

Interestingness Measures In Rule Mining: A Valuation

Garima*, Jyoti Vashishtha**

*(Department of Computer Science, Guru Jambheshwar University of Science and Technology, Hisar, India)

** (Department of Computer Science, Guru Jambheshwar University of Science and Technology, Hisar, India)

ABSTRACT

In data mining it is normally desirable that discovered knowledge should possess characteristics such as accuracy, comprehensibility and interestingness. The vast majority of data mining algorithms generate patterns that are accurate and reliable but they might not be interesting. Interestingness measures are used to find the truly interesting rules which will help the user in decision making in exceptional state of affairs. A variety of interestingness measures for rule mining have been suggested by researchers in the field of data mining. In this paper we are going to carry out a valuation of these interestingness measures and classify them from numerous perspectives.

Keywords: Accuracy, Comprehensible, Data Mining, Interestingness Measures

I. INTRODUCTION

Data mining is an algorithmic process for extracting valuable patterns from data, as a large amount of data is collected daily which contain a lot of information/knowledge that can help us in decision making [1]. It is a step in Knowledge Discovery in Databases (KDD) process. KDD is 'the non-trivial process (complex computation required) of identifying valid (true for new data), novel (new to user), potentially useful (actionable), and ultimately comprehensible (understandable) knowledge from databases' [2].

According to data mining tasks, patterns can be represented in many different forms, including classification rules and association rules which are usually the two most popular techniques in data mining [1]. **Association rule** is an implication of the form $Y \rightarrow Z$, where $Y \cap Z = \emptyset$. For example, $\{\text{Hardware}\} \rightarrow \{\text{Software}\}$ is an association rule which says that when Hardware is purchased, software is likely to be purchased as well. **Classification rule** is an implication of the form $Y_1 \text{ oper } y_1, Y_2 \text{ oper } y_2, \dots, Y_n \text{ oper } y_n \rightarrow Z = z$, where Y_i is a conditional attribute, y_i is a value that belongs to the domain of Y_i , Z is the class attribute, z is a class value, and oper is a relational operator such as = or >. For example, Marital Status = yes, Returned = no > Risky = no, is a classification rule which says that a client who is married and not returned loan is classified as loan application is not risky [3].

Although association and classification rule both can be represented as "if-then" rules, but their purposes are different. Association rules are ordinarily used as **descriptive** tools for finding association relationship among a set of objects in database while Classification rules as **predictive**

tools for understanding existing data and predicting classifications for unseen data [4,5,6]. In data mining it is desirable that mined patterns should satisfy some **properties** such as predictive accuracy, comprehensibility and interestingness [7]. *Predictive Accuracy* should be high for the discovered knowledge. This is most important property of mined rules. It can be calculated as $PreAcc = \text{Number of testing examples correctly classified by the rule set} / \text{Total number of testing example}$. [3]. *Comprehensibility* means the discovered rules should be understandable to the user as these are ultimately used by the user in decision making. *Interestingness* means the discovered rules should be novel or surprising to the user.

The discovered patterns that are accurate and reliable are not necessarily interesting if these are previously known to the user [8]. As an example a rule that is accurate and comprehensible is:

IF (person is 5 year old) THEN (he cannot drive)
But this rule is not interesting as everyone knows this. So an example of interesting rule is:

IF(refund=no, marital_status = married)
THEN(cheat= no).

As this rule is previously not known as well as accurate and comprehensible, it is interesting rule which we are concerned about.

Interestingness measures are classified as subjective or objective [3,9,10]. *Objective Measures* (also called as data driven) are based only on raw data, and no knowledge about the user or application is required. Support, Confidence, Coverage etc. are the objective measures. *Subjective Measures* (also called as user driven) are based on both data and user of data, knowledge about the background or user's domain is required. Silberschatz and Tuzhilin's interestingness is the subjective measure. The word "interestingness" has several different meanings in

the data mining literature. In some cases, a particular behavior in a domain might be interesting while the same behavior in another domain may not be interesting. So, different interestingness measures can be used in different situations to find the strongly correlated rules.

II. INTERESTINGNESS MEASURES / LITERATURE REVIEW

Interestingness measures are very important area of research these days. Researchers have discovered a lot of interestingness measures that can be used to find interesting rules or to reduce the number of mined rules. As classification rules generated as a result of data mining are used for the prediction of unseen data, the most common measure that is used to evaluate the quality of classification rules is **predictive accuracy**, which is defined as [3] in equation 1:

$$\text{Accuracy} = \frac{\text{Number of testing examples correctly classified by the ruleset}}{\text{Total no. of testing examples}}$$

Predictive Accuracy, support and confidence are the basic measures for association and classification rules but many other measures have also been proposed which have their own importance and use. Most of these are derived from these basic measures.

2.1. Piatetsky-Shapiro's Rule interest function [11] is used to evaluate the correlation b/w attributes in simple classification rule or between antecedent and consequent. A simple classification rule is one where the left and right hand side of the rule $X \rightarrow Y$ contains single attributes. The rule interest function can be evaluated by equation 2:

$$RI = \frac{|X \cap Y| - (|X| |Y|) / N}{N} \quad (2)$$

Where N denotes the total no. of tuples, $|X|$ and $|Y|$ denotes the number of tuples satisfying conditions X and Y respectively, $|X \cap Y|$ denotes the number of tuples satisfying $X \rightarrow Y$, and $(|X| |Y|) / N$ denotes the no. of tuples expected if X and Y are independent.

1. $RI=0$ if $|X \cap Y| = (|X| |Y|) / N$
 2. RI monotonically increases with $|X \cap Y|$ when other parameters are fixed;
 3. RI monotonically decreases with $|X|$ or $|Y|$ when other parameters are fixed;
- When $RI=0$, then X and Y are independent to each other and the rule is not interesting.
 When $RI < 0$ ($RI > 0$) then X is negatively (positively) correlated to Y .

2.2. Smyth and Goodman's J-Measure [12] is used to represent the average information content of a probabilistic classification rule. This measure is used to find the best rules in context of discrete-valued

attributes. Probabilistic classification rule is the rule of form $X \rightarrow Y$ having some probability p , where the left and right hand sides contain single attributes. The right hand side is restricted to have single valued assignment expressions, while the left hand side may be a combination of these simple expressions. The J-measure can be denoted by :

$$J(x; y) = p(y) \left[p(x/y) \log\left(\frac{p(x/y)}{p(x)}\right) + (1 - p(x/y)) \log\left(\frac{1 - p(x/y)}{1 - p(x)}\right) \right] \quad (3)$$

Where $p(x)$, $p(y)$ and $p(x/y)$ represent the probabilities of occurrence of x , y and x given y respectively. The term which is inside the square bracket is relative (cross) entropy. Relative entropy is the similarity of two probability distributions.

High value for $J(x; y)$ is desirable, but it is not necessary that its high value give the best rule.

2.3. Major and Mangano's rule refinement [13] is a strategy which is used for induction of interesting classification rules from a database of classification rules. It consists of three phases: identifying potentially interesting rules (are those that satisfy specified confidence, coverage, and simplicity (i.e. rule length) criteria), identifying technically interesting rules (are selected from potentially interesting rules according to simplicity and statistical significance (i.e. chi-square test) criteria), removing rules that are not genuinely interesting (is a manual task performed by the domain expert. This task involves keeping the simplest and most general rules and removing other similar rules).

2.4. Silberschatz and Tuzhilin's interestingness [14] is used to determine the extent to which a soft belief (is one that a user is willing to change as new evidence encountered) is changed as a result of encountering new evidence (i.e. discovered knowledge). Interestingness within the context of soft beliefs can be calculated by:

$$I = \frac{\sum \partial |P(\partial|E, \xi) - P(\partial|\xi)|}{P(\partial|\xi)} \quad (4)$$

Where E is the new evidence, ∂ is the belief, ξ is the previous element supporting belief ∂ , $P(\partial|\xi)$ is the confidence in belief ∂ , $P(\partial|E, \xi)$ is the new confidence in belief ∂ given the new evidence E and Summation is over all beliefs. Bayes theorem is used to determine the new confidence and can be calculated by:

$$P(\partial|E, \xi) = \frac{P(E|\partial, \xi)P(\partial|\xi)}{P(E|\partial, \xi)P(\partial|\xi) + P(E|\neg\partial, \xi)P(\neg\partial|\xi)} \quad (5)$$

Positive (Negative) evidence strengthens (weakens) the belief.

2.5. Agarwal and Shrikant's itemset measures [15,16] are used for the identification of frequently

occurring association rules from set of items in large databases. An association rule is defined as: Let $I = \{i_1, i_2, \dots, i_m\}$ be a set of items and D be a set of transactions, where each transaction T is a set of items such that $T \subseteq I$. An association rule is an implication of the form $A \rightarrow B$, where $A \subset I$, $B \subset I$, and $A \cap B = \varnothing$.

The rule $A \rightarrow B$ holds for the dataset D with support s and confidence c if $s\%$ of transactions in D contain $A \cup B$ and $c\%$ of transactions in D that contain A also contain B . From these definitions we can say that confidence correspond to the strength of the rule, while support corresponds to the statistical significance. Those rules which satisfy a predetermined minimum threshold for support and confidence are considered to be interesting.

2.6. Matheus and Piatetsky-Shapiro's Projected savings [17] is an interestingness measure that estimate the financial impact of cost deviations from some normative or expected value or to forecast the percentage savings in medical domain. Projected saving can be calculated by :

$$PS = PI * SP \quad (6)$$

Where SP is the saving percentage and PI is the projected impact. The projected impact can be calculated by:

$$PI = PD * IF \quad (7)$$

Where IF is impact factor (which can be the no. of units sold) that relates to increased profit and PD is difference b/w the current average cost or the expected or normative cost for some service or product. Saving percentage (SP) is the value of the percentage decrease in the deviation specified by the domain expert that would result following some relevant intervention strategy. Interestingness of a deviation is directly related to the projected saving that is achieved as a result of this strategy.

2.7. Hamilton et al. Creditability [18] is used to determine the extent to which a classification (i.e. generalized relation) provides decisions for all or nearly all possible values of condition attributes, based upon evidence supported adequately.

2.8. Liu et al. General Impressions [19] is proposed as an approach for evaluating the importance of classification rules. It compares the discovered rules to an approximate or vague description of what is considered to be interesting. Thus a general impression can be considered as a kind of specification language. General impression is a rule of the form:

$$B_1 OP_1, B_2 OP_2, \dots, B_x OP_x \rightarrow C_j.$$

Where $B_i OP_i$ is called an impression term, B_i is an attribute, each OP_i is an impression descriptor from like $<, >, |$ etc. and C_j is the class. The $>$ ($<$) impression descriptor means larger or smaller

attribute values are more likely to be included in class C_j .

2.9. Freitas AttSurp (Attribute Surprisingness) [20, 21] is based on the degree of surprisingness associated with the individual attributes which occurred in rule antecedent [21]. The degree of surprisingness of an attribute can be calculated as the inverse of its information gain [31]. So, an attribute having low information gain in a rule tends to be more surprising, because this kind of attribute is usually considered to be of little relevant for classification purposes. Although an attribute can have a low information gain individually, it is possible that, when it is combined with other attributes, then the attribute interaction makes the former relevant for classification, and this kind of attribute interaction has the potential to be very surprising to the user. Mathematically, AttSurp was originally defined by:

$$AttSurp = 1 / \sum_{i=1}^k \text{InformationGain}(A_i) / k \quad (8)$$

where $\text{InformationGain}(A_i)$ is the information gain of the i -th attribute in the rule antecedent and K is the total number of attributes in the rule antecedent. Here when the information gain values are very low then the value of AttSurp can be very large, which makes it difficult to compare the value of this formula with other rule surprisingness measures.

2.10. Gago and Bento's Distant matrix [22] measures the distance b/w the classification rules and determines the rules that provide the highest coverage for the given data. It is given in equation 9:

$$D(r_i, r_j) = \begin{cases} \frac{DA(r_i, r_j) + 2DV(r_i, r_j) - 2EV(r_i, r_j)}{N(r_i) + N(r_j)}, & NO(r_i, r_j) = 0 \\ 2, & \text{otherwise} \end{cases}$$

Where r_i, r_j are the rules i, j respectively, $DA(r_i, r_j)$ is the sum of number of attributes in r_i not in r_j and number of attributes in r_j not in r_i , $DV(r_i, r_j)$ is the number of attributes in r_i and r_j that have slightly overlapping (less than 66%) values in range condition, $EV(r_i, r_j)$ is the number of attributes in r_i and r_j that have overlapping (more than 66%) values in range condition, $N(r_i), N(r_j)$ is the number of attributes in r_i and r_j and $NO(r_i, r_j)$ is the number of attributes in r_i and r_j with non overlapping values. Its range is $[-1, 1]$ (strong or slightly overlap) or 2 (no overlap). The rules having the highest average distance to are considered to be most interesting.

2.11. Gray and Orlowska's interestingness [23] is used to evaluate the strength of associations b/w association rules or sets of items in retail transactions.

Support and confidence are the most basic measures for association rules, but interestingness contains a discrimination component that gives an indication of the independence of antecedent and consequent. Interestingness can be calculated by:

$$I = \left\{ \left(\frac{P(X \cap Y)}{P(X) * P(Y)} \right)^k - 1 \right\} * (P(X) * P(Y))^m \quad (10)$$

where $P(X) * P(Y)$ is the support, $P(X \cap Y)$ is the confidence, $P(X \cap Y) / (P(X) * P(Y))$ is discrimination, and k and m are the parameters that calculate the relative importance of discrimination and support component respectively. Higher the value of interestingness more interesting the rules are.

2.12. Zhong et al. Peculiarity [24] determines the extent to which one data object differs from other similar data objects. The Peculiarity Factor can be calculated by:

$$PF(x_i) = \sum_{j=1}^n N(x_i, x_j)^a \quad (11)$$

where x_i, x_j are the attributes values, N is the total no. of different attribute values, $N(x_i, x_j)$ is the conceptual distance b/w x_i and x_j and a is the user defined parameter. The conceptual difference can be calculated by:

$$N(x_i, x_j) = |x_i - x_j| \quad (12)$$

2.13. Lavrac et al. Novelty [25] means a person did not know about the pattern before and is not able to infer it from other known patterns, i.e. the discovered patterns should be new to the organization. A rule $A \rightarrow B$ is novel if $P(AB)$ cannot be inferred from $P(A)$ and $P(B)$. So, novelty can be calculated as:

$$Novelty = P(AB) - P(A)P(B) \quad (13)$$

2.14. Noda et al. Normalized AttSurp [26] in this the original formula was normalized to return values in the range [0, 1].

$$AttSurp = 1 - \left[\frac{\sum_{i=1}^k \text{InformationGain}(A_i)/k}{\log_2(\text{number of classes})} \right] \quad (14)$$

It is well-known that the information gain measure is biased towards attributes having many values. As the AttSurp measure favors attributes with a small information gain, AttSurp is biased towards attributes having few values. As it favors the attributes with a low information gain, so it favors rules where accuracy is not so large. It is essential that it be used together with another rule quality criterion instead of using alone to evaluate rule quality favoring more accurate rules.

2.15. Freitas' Surprisingness [20, 27] is a measure that determines the interestingness of discovered knowledge via the explicit detection of occurrences of Simpson's paradox and then calculates its

magnitude, using the ranking as an indication of the degree of surprisingness. Simpson's paradox is a phenomena according to which an event increases the probability in the super-population but also decreases the probability in the sub-populations comprising the data. The effects can also be experienced in the opposite direction, i.e. the sub-populations can seemingly have the opposite effect to the super-population.

2.16. Korn et al. Ratio rules [28] are a technique that employs eigensystem analysis to calculate correlations between values of attributes, which reveals the axes of greatest variation and thus the most important correlations. The interestingness measure proposed by Korn to assess the quality of the ratio rules is called as a 'guessing error' (GE). The GE refers to the estimation of calculating missing values in the data matrix. The GE can be calculated as

$$GE = \sqrt{\frac{1}{NM} \sum_{i=1}^N \sum_{j=1}^M (\widehat{x}_{ij} - x_{ij})^2} \quad (15)$$

Where M is the number of products, N is the number of customers, \widehat{x}_{ij} is the estimated missing value and x_{ij} is the actual value. A low value for GE for a ratio rule implies that it has identified a novel pattern in the data with high confidence.

2.17. Chen, Liu, Li's Influence [29] Let $p(X) = |X|/|T|$ and $p(Y/X) = p(XY)/p(X)$, where $X, Y \subset I$ (I is the itemset) and $X \cap Y = \emptyset$. The influence of association rule $X \rightarrow Y$ is given as in equation 16:

$$\begin{aligned} \text{Influence}(X \rightarrow Y) &= \text{Influence}(XY) \\ &= \log \frac{p(Y/X)/p(\sim Y/X)}{p(Y)/p(\sim Y)} \\ &= \log \frac{\text{confidence}(X \rightarrow Y)/\text{confidence}(X \rightarrow \sim Y)}{\text{support}(Y)/\text{support}(\sim Y)} \end{aligned}$$

Where $p(Y)/p(\sim Y)$ is the non-conditional contrast between positive and negative facts, while $p(Y/X)/p(\sim Y/X)$ is the contrast with condition X . The change in contrast that is caused by X can reflect the influence of X on Y .

When influence = 0 then the antecedent lacks association with the consequent.

When influence > 0 i.e. positive then antecedent is positively associated with the consequent (i.e. positive influence) and

When influence < 0 i.e. is negative the antecedent is negatively associated with the consequent (negative influence).

2.18. Chen, Liu, Li's Conditional Influence [29] of itemset Z on itemset Y on the condition of X , is defined as equation 17:

$$\begin{aligned} & \text{Influence}(Z, Y/X) \\ &= \text{Influence}(XZ, Y) \\ & \quad - \text{Influence}(XY) \\ &= \text{Influence}(XZ \rightarrow Y) - \text{Influence}(X \rightarrow Y) \\ &= \log \frac{\text{confidence}(XZ \rightarrow Y)/\text{confidence}(XZ \rightarrow \sim Y)}{\text{confidence}(X \rightarrow Y)/\text{confidence}(X \rightarrow \sim Y)} \end{aligned}$$

This tells us that conditional influence of Z on Y is the difference between influence of XZ on Y and that of X on Y; it is clear to see that influence can reflect confidence deviation and when $X = \varphi$, conditional influence degenerates to influence. So we can say that influence is a special form of conditional influence. Similar to influence, When $\text{Influence}(Z, Y/X) > 0$ then Z has positive influence on Y with condition X. When $\text{Influence}(Z, Y/X) < 0$ then Z has negative influence on Y with condition X and When $\text{Influence}(Z, Y/X) = 0$ then Z has no influence on Y with condition X.

2.19. Malone et al. Differential ratio rules [30] is an improvement over Kort et al. Ratio rules by adding a temporal element to them in the form of differential ratio (DR) rules which is capable of detecting interesting patterns in spatio-temporal data. DR data mining measures variation of a given object in terms of the pair wise ratios of the elements describing the data over time. Consider two variables *a* and *b* as elements of a given object. Calculation of a single DR (here, DR will be referred to as the measure of difference calculated by this process) between two time points *t* and *t+1* can be calculated by:

$$DR_t = \begin{cases} \log \frac{a_t/b_t}{a_{t+1}/b_{t+1}} & \text{when } a \leq b \\ \log \frac{b_t/a_t}{b_{t+1}/a_{t+1}} & \text{when } b < a \end{cases} \quad (21)$$

These kind of calculations can be performed for a time series ($t=1, \dots, n$). Using the definition of interestingness, the ratio between variable *a* and *b* over time point *t* and *t+1*:

$DR_t \approx 0$, ratio has remained constant i.e. the ratios between the variables has barely altered over time.

$DR_t = 0$ exactly zero means no difference at all.

$DR_t < 0$, ratio of difference has decreased over time i.e. the two variables values are becoming closer together in terms of the two ratios over time.

$DR_t > 0$, ratio of difference has increased over time i.e. the two variables values are growing further apart in terms of the two ratios over time.

The magnitude of the measure also has a proportional meaning since the greater the value the more change has occurred.

2.20. Blanchard et al. Reduced Entropy [31] In order to remove the symmetry introduced by the

entropy in the measure *i*, directed entropic function $\hat{H}(A)$ also called reduced entropy was introduced. The reduced entropy $\hat{H}(A)$ of a variable *A* is defined by:

$$\text{if } p(A = 1) \begin{cases} \leq 1/2 \text{ then } \hat{H}(A) = 1 \\ \geq 1/2 \text{ then } \hat{H}(A) = H(A) \end{cases} \quad (18)$$

One similarly that defines the conditional reduced entropy of the variable B given the realization of A is shown in equation 19:

$$\text{if } p \left(B = \frac{1}{A} = \begin{cases} \leq 1/2 \text{ then } H(B/A=1) = 1 \\ \geq 1/2 \text{ then } H(B/A=1) = H(B/A=1) \end{cases}$$

The entropy $H(A)$ of a variable *A* can be written as the sum of two reduced entropies:

$$H(A) = \hat{H}(A) + \hat{H}(\bar{A}) - 1 \quad (20)$$

Where \bar{A} is the negation of *A*. Contrary to H , \hat{H} is an asymmetric measure which evaluates an imbalance in favor of $A = 1$ and in favor of $A = 0$: $\hat{H}(A) \neq \hat{H}(\bar{A})$. More precisely, if $A = 1$ is more frequent than $A = 0$, then the reduced entropy $\hat{H}(A)$ measures the entropy of A : $\hat{H}(A) = \hat{H}(\bar{A})$ and the reduced entropy $\hat{H}(A) = 1$. If $A = 1$ is less frequent than $A = 0$, then their roles are reversed. Or we can say, \hat{H} measures a "directed uncertainty" in favor of one of the values, in the sense that if this value is not the more likely, then the uncertainty is considered to be maximal.

2.21. Blanchard et al. DIR (Directed Information ratio) [31] is a new measure which is based on information theory. It is designed for association rule (useful for post processing) and it differentiate two opposite rules $A \rightarrow B$ and $A \rightarrow \bar{B}$. it rejects both, the rules whose antecedent and consequent are negatively correlated, and the rules which have more counter-examples than examples. DIR of a rule $A \rightarrow B$ is defined as in equation 22:

$$DIR(A \rightarrow B) = \frac{\hat{H}(B) - \hat{H}\left(\frac{B}{A=1}\right)}{\hat{H}(B)}$$

if $p(B = 1) \neq 1$

if $p(B = 1) = 1$, then $\hat{H}(B) = 0$ and DIR is not defined. However, such rules should be discarded as these are completely expected. If DIR is strictly positive then the rule is said to be informative. DIR is asymmetric as it does not assign the same value to $A \rightarrow B$ and to its opposite $A \rightarrow \bar{B}$ and it also does not assign the same value to $A \rightarrow B$ and to its converse $B \rightarrow A$.

2.22. J. Vashishtha's Intra Class Exception [32] determines the rare features of an object within its class. For default rule, the following representation is used:

$$\text{IF } P \text{ THEN } d_i \dot{w}(E_1 \cup E_2 \cup E_3 \dots) : \partial_1 \partial_2 \partial_3 \partial_4$$

Where E is the unique intra class exception with respect to class d_i , \hat{w} is with operator that is used for augmentation of exception with default rule. $\partial_1, \partial_2, \partial_3, \partial_4$ are the point parameters. ∂_1 is precision, ∂_2 is recall, ∂_3 is support of intra class exception w.r.t. default rule and $\partial_4 = 1$. If a rule satisfies $\partial_1, \partial_2, \partial_3, \partial_4$ then it gives interesting rules.

2.23. J. Vashishtha's Inter-class Exception [32] are the exceptional features that change the class of an object. For default rule in multi class dataset, the following representation is used:

$$IF P THEN d_i U'(E_j \cup (d_k) \cup \dots) : \partial_1 \partial_2 \partial_5 \partial_6, \\ i, k \in [1, m], m = \text{number of class}, i \neq k$$

Where d_i is default class predicted by P which changes to another class d_k by enclosure. $P \cap E_j = \emptyset$, U' is unless operator for augmentation. If a rule satisfies $\partial_1, \partial_2, \partial_5, \partial_6$ then it gives interesting rules. The classification of different interestingness measures is shown in table 1.

III. CONCLUSION

Data mining algorithms can generate large quantity of rules, most of which are of no interest to the user. For the discovery of truly interesting rules, various interestingness measures are suggested in data mining literature. These interestingness measures are classified as objective or subjective.

Table 1: Classification of different interestingness measures

Interestingness measure	Rule Type	Measure Type	Year
Piatetsky-Shapiro's Rule interest function	Classification Rules	Objective	1991
Smyth and Goodman's J-Measure	Classification Rules	Objective	1991
Major and Mangano's rule refinement	Classification Rules	Objective	1995
Silberschatz and Tuzhilin's interestingness	Format-Independent	Subjective	1995
Agarwal and Shrikant's itemset measures	Association Rules	Objective	1994, 1996
Matheus and Piatetsky-Shapiro's Projected savings	Summary	Subjective	1996
Hamilton et al. Creditability	Generalized Relation	Objective	1997
Liu et al. General Impressions	Classification Rules	Subjective	1997
Freitas AttSurp (Attribute Surprisingness)	Format-Independent	Objective	1997
Gago and Bento's Distant matric	Classification Rules	Objective	1998
Gray and Orłowska's interestingness	Association Rules	Objective	1998
Zhong et al. Peculiarity	Association Rules	Objective	1999
Lavrac et al. Novelty	Association Rules	Subjective	1999
Noda et al. Normalized AttSurp	Format-Independent	Objective	1999
Freitas' Surprisingness	Format-Independent	Objective	1998,2000
Korn et al. Ratio rules	Ratio Rules	Objective	000
Chen, Liu, Li's Influence	Association Rules	Objective	2001
Chen, Liu, Li's Conditional Influence	Association Rules	Objective	2001
Malone et al. Differential ratio rules	Ratio Rules	Objective	2004
Blanchard et al. Reduced Entropy	Association Rules	Objective	2005
Blanchard et al. DIR	Association Rules	Objective	2005
J. Vashishtha's Intra Class Exception	Classification Rules	Objective	2013
J. Vashishtha's Inter Class Exception	Classification Rules	Objective	2013

Objective measures (data driven) are based only on raw data and no knowledge of domain is required. Subjective measures (user driven) are based on both raw data and knowledge of user's domain is required. Selecting interestingness measure is an important issue of human interest and this valuation will help

user to select appropriate measure. The future perspective could be the combination of objective and subjective measures and also to design a method which will help user to automatically select suitable measure.

REFERENCE

- [1] J. Han, M. Kamber, and J. Pei, *Data Mining Concepts and Techniques Third Edition, 3rd ed.* (Morgan Kaufmann 2011).
- [2] Fayyad, Usama M., et al. "Advances in knowledge discovery and data mining." (1996).
- [3] Geng, Liqiang, and Howard J. Hamilton, *Interestingness measures for data mining: A survey*, *ACM Computing Surveys (CSUR)* 38.3 (2006): 9.
- [4] Vashishtha, Jyoti, Dharminder Kumar, and Saroj Ratnoo, *Revisiting Interestingness Measures for Knowledge Discovery in Databases*, *Advanced Computing & Communication Technologies (ACCT)*, 2012 *Second International Conference on*. IEEE, 2012.
- [5] Gopalan, Janaki, Reda Alhajj, and Ken Barker, *Discovering Accurate and Interesting Classification Rules Using Genetic Algorithm*, *DMIN*. 2006.
- [6] Liu, Bing, Minqing Hu, and Wynne Hsu, *Multi-level organization and summarization of the discovered rules*, *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2000.
- [7] Triantaphyllou, Evangelos, and Giovanni Felici, *Data mining and knowledge discovery approaches based on rule induction techniques* (Vol. 6. Springer, 2006).
- [8] Carvalho, D. R., A. A. Freitas, and N. F. F. Ebecken. "A critical review of rule surprisingness measures." *Proc. Data Mining IV-Int. Conf. on Data Mining*. Vol. 7. WIT Press, 2003.
- [9] Silberschatz, Abraham, and Alexander Tuzhilin, *What makes patterns interesting in knowledge discovery systems*, *Knowledge and Data Engineering, IEEE Transactions on* 8.6 (1996): 970-974.
- [10] Sahar, Sigal, *Interestingness via what is not interesting*, *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1999.
- [11] Piatetsky-Shapiro, Gregory, *Discovery, analysis and presentation of strong rules*, *Knowledge discovery in databases* (1991): 229-238.
- [12] Smyth, Padhraic, and Rodney M. Goodman, *Rule Induction Using Information Theory*, *Knowledge discovery in databases* 1991 (1991).
- [13] Major, John A., and John J. Mangano, *Selecting among rules induced from a hurricane database*, *Journal of Intelligent Information Systems* 4.1 (1995): 39-52.
- [14] Silberschatz, Abraham, and Alexander Tuzhilin, *On subjective measures of interestingness in knowledge discovery*, *KDD*. Vol. 95. 1995.
- [15] Agrawal, Rakesh, and Ramakrishnan Srikant, *Fast algorithms for mining association rules*, *Proc. 20th int. conf. very large data bases, VLDB*. Vol. 1215. 1994.
- [16] Srikant, Ramakrishnan, and Rakesh Agrawal, *Mining quantitative association rules in large relational tables*, *ACM SIGMOD Record*. Vol. 25. No. 2. ACM, 1996.
- [17] Matheus, Christopher J., Gregory Piatetsky-Shapiro, and Dwight McNeill, *20 Selecting and Reporting What is Interesting: The KEFIR Application to Healthcare Data* (1996).
- [18] Hamilton, Howard J., Ning Shan, and Wojciech Ziarko, *Machine learning of credible classifications*, *Advanced Topics in Artificial Intelligence*. Springer Berlin Heidelberg, 1997. 330-339.
- [19] Liu, W. Hsu, and S chen, *Using general impressions to analyse discovered classification Rules*, *In proceedings of third international conference on knowledge discovery and data mining (KDD' 97)*, Newport Beach, California, august 1997.
- [20] Freitas, Alex A, *On objective measures of rule surprisingness*, *Principles of Data Mining and Knowledge Discovery*. Springer Berlin Heidelberg, 1998. 1-9.
- [21] Mitchell, Tom M, *Machine Learning*, MacGraw-Hill, USA ,1997.
- [22] Gago, Pedro, and Carlos Bento, *A metric for selection of the most promising rules*, *Principles of Data Mining and Knowledge Discovery*. Springer Berlin Heidelberg, 1998. 19-27.
- [23] Gray, Brett, and Maria E. Orlowska, *CCAIIA: Clustering categorical attributes into interesting association rules*, *Research and Development in Knowledge Discovery and Data Mining*. Springer Berlin Heidelberg, 1998. 132-143.
- [24] Zhong, Ning, Y. Y. Yao, and Setsuo Ohsuga, *Peculiarity oriented multi-database mining*, *Principles of Data Mining and Knowledge Discovery*. Springer Berlin Heidelberg, 1999. 136-146.
- [25] Lavrac, N., Flach, P., and Zupan, B. *Rule evaluation measures: A unifying view*. *In Proceedings of the Ninth International Workshop on Inductive Logic Programming (ILP'99)*, (Dzeroski and Flach eds), pages

- 174–185, Bled, Slovenia, 1999. Springer-Verlag.
- [26] Noda, E; Freitas, A.A.; Lopes, H.S. (1999) *Discovering interesting prediction rules with a genetic algorithm*, *Proc. Congress on Evolutionary Computation(CEC-99)*, 1322-1329. Washington D.C., USA.
- [27] Pearl, J, *Causality*. Cambridge: Cambridge University Press, 2000.
- [28] Korn, Flip, et al., *Quantifiable data mining using ratio rules*, *The VLDB journal* 8.3-4 (2000): 254-266.
- [29] Chen, Guoqing, De Liu, and Jiexun Li, *Influence and conditional influence-new interestingness measures in association rule mining*, *Fuzzy Systems, 2001. The 10th IEEE International Conference on*. Vol. 3. IEEE, 2001.
- [30] Malone, J, McGarry, K and Bowerman, C, 2004, *Performing trend analysis on spatio-temporal proteomics data using differential ratio rules*. In *Proceedings of the 6th EPSRC Conference on Postgraduate Research in Electronics, Photonics, Communications and Software*, University of Hertfordshire, UK, pp. 103–105.
- [31] Blanchard, Julien, et al, *Using information-theoretic measures to assess association rule interestingness*, *Data Mining, Fifth IEEE International Conference on*. IEEE, 2005. *l'Information*, E-2:287–298, 2004. Actes EGC2004.
- [32] Vashishtha, Jyoti, Dharminder Kumar, and Saroj Ratnoo, *An evolutionary approach to discover intra-and inter-class exceptions in databases*, *International Journal of Intelligent Systems Technologies and Applications* 12.3 (2013): 283-300.