasi, A., eds., Proceedings of Data Mining IV - International Conference on Data Mining*, WIT Press (December 2003) pp. 545-556

[20] Freitas, A.A.: "Are we really discovering interesting knowledge from data?" (2006)

[21] Lenca, P., Meyer, P., Vaillant, B., Lallich, S., "On selecting interestingness measures for association rules: User oriented description and multiple criteria decision aid." *European Journal of Operational Research* 184(2) (2008) pp. 610-626

[22] Blanchard, J., Guillet, F., Kuntz, P., "Semantics-based classification of rule interestingness measures." *In: Post-mining of Association Rules: Techniques for Effective Knowledge Extraction. Information Science Reference* - Imprint of: IGI Publishing, Hershey, PA (2009)

[23] Han, J., Pei, J., Yin, Y., Mao, R., "Mining frequent patterns without candidate generation: A frequent-pattern tree approach." *Data Mining and Knowledge Discovery 8* (January 2004) pp. 53-87

[24] Aljandal, W.A., Hsu, W.H., Bahirwani, V., Caragea, D., Weninger, T., "Validation-based normalization and selection of interestingness measures for association rules." *In Proceedings of the 18th International Conference on Artificial Neural Networks in Engineering* (2008) pp. 517-524

[25] Shekar, B., Natarajan, R., "A transaction based neighbourhood-driven approach to quantifying interestingness of association rules." *In: Proceedings of the 4th IEEE International Conference on Data Mining. ICDM '04*, Washington, DC, USA, IEEE Computer Society (2004) pp. 194-201

[26] Blanchard, J., Guillet, F., Briand, H., Gras, R., "Assessing rule interestingness with a probabilistic measure of deviation from equilibrium." (2005)

[27] Jalali-Heravi, M., Zaïane, R.O., "A study on interestingness measures for associative classifiers." *In: Proceedings of the 2010 ACM Symposium on Applied Computing*. SAC '10, New York, NY, USA, ACM (2010) pp. 1039-1046

[28] Wu, T., Chen, Y., Han, J., "Association mining in large databases: A reexamination of its measures." *In Kok, J., Koronacki, J., Lopez de Mantaras, R., Matwin, S., Mladenic, D., Skowron, A., eds.: Knowledge Discovery in Databases: PKDD 2007*. Volume 4702 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg (2007) pp. 621-628

[29] Li, J., "On optimal rule discovery." *The IEEE Transactions on Knowledge and Data Engineering 18* (April 2006) pp. 460-471

[30] Jiuyong, L., Yanchun, Z., "Direct interesting rule generation." *IEEE International Conference on Data Mining* (2003)

[31] Vaillant, B., Lallich, S., Lenca, P., "Modeling of the counter-examples and association rules interestingness measures behavior." *In Crone, S.F., Lessmann, S., Stahlbock, R., eds.: The 2nd International Conference on Data Mining (DMIN)*, CSREA Press (2006) pp. 132-137

[32] Delpisheh, E., Zhang, J.Z., "Evaluating association rules by quantitative pairwise property comparisons." *In: Proceedings of the IEEE International Conference on Data Mining Workshops. ICDMW '10*, Washington, DC, USA, IEEE Computer Society (2010) pp. 927-934

[33] Vaillant, B., Lenca, P., Lallich, S., "A clustering of interestingness measures." *In Suzuki, E., Arikawa, S., eds.: Discovery Science. Volume 3245 of Lecture Notes in Computer Science*. Springer Berlin / Heidelberg (2004) pp. 290-297

[34] Abe, H., Tsumoto, S., "Analyzing behavior of objective rule evaluation indices based on a correlation coefficient." *In: Proceedings of the 12th International Conference on Knowledge-based Intelligent Information and Engineering Systems, Part II. KES '08*, Berlin, Heidelberg, Springer-Verlag (2008) pp. 758-765

[35] Xuan-Hiep, H., Guillet, F., Briand, H., "ARQAT: An exploratory analysis tool for interestingness measures." *In: International Symposium on Applied Stochastic Models and Data Analysis.* (2005)

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[36] Grissa, D., Guillaume, S., Nguifo, E.M., "Combining clustering techniques and formal concept analysis to characterize interestingness measures." *Computing Research Repository (CoRR)* abs/1008.3629 (2010)

[37] Sahar, S., "Exploring interestingness through clustering: A framework." *In: Proceedings of the IEEE International Conference on Data Mining. ICDM '02*, Washington, DC, USA, IEEE Computer Society (2002)

[38] Kannan, S., Bhaskaran, R., "Association rule pruning based on interestingness measures with clustering." *International Journal of Computer Science Issues* abs/0912.1822 (2009)

[39] Belohlavek, R., Grissa, D., Guillaume, S., Nguifo, E.M., Outrata, J., "Boolean factors as a means of clustering of interestingness measures of association rules." *In: Proceedings of the 8th International Conference on Concept Lattices and Their Applications.* (2011)

[40] Rodgers, J.L., Nicewander, W.A., „Thirteen ways to look at the correlation coefficient." *The American Statistician 42* (1988) pp. 59-66

[41] Hébert, C., Crmilleux, B., "A unified view of objective interestingness measures." *In: The 5th International Conference on Machine Learning and Data Mining (MLDM'07)*, Leipzig, Germany, Springer-Verlag (July 2007)

[42] Bayardo, R.J.J., Agrawal, R., "Mining the most interesting rules." *In: Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, New York, NY, USA, ACM (1999) pp. 145-154

[43] Francisci, D., Collard, M., "Multi-criteria evaluation of interesting dependencies according to a data mining approach." *In: Congress on Evolutionary Computation*, IEEE Press (2003) pp. 1568-1574

[44] Martínez-Ballesteros, M., Riquelme, J.C., "Analysis of measures of quantitative association rules." *In: Proceedings of the 6th International Conference on Hybrid Artificial Intelligent Systems - Volume PartII. HAIS'11*, Berlin, Heidelberg, Springer-Verlag (2011) pp. 319-326

[45] Anand, R., Vaid, A., Singh, P.K., "Association rule mining using multi-objective evolutionary algorithms: Strengths and challenges." *In: The IEEE World Congress on Nature and Biologically Inspired Computing (NaBIC)*, IEEE (2009) pp. 385-390

[46] Yang, G., Mabu, S., Shimada, K., Hirasawa, K., "An evolutionary approach to rank class association rules with feedback mechanism." *Expert System Applications 38(12)* (2011) pp. 15040-15048

[47] Nandhini, M., Janani, M., Sivanandham, S., "Association rule mining using swarm intelligence and domain ontology." *In: Recent Trends In Information Technology (ICRTIT)*, 2012 International Conference on. (April 2012) pp. 537-541

[48] Bouker, S., Saidi, R., Yahia, S.B., Nguifo, E.M., "Ranking and selecting association rules based on dominance relationship." *Technical report*, Clermont Universit, Universit Blaise Pascal, LIMOS, BP 10448, F-63000 CLERMONT-FERRAND, http://hal.archives-ouvertes.fr/hal-00677853. (May 2012)

[49] Suzuki, E., "Pitfalls for categorizations of objective interestingness measures for rule discovery." *In: Statistical Implicative Analysis, Theory and Applications. Volume 127 of Studies in Computational Intelligence.* Springer (2008) pp. 383-395

[50] Carvalho, D.R., Freitas, A.A., Ebecken, N.: "Evaluating the correlation between objective rule interestingness measures and real human interest." *In: Proceedings of the 9th European Conference on Principles and Practice of Knowledge Discovery in Databases. PKDD'05*, Berlin, Heidelberg, Springer-Verlag (2005) pp. 453-461

[51] Tan, P.N., Kumar, V., "Interestingness measures for association patterns: A perspective." *In: The 6th ACM SIGKDD Workshop on Postprocessing in Machine Learning and Data Mining: Interpretation Visualization, Integration and Related Topics. (2000),*

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 3: Criteria/Properties For Evaluating OIM [21]

Property	Description	Possible Values
g1 – Treatment of antecedent and consequent	How the OIM evaluates "A→B" and "B→"?	Symmetric / Asymmetric
g2 – Decrease with N_b	Is the interestingness calculated by a particular OIM a decreasing function of N_b?	Y / N
g3 – Independence value	Is the value of interestingness of an association rule is a constant or variable when the antecedent is independent (probabilistically) of the consequent.	Constant / Variable
g4 – Logical rule value	Is the value of interestingness of an association rule is a constant or variable when there are no counter examples to the rule.	Constant / Variable
g5 – Sensitivity towards counter example	How much will the interestingness value change when small addition of counter examples?	Change Type (Convex, Linear, Concave)
g6 – Sensitivity towards total records	How will the interestingness value change when more rules are added?	Increase / Invariant
g7 – Ease of threshold determination	How easy is it to set a threshold for the value of the OIM to filter association rules?	Easy / Hard
g8 – Intelligibility of the OIM	How easily the semantics of the OIM be understood?	A / B / C

Table 4: Additional Criteria/Properties Used By [32] For Evaluating OIM.

Property	Description	Possible Values
Treatment of counter example	Interestingness stays constant when there exists no counter example for the association rule.	1 / 0
Response to row or column permutation	Interestingness values changes its sign when either one of the row and column of the contingency table are permuted.	1 / 0
Response to both row and column permutation	Interestingness values sign remain unchanged if both the row and column of the contingency table are permuted.	1 / 0
Null invariance	Interestingness value must be null invariant. In the contingency table context, the cell Nab does not have an effect on the interestingness.	1 / 0

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 5: Criteria/Properties For Evaluating OIM [36].

Property	Description
Invariance in quantity expansion	The interestingness value produced by the OIM should not change when the cells Nab, Nab are expanded by constant K_1 and cells Nab, Nab by constant K_2.
Asymmetric in conclusion negation	The interestingness is different for a rule A→ B and its counter example A→B
Attraction case between A and B	A and B are said to be an attraction case if the joint probability of A and B is more than the product of their individual probability.
Repulsion case between A and B	Repulsion of A and B occurs when the joint probability of A and B is less than the product of their individual probability.
Relationship between A→B and Ā→B	An OIM with this property must hold true for the interestingness of Ā→B is the negation (opposite) of the interestingness A→B.
Relationship between A→B and A→B	An OIM with this property must hold true for the interestingness of Ā→B is the negation (opposite) of the interestingness A→B.
Relationship between A→B and A→B	An OIM with this property must hold true for the interestingness of Ā→B is the negation (opposite) of the interestingness A→B.
Descriptive statistical measure	A descriptive OIM is invariant if the data is dilated by a constant K. Otherwise, it is a statistical measure.
Antecedent size is fixed or random	The antecedent size is uncertain when the OIM is based on probabilistic models.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 7: OIM Formulas And Lower Bounds In Terms Of Γ, Δ, H [41]. The Original Form Are Written With Contingency Table Notation From Section 1.

OIM	Formula	Lower Bounds
Support	$\frac{N_{ab}}{N}$	$\frac{\gamma-\delta}{N}$
Confidence	$\frac{N_{ab}}{N_a}$	$1-\frac{\delta}{\gamma}$
Sensitivity	$\frac{N_{ab}}{N_b}$	$\frac{\gamma-\delta}{\eta}$
Specificity	$1-\frac{N_a-N_{ab}}{N-N_b}$	$1-\frac{\delta}{\mathrm{N}-\eta}$
Success Rate	$\frac{N-N_b-N_a+2N_{ab}}{N}$	$1+\frac{\gamma-2\delta-\eta}{N}$
Lift	$\frac{N\cdot N_{ab}}{N_a\cdot N_b}$	$\left(1-\frac{\delta}{\gamma}\right)\cdot\frac{N}{\eta}$
Rule Interest	$N_{ab}-\frac{N_a\cdot N_b}{N}$	$\gamma-\delta-\frac{\gamma\cdot\eta}{N}$
Laplace (K=2)	$\frac{N_{ab}+1}{N_a+2}$	$\frac{\gamma-\delta+1}{\gamma+2}$
Odds Ratio	$\frac{N_{ab}}{N_a-N_{ab}}-\frac{N-N_b-N_a+N_{ab}}{N_b-N_{ab}}$	$\left(\frac{\gamma-\delta}{\eta-\gamma+\delta}\right)\cdot\left(\frac{N-\eta-\delta}{\delta}\right)$
Growth Rate	$\left(\frac{N_{ab}}{N_a-N_{ab}}\right)\cdot\left(\frac{N-N_b}{N_b}\right)$	$\left(\frac{\gamma-\delta}{\delta}\right)\cdot\left(\frac{N-\eta}{\eta}\right)$
Sebag & Schoenauer	$\frac{N_{ab}}{N_a-N_{ab}}$	$\frac{\gamma-\delta}{\delta}$
Jaccard	$\frac{N_{ab}}{N_b+N_a-N_{ab}}$	$\frac{\gamma-\delta}{\eta+\delta}$
Conviction	$\left(\frac{N-N_b}{N}\right)\cdot\left(\frac{N_a}{N_a-N_{ab}}\right)$	$\left(\frac{N-\eta}{N}\right)\cdot\left(\frac{\gamma}{\delta}\right)$
Θ-coefficient	$\frac{N\cdot N_{ab}-N_b\cdot N_a}{\sqrt{N_a\cdot N_b\cdot(N-N_b)\cdot(N)-N_b}}$	$\frac{\gamma\cdot\left(N-\eta\right)-\delta\cdot N}{\sqrt{\gamma\cdot(N-\gamma)\cdot\eta\cdot\left(\mathrm{N}-\eta\right)}}$
Added Value	$\frac{N_{ab}}{N_a}-\frac{N_b}{N}$	$\left(\frac{\gamma-\delta}{\gamma}\right)\cdot\left(\frac{\eta}{N}\right)$
Certainty Factor	$\frac{N_{ab}\cdot N-N_a\cdot N_b}{N_a\cdot(N-N_b)}$	$\frac{\gamma\cdot\left(N-\eta\right)-\delta\cdot N}{\gamma\cdot\left(N-\eta\right)}$
Information Gain	$\log\left(\frac{N_{ab}}{N_a}\cdot\frac{N}{N_b}\right)$	$\log\left(\frac{\gamma-\delta}{\gamma}\cdot\frac{N}{\eta}\right)$