EAI Endorsed Transactions
on Context-aware Systems and Applications
Research Article ICST.ORG

Classification of objective interestingness measures

Lan Phuong Phan[1,*], Nghia Quoc Phan[2], Vinh Cong Phan[3], Hung Huu Huynh[4], Hiep Xuan Huynh[1], and Fabrice Guillet[5]

[1] Can Tho University, Campus 2- 3/2 Street, Ninh Kieu District, Can Tho City, Vietnam
[2] Tra Vinh University, No. 126 National Road 53, Ward 5, Tra Vinh City, Vietnam
[3] Nguyen Tat Thanh University, 300A Nguyen Tat Thanh Street, Ward 13, District 4, Ho Chi Minh City, Vietnam
[4] VNUK Institute for Research and Executive Education, The University of Danang, 41 Le Duan Street, Hai Chau District, Danang city, Vietnam
[5] Polytech Nantes, University of Nantes, La Chantrerie rue Christian Pauc BP 50609 F-44306 Nantes Cedex 3, France

Abstract

The creation of the interestingness measures for evaluating the quality of the association rule - based knowledge plays an important role in the post-processing of the Knowledge Discovery from Databases. More and more interestingness measures are proposed by two approaches (subjective assessment and objective assessment), studying the properties or the attributes of the interestingness measures is important in understanding the nature of the objective interestingness measures. In this paper, we focus primarily on the objective interestingness measures to obtain a general view of recent researches on the nature of the objective interestingness measures, as well as complete a new classification on 109 selected objective interestingness measures on 6 criterions (independence, equilibrium, symmetry, variation, description, and statistics).

Keywords: objective interestingness measures, classification, property/criterion of interestingness measures, association rules.

Received on 19 April 2016, accepted on 14 July 2016, published on 12 September 2016

doi: 10.4108/eai.12-9-2016.151678

1. Introduction

The process of knowledge discovery from databases (KDD) (Fayyad et al., 1996) is usually divided into three main stages: preprocessing, processing or forming knowledge patterns (mining), and post-processing those patterns. The evaluation of the interestingness or the quality of the patterns found in the processing stage is always one of the contents attracting researchers. During the last decade, the research community in the KDD field recognized the post-processing stage to evaluate the interestingness or the quality of the knowledge patterns generated from the processing stage to be a complex and important part of the KDD process (Silberschartz and Tuzhilin, 1996; Liu et al., 1999; Hilderman and Hamilton, 2001; Tan et al., 2004). For solving this problem, most of approaches are based on the creation of the interestingness measures. From the initial approaches (Piatetsky-Shapiro, 1994; Piatetsky-Shapiro and Matheus, 1991; Agrawal and Srikant, 1994) to the recent approaches, many interestingness measures with reciprocal nature has been proposed to search the best knowledge with many views, perspectives and different evaluations (Sahar and Mansour, 1999) such as summarization (Hildermand and Hamilton, 2001), objectiveness (Tan et al., 2004; Huynh et al., 2007; Bayardo and Agrawal, 1999; Guillet and Hamilton, 2007; Tamir and Singer, 2006; McGarry, 2005; Geng and Hamilton, 2006; Omiecinski, 2003; Weng et al., 2010; Shaharanee et al., 2011; McGrane and Poon, 2010; Jalalvand et al., 2008; Huynh et al., 2008) and subjectiveness (Silberschatz and Tuzhilin, 1996).

The interestingness measures can be divided into two types (Silberschatz and Tuzhilin, 1996): subjective interestingness measures and objective interestingness

*Corresponding author. Email: pplan@ctu.edu.vn

measures. The subjective measures evaluate the found knowledge patterns by basing on the target, the knowledge, and the belief of user. The objective measures evaluate knowledge patterns by basing on the distribution of data.

This article focuses on studying the evaluation criteria in theory for objective measures. These objective measures are commonly used for evaluating the quality of knowledge patterns in the association rule form X ⟶ Y (Agrawal and Srikant, 1994).

The article is organized into six sections. Section 1 introduces the approaches of interestingness measures generally. Section 2 is about an overview of subjective interestingness measures. Section 3 presents objective interestingness measures, and the method to calculate their values by using association rules. Section 4 analyses and summarizes the basic criteria in evaluating the quality of objective measures. Section 5 classifies those objective measures by using some key criterions, and raises the comments concerning the measure nature. The last section summarizes some achieved important results.

2. Subjective interestingness measures

Subjective measures (Piatetsky-Shapiro and Matheus, 1994; Silberschatz and Tuzhilin, 1995, Silberschatz and Tuzhilin, 1996) were studied in the domain-independent context. The interestingness or the benefit from an achieved knowledge pattern (e.g., an association rule, classification rule, etc.) is subjectively evaluated by the view and the perspective of user. A knowledge pattern is usually identified as an interesting or useful one by basing on two approaches (Silberschatz and Tuzhilin, 1996): (i) a knowledge pattern is considered to be unexpectedness if it causes users to surprise (Silberschatz and Tuzhilin, 1995); (ii) and a knowledge pattern is considered to be actionability if users can build actions from the found knowledge, and those actions bring benefit to users (Piatetsky-Shapiro and Matheus, 1994).

2.1. Actionability

Actionability is a subjective interestingness measure allowing users to create some actions in response to the newly found knowledge (Silberschatz and Tuzhilin, 1996). The method for capturing association rules and using them to propose the actionable patterns is always a difficult issue. One of the important factors affecting the above mentioned issue is the required actions (i.e., from the perspective of the individual) which can change over time, and are also very difficult to retain.

The found knowledge patterns resulting in suggested actions can be found via the system exploring the change of rules (Piatetsky-Shapiro and Matheus,1994), the hierarchical structure of actions, or the extraction of patterns responding to actions.

2.2. Unexpectedness

Unexpectedness is a subjective interestingness measure which provides the knowledge patterns not previously anticipated, and being contradictory to the users' expectation (Silberschatz and Tuzhilin, 1996). The users' expectation depends strongly on the user's belief. The belief can be divided into two types: (i) the hard belief – the belief constraints are unchanged and depend strongly on the users' perspective, and (ii) the soft belief - the user wants to change to a certain allowed level of the belief. The level of the soft belief can be associated with different approaches such as Bayesian, Dempster-Shafer, frequency of the occurrence, or statistics.

An association rule (i.e., a knowledge pattern) will always be interesting or beneficial if it is contrary to the existing hard belief of users. For the soft belief, the interestingness of a knowledge pattern p can be calculated as the follow $I(p,B,\xi)=\sum_{\alpha_i\in B} w_i|d(\propto_i|p,\xi)-d(\alpha_i|\xi)|$ with w_i is the weight function associated with each the soft belief α_i in the soft belief system B, $\sum_{\propto_i\in B} w_i=1$ và ξ are the events occurring before.

3. Objective interestingness measures

Suppose that τ is a finite set of transactions (e.g., transactions of customers in a supermarket (Agrawal and Srikant, 1994)). An association rule is represented in the form $X\longrightarrow Y$ where X and Y are two disjoint sets $X\cap Y=\emptyset$. Set X (set Y) is attached to a subset of transactions $t_X=\tau(X)=\{T\in\tau, X\subseteq T\}$ $(t_Y=\tau(Y) resp.)$. Set $\bar{X}$ $(\bar{Y})$ is attached to $t_{\bar{X}}=\tau(\bar{X})=\tau-\tau(X)=\{T\in\tau, X\subsetneq T\}$ $(t_{\bar{Y}}=\tau(\bar{Y})\ resp.)$. In order to accept or reject the tendency of Y's appearance when X has appeared, normally $n_{X\bar{Y}}$ (negative examples, contra-examples) which does not tend to support the rule formation $X\longrightarrow Y$ would be interested. Each rule is characterized by 4 parameters: $n=|\tau|, n_X=|t_X|, n_Y=|t_Y|, n_{\bar{X}}=|t_{\bar{X}}|, n_{\bar{Y}}=|t_{\bar{Y}}|$ (Figure 1).

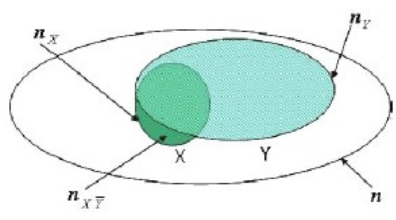

Figure 1. The cardinality of an association rule $X\longrightarrow Y$

For more clearly, notations $p(X)$ $(p(Y), p(X\cap Y), p(X\cap \bar{Y}))$ representing probabilities of X $(Y, X\cap Y, X\cap\bar{Y})$ respectively are retained. This probability is calculated by the frequency of X: $p(X)=\frac{n_X}{n}$ $(p(Y)=\frac{n_Y}{n}, p(X\cap Y)=\frac{n_{XY}}{n}, p(X\cap\bar{Y})=\frac{n_{X\bar{Y}}}{n}\ resp.)$

The interestingness value of an association rule based on an objective interestingness measure will then be calculated by using the cardinality of the rule: $m(X \longrightarrow Y) = f(n, n_X, n_Y, n_{X\bar{Y}}) \in R$. To calculate easily, the following equivalent transformations should be used: $n_{XY} = n_X - n_{X\bar{Y}}$, $n_{\bar{X}} = n - n_X$, $n_{\bar{Y}} = n - n_Y$, $n_{\bar{X}Y} = n_Y - n_X + n_{X\bar{Y}}$, $n_{\bar{X}\bar{Y}} = n - n_Y - n_{X\bar{Y}}$.

For example, two given sets X and Y and an association rule is in the form $X \longrightarrow Y$ where $n = 100, n_X = 50, n_Y = 80, n_{X\bar{Y}} = 10$; and the objective interestingness measure, Pavillon, is identified by the formula: $m(X \longrightarrow Y) = f(n, n_X, n_Y, n_{X\bar{Y}}) = \frac{n_{\bar{Y}}}{n} - \frac{n_{X\bar{Y}}}{n_X}$, the interestingness value is: $m(X \longrightarrow Y) = \frac{80-10}{100} - \frac{10}{50} = 0.5$.

The formulae of interestingness measures calculated by using the cardinality $(n, n_X, n_Y, n_{X\bar{Y}})$ are collected and presented in Table 1 (see Appendix).

4. Evaluation criteria

In order to understand how an objective interestingness measure is "good", several criteria have been proposed (Bayardo and Agrawal, 1999; Hilderman and Hamilton, 2001; Guillet and Hamilton 2007; Lallich and Teytaud, 2004; Lallich et al., 2005; Piatetsky-Shapiro, 1991; Silberschatz and Tuzhilin, 1995; Tan et al., 2004; Geng and Hamilton, 2006). The basic criteria will be discussed in the following content.

4.1. Value variation

Determining the variation of interestingness values is always one of the most important criteria in evaluating interestingness measures. The interestingness value increases monotonically with n_{XY} and decreases monotonically with $n_{X\bar{Y}}$ or $n_{\bar{X}Y}$. It should be noted that values of $n_\omega(n_{XY}, n_{X\bar{Y}}, n_{\bar{X}Y})$ vary while the other parameters are the fixed values. This helps us track the variation of interestingness values clearly and homogeneously.

The trend of the values decline of an interestingness measure should start slowly when there are appearances of elements or transactions that do not support the existence of the studied association rules by the reasons such as a change, a noise, and an error (Figure 2). These interestingness values then should decrease rapidly when there are more and more appearances of elements that do not support the formation of rules, and strongly threaten to the formation of the existence of association rules being reviewed and evaluated. The interestingness values of an objective measure have to also decline as there are more and more appearances of the unimportant transactions (i.e., it does not contain any useful information according to Shannon entropy), which does not contain any information about the formation of association rules.

In addition, a good objective interestingness measure is not allowed to output interestingness values varying linearly with the number of elements that do not support the formation of the corresponding rule.

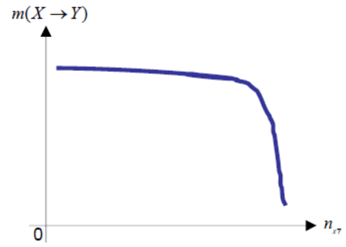

Figure 2. A “good” variation of an interestingness measure

4.2. Particular situation

Observing and evaluating particular situations that occur during the variation of interestingness values is an important method to understand the behaviour of interestingness measures effecting on association rules deeply. Two important particular situations are investigated: independence and equilibrium. Both situations are called the subject of an objective interestingness measure.

Independence occurs when the antecedent and the consequent of an association rule are independent together according to statistical factors. This situation occurs when $n_{XY} = \frac{n_X n_Y}{n}$ or $n_{X\bar{Y}} = \frac{n_X n_{\bar{Y}}}{n}$, then the interestingness value of the rule is a constant.

$$m(X \longrightarrow Y) = f\left(n, n_X, n_Y, \frac{n_X n_{\bar{Y}}}{n}\right) = constant$$

Equilibrium occurs when the number of elements that support the formation of a rule and the number of elements that does not support the formation of that rule are equal. This situation occurs when $n_{XY} = n_{X\bar{Y}} = \frac{n_X}{2}$, then the interestingness value of the rule is a constant.

$$m(X \longrightarrow Y) = f\left(n, n_X, n_Y, \frac{n_X}{2}\right) = constant$$

By considering the variation of interestingness values from independence value or equilibrium value, the interestingness measure will be evaluated as the change tendency from independence value or equilibrium value.

Moreover, the determination of a threshold of an interestingness value will be necessary if we wish to observe a limited range of the benefit value. When $n_{X\bar{Y}} = 0$, the association rule tends to become a logical rule. In this case, the implicative tendency of an association rule will not exist, and the association rule is not itself as well as loses its interestingness.

4.3. Paradoxical situation

The interestingness values of a measure are not the same when the paradoxical situation occurs such as in the *symmetric* situation $m(X \longrightarrow Y) = m(Y \longrightarrow X)$ or in the inverse situation $m(X \longrightarrow Y) = m(X \longrightarrow \bar{Y})$.

4.4. Countable situation

The analysable criterion of an interestingness measure (i.e., countable) helps determine the order or create a pre-order structure.

4.5. Diversification

Interestingness measures have to be fully analysed on the flexibility and the generality when they are handled and applied on the different types of variables.

4.6. Discriminative ability

The discriminative ability of an objective interestingness measure is not affected by a noise or a big capacity data (i.e., n increases). If the interestingness value of a measure is not vary when its input parameters vary with a certain coefficient α: $m(X \longrightarrow Y) = f(n, n_X, n_Y, n_{X\bar{Y}}) = f(\propto \cdot n, \alpha \cdot n_X, \alpha \cdot n_Y, \alpha \cdot n_{X\bar{Y}})$, then that measure is called a *descriptive* measure (a *statistical* measure in the otherwise).

The descriptive or the statistical aspect of a measure is also known as the nature of the measure.

4.7. Interpretable situation

The execution time of formulas and algorithms used to calculate the interestingness values of association rules is not been too long. Their definitions have to be assessed visually, and the obtained values have to be explainable.

4.8. Imbalance

The unbalanced problem will be interested when the effect of a little number of elements that does not support the formation of association rules (i.e. $n_{X\bar{Y}} \ll n$) is observed. This attention is essential because it can bring the extremely valuable knowledge.

4.9. Attribute interestingness

An interested association rule in the entire set of rules may lead to the situation in which two rules will have the same interestingness value. These two rules can have two different degrees of interestingness for users. The distinction is based on the appearance of the attribute in the rule antecedent. To solve this problem, the degrees of interestingness of each attribute appearing in the rule antecedent of an association rule need to be interested.

4.10. Quasi-

Determining the quasi- relationships in calculating the interestingness values is placed in the context to be determined, in some cases, some of the relationships among objective interestingness measures. Relationships to be considered are quasi-implication, quasi-conjunction and quasi-equivalence.

An interestingness measure is a quasi-implication if that measure satisfies the condition $m(X \longrightarrow Y) = m(\bar{Y} \longrightarrow \bar{X})$ where $f(n, n_X, n_Y, n_{X\bar{Y}}) = f(n, n - n_Y, n - n_X, n_{X\bar{Y}}) = f(n, n_{\bar{Y}}, n_{\bar{X}}, n_{X\bar{Y}})$.

An interestingness measure is a quasi-conjunction if that measure satisfies the condition $m(X \longrightarrow Y) = m(Y \longrightarrow X)$ where $f(n, n_X, n_Y, n_{X\bar{Y}}) = f(n, n_Y, n_X, n_{X\bar{Y}})$.

An interestingness measure is quasi-equivalence if that measure satisfies the condition $m(X \longrightarrow Y) = m(Y \longrightarrow X) = m(\bar{Y} \longrightarrow \bar{X}) = m(\bar{X} \longrightarrow \bar{Y})$ where $f(n, n_X, n_Y, n_{X\bar{Y}}) = f(n, n_Y, n_X, n_{X\bar{Y}}) = f(n, n_{\bar{Y}}, n_{\bar{X}}, n_{X\bar{Y}}) = f(n, n_{\bar{X}}, n_{\bar{Y}}, n_{\bar{X}Y})$.

5. Classification of interestingness measures

In this research, to collect the interestingness measures, the selected articles have to own the following criteria: (i) studying the interestingness measures and being cited by many others articles, (ii) being published by the reliable sources such as IEEE, Springer, ACM, Science Direct, (iii) being researched and analysed by the research groups independently.

The collected result shows that there are: (i) 21 groups of interestingness measures in which each group consists of the measures called by different names but having the same formula (in Table 2); (ii) 109 interestingness measures presented (in Table 3).

Table 2. Interestingness measures called by different names but having the same formula

N	Group	N	Group
1	Phi-Coefficient, Correlation Coefficient, Pearson's correlation coefficient, Linear-Correlation, Newrelevancy	5	Jaccard, Coherence
2	Cosine, Ochia, IS Measure	6	Added value, Pavillon, Centred confidence
3	Loevinger, Certainty Factor, Satisfaction	7	Bayes factor, Odd multiplier
4	Piatetsky-Shapiro, Pearl, Leverage 2, Carnap, Novelty	8	Kappa coefficient , Cohen

9	Examples and counter-examples rate, Example and contra-example rate, Encountered rate	16	Specificity 1, Negative Reliability
10	Accuracy, Causal support	17	Relative Risk , Class correlation ratio
11	Descriptive Confirmed-Confidence , Ganascia Index	18	Lerman similarity index, Directed Contribution to Chi square
12	Lift, Interest	19	Gray and Orlowska's Interestingness Weighting Dependency, I - Measure
13	Mutual Information, 2-way Support variation	20	Kulczynski 1, Agreement–Disagreement Index
14	F-Measure, Dice Index, Czekanowski Dice	21	Normalized difference, Match
15	Probabilistic measure of deviation from equilibrium(IPEE), Indice Probabiliste d'Ecart d'Equilibre		

Table 3. Interestingness measures

N	Interestingness Measure	N	Interestingness Measure
1	1-way Support	9	Causal-Confidence
2	2-way Support	10	Causal-Confirmed confidence
3	Accuracy, Causal Support, Sokal-Michener Index	11	Loevinger, Certainty Factor, Satisfaction
4	Added value, Pavillon, Centred Confidence	12	Chi-square
5	All Confidence	13	Relative Risk , Class correlation ratio
6	All Confidence Difference	14	Kappa coefficient , Cohen
7	Bayes factor, Odd multiplier	15	Jaccard, Coherence
8	Brin's Conviction	16	Collective strength
17	Complement Class Support	36	Examples and counter-examples rate, Example and contra-example rate, Encountered rate
18	Conditional Entropy Measure	37	Expected Frequency
19	Confidence	38	Gain
20	Causal Confirm	39	Gini index
21	Confidence Expectation	40	Goodman–Kruskal
22	Confidence Interestingness	41	Implication index
23	Confidence Difference	42	Implication Intensity
24	Contramin	43	Improvement
25	Conviction	44	Probabilistic measure of deviation from equilibrium(IPEE), Indice Probabiliste d'Ecart d'Equilibre
26	Phi-Coefficient, Correlation Coefficient, Pearson's correlation coefficient, Linear Correlation, Newrelevancy	45	Information gain
27	Cosine, Ochia, IS Measure	46	Information Ratio Contraposé (TIC)
28	Coverage	47	(Gray and Orlowska's) Interestingness Weighting Dependency, I - Measure
29	F-Measure, Dice Index, Czekanowski Dice	48	Ion
30	Descriptive Confirmed-Confidence , Ganascia Index	49	J-measure
31	Descriptive-Confirm	50	J1-measure
32	Dilated Chi-square	51	Jaccard Difference
33	Lerman similarity index, Directed Contribution to Chi square	52	Jaccard Expectation
34	Entropic Implication Intensity 1	53	Kemeny–Oppenheim
35	Entropic Implication Intensity 2	54	Klosgen

55	K-measure	82	Pointwise Mutual Information
56	Kulczynski1, Agreement–Disagreement Index	83	Prevalence
57	Kulczynski 2	84	Putative Causal Dependency
58	Kulczynski index	85	Quadratic entropy
59	Laplace	86	Quotient
60	Least contradiction	87	R Cost
61	Leverage (1)	88	Ralambrodrainy
62	Lift, Interest	89	Ratio of link
63	Logical Necessity	90	Recall, Completeness
64	Log-likelihood	91	Rectangular Gain
65	Log-ratio	92	Rogers and Tanimoto index
66	Max Confidence	93	Rule Interest
67	Mutual Expectation	94	S Cost
68	Mutual Information MI, 2-way Support Variation	95	Sebag and Schoenauer
69	NConf	96	Shannon conditional entropy
70	Newl	97	Specificity 1, Negative Reliability
71	Normalized difference, Match	98	Specificity 2
72	Normalized Expectation	99	Support
73	Normalized Mutual Information	100	Support Error
74	ϕ AllConfidence	101	Support Expectation
75	ϕ Confidence	102	Support Interestingness
76	ϕ Jaccard	103	T Combined Cost
77	ϕ Jaccard Difference	104	Theil Uncertainty Coefficient
78	ϕ Kappa	105	U Cost
79	Odd's ratio	106	Wang index
80	Ohsaki's Conviction	107	Yule's Q (indice de Yule)
81	Piatetsky-Shapiro, Pearl, Leverage 2, Carnap, Novelty	108	Yule's Y
		109	Zhang Zhang

This research focuses on some important criteria mentioned in the previous section. They are Variation (VAR.), Independence (IND.), Equilibrium (EQU.), Symmetric (SYM.), Descriptive (DES.), and Statistical (STA.). Table 4 presents the responses of 109 interestingness measures for these criterions where 1 is responsive, and 0 is unresponsive.

Table 4. The responses of 109 interestingness measures for 6 criterions

N	VAR.	IND.	EQU.	SYM.	DES.	STA.
1	0	0	0	0	1	0
2	0	0	0	1	1	0
3	1	0	0	1	1	0
4	1	1	0	0	1	0
5	0	0	0	0	1	0
6	1	0	0	0	1	0
7	1	1	0	0	1	0
8	0	0	1	0	1	0
9	1	0	0	0	1	0
10	0	0	0	0	1	0
11	1	1	0	0	1	0
12	0	0	0	1	0	1
13	1	0	0	0	1	0
14	0	1	0	1	1	0
15	1	0	0	1	1	0
16	1	0	0	0	0	1
17	0	0	0	0	1	0
18	0	0	0	0	1	0
19	0	0	1	0	1	0
20	1	0	0	0	1	0
21	0	0	0	0	1	0
22	1	0	0	0	1	0
23	0	0	0	0	1	0
24	1	0	0	0	1	0
25	0	1	0	0	1	0
26	1	1	0	1	1	0
27	1	0	0	1	1	0
28	0	0	0	0	1	0
29	1	0	0	1	1	0
30	0	0	1	0	1	0
31	0	0	1	0	1	0
32	0	0	0	1	0	1
33	0	1	0	1	0	1
34	1	1	0	0	0	1
35	1	1	0	0	0	1
36	1	0	1	0	1	0
37	0	0	0	1	1	0
38	0	0	1	0	1	0
39	0	0	0	0	1	0
40	0	0	0	0	1	0
41	0	0	0	0	0	1
42	1	1	0	0	0	1
43	0	0	0	0	1	0
44	0	0	1	0	0	1
45	1	1	0	1	1	0
46	1	1	0	0	1	0
47	0	1	0	1	1	0
48	0	1	0	0	1	0

49	0	1	0	0	1	0
50	0	1	0	1	1	0
51	0	1	0	0	1	0
52	0	0	0	0	1	0
53	0	1	0	0	1	0
54	0	1	0	0	1	0
55	0	0	0	0	1	0
56	0	0	0	1	1	0
57	0	0	0	1	1	0
58	1	0	0	0	1	0
59	0	0	1	0	0	1
60	0	0	1	0	1	0
61	0	0	0	0	1	0
62	0	1	0	1	1	0
63	1	0	0	0	1	0
64	0	1	0	0	1	0
65	0	1	0	1	1	0
66	0	0	0	1	1	0
67	0	0	0	1	1	0
68	0	0	0	0	1	0
69	0	1	0	0	1	0
70	0	0	0	0	0	1
71	0	1	0	0	1	0
72	0	0	0	1	1	0
73	0	0	0	0	1	0
74	0	1	0	1	1	0
75	0	1	0	0	1	0
76	0	1	0	1	1	0
77	0	1	0	0	1	0
78	0	1	0	1	1	0
79	1	1	0	1	1	0
80	1	0	0	0	0	1
81	0	1	0	1	1	0
82	0	1	0	1	1	0
83	0	0	0	0	1	0
84	0	1	0	0	1	0
85	0	0	1	0	1	0
86	0	0	0	0	1	0
87	0	0	0	0	1	0
88	0	0	0	0	1	0
89	0	1	0	1	1	0
90	0	0	0	0	1	0
91	0	0	0	1	0	1
92	0	0	0	1	1	0
93	0	1	0	1	0	1
94	0	0	0	1	0	1
95	1	0	1	0	1	0
96	0	0	1	0	1	0
97	0	0	0	0	1	0
98	0	0	0	0	1	0
99	0	0	0	1	1	0
100	0	0	0	1	0	1
101	0	0	0	0	1	0
102	0	1	0	0	1	0
103	0	0	0	0	0	1
104	0	0	0	0	1	0
105	0	0	0	1	1	0
106	0	0	0	1	1	0
107	0	1	0	1	1	0
108	0	1	0	1	1	0
109	0	1	0	0	1	0

Based on the results in Table 4, the interestingness measures for each criterion is listed in Table 5, and the classification of these 109 objective interestingness measures is shown in Table 6.

Table 5. The interestingness measures of each criterion

Criterion	Interestingness measure
VAR.	3, 4, 6, 7, 9, 11, 13, 15, 16, 20, 22, 24, 26, 27, 29, 34, 35, 36, 42, 45, 46, 58, 63, 79, 80, 95
IND.	4, 7, 11, 14, 25, 26,33, 34, 35, 42, 45, 46, 47, 48, 49, 50, 51, 53, 54, 62, 64, 65, 69, 71, 74, 75, 76, 77, 78, 79, 81, 82, 84, 89, 93, 102, 107, 108, 109
EQU.	8, 19, 30, 31, 36, 38, 44, 59, 60, 85, 95, 96
SYM.	2, 3, 12, 14, 15, 26, 27, 29, 32, 33, 37, 45, 47, 50, 56, 57, 62, 65, 66, 67, 72, 74, 76, 78, 79, 81, 82, 89, 91, 92, 93, 94, 99, 100, 105, 106, 107, 108
DES.	1, 2, 3, 4, 5, 6,7, 8, 9, 10, 11, 13, 14, 15, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 36, 37, 38, 39, 40, 43, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 71, 72, 73, 74, 75, 76, 77, 78, 79, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 92, 95, 96, 97, 98, 99, 101, 102, 104, 105, 106, 107, 108, 109
STA.	12, 16, 32,33, 34, 35, 41, 42, 44, 59, 70, 80, 91, 93, 94, 100, 103

Table 6. The classification of interestingness measures

NATURE SUBJECT	Descriptive	Statistical
Equilibrium	8, 19, 30, 31, 36, 38, 60, 85, 95, 96	44, 59
Independence	4, 7, 11, 14, 25, 26, 45, 46, 47, 48, 49, 50, 51, 53, 54, 62, 64, 65, 69, 71, 74, 75, 76, 77, 78, 79, 81, 82, 84, 89, 102, 107, 108, 109	33, 34, 35, 42, 93
Others	1, 2, 3, 5, 6, 9, 10, 13, 15, 17, 18, 20, 21, 22, 23, 24, 27, 28, 29, 37, 39, 40, 43, 52, 55, 56, 57, 58, 61, 63, 66, 67, 68, 72, 73, 83, 86, 87, 88, 90, 92, 97, 98, 99, 101, 104, 105, 106	12, 16, 32, 41, 70, 80, 91, 94, 100, 103

Table 6 shows that the most of objective interestingness measures are descriptive measures; IPEE (44) and Laplace (59) are two statistical measures that calculate the interestingness values from the equilibrium position; Lerman similarity index (33), Entropic Implication Intensity 1(34), Entropic Implication Intensity 2 (35), Implication Intensity (42), and Rule Interest (93) are statistical measures that calculate the interestingness values from the independence position

The classification also gives a quick look at the mutual relationships among objective interestingness measures. This view is very useful to deeply understand on the formation of the clusters of interestingness measures when the clustering is influenced by the set of association rules. For example, measures influenced by the measure Confidence such as Confidence (19), Descriptive Confirmed-Confidence (30), Descriptive-Confirm (31), Example and counter-examples rate (36) belong to descriptive measures and tend to be varied from equilibrium position.

6. Conclusions

A lot of researchers in field KDD focus on ranking association rules by using the interestingness measures. Two types of interestingness measures studied in those researches are: subjective measures and objective measure. This article searched 109 objective interestingness measures which are discussed widely, transformed their formulae into a generic form using the cardinality $(n, n_X, n_Y, n_{X\bar{Y}})$, learned the evaluation criteria, and classified those interestingness measures based on 6 criterions. This classification is also evaluated closely to show the relationship among measures with common and particular characteristics.

Appendix.

Table 1. The formulae of interestingness measures calculated by using the cardinality $(n, n_X, n_Y, n_{X\bar{Y}})$

1. $$\frac{n_X - n_{X\bar{Y}}}{n_X}\log_2\frac{n(n_X - n_{X\bar{Y}})}{n_X n_Y}$$
2. $$\frac{n_X - n_{X\bar{Y}}}{n}\log_2\frac{n(n_X - n_{X\bar{Y}})}{n_X n_Y}$$
3. $$\frac{n + n_X - n_Y - 2n_{X\bar{Y}}}{n}$$
4. $$\frac{n_X - n_{X\bar{Y}}}{n_X} - \frac{n_Y}{n}$$
5. $$min(\frac{n_X - n_{X\bar{Y}}}{n_X}, \frac{n_X - n_{X\bar{Y}}}{n_Y})$$
6. $$\min(\frac{n_X - n_{X\bar{Y}}}{n_X}, \frac{n_X - n_{X\bar{Y}}}{n_Y})\left(\min(\frac{n_X - n_{X\bar{Y}}}{n_X}, \frac{n_X - n_{X\bar{Y}}}{n_Y}) - \frac{n - n_X + n_{X\bar{Y}}}{(n - n_X)}\right)$$
7. $$\frac{nn_X - n_X n_Y - nn_{X\bar{Y}} + n_Y n_{X\bar{Y}}}{n_Y n_{X\bar{Y}}}$$
8. $\propto \left(1 - \frac{n_X - n_{X\bar{Y}}}{n_X}\right)$ where α is a constant
9. $$1 - \frac{1}{2}\left(\frac{1}{n_X} + \frac{1}{n - n_Y}\right)n_{X\bar{Y}}$$
10. $$1 - \frac{1}{2}\left(\frac{3}{n_X} + \frac{1}{n - n_Y}\right)n_{X\bar{Y}}$$
11. $$1 - \frac{nn_{X\bar{Y}}}{n_X(n - n_Y)}$$
12. $$\frac{n(nn_X - nn_{X\bar{Y}} - n_X n_Y)^2}{n_X n_Y(n - n_X)(n - n_Y))}$$
13. $$\frac{(n_X - n_{X\bar{Y}})(n - n_X)}{n_X(n_Y - n_X + n_{X\bar{Y}})}$$
14. $$\frac{2(nn_X - n_X n_Y - nn_{X\bar{Y}})}{nn_X + nn_Y - 2n_X n_Y}$$
15. $$\frac{n_X - n_{X\bar{Y}}}{n_Y + n_{X\bar{Y}}}$$
16. $$\frac{(n_X - n_{X\bar{Y}})(n - n_Y - n_{X\bar{Y}})(n_X(n - n_Y) + n_Y(n - n_X))}{((n - n_X)(n - n_Y) + n_X n_Y)(n_Y - n_X + 2n_{X\bar{Y}})}$$
17. $$\frac{n_{X\bar{Y}}}{n - n_Y}$$
18. $$-\left(1 - \frac{n_{X\bar{Y}}}{n_X}\right)\log_2(n_X - n_{X\bar{Y}}) + \log_2 n_X - \frac{n_{X\bar{Y}}}{n_X}\log_2 n_{X\bar{Y}}$$
19. $$\frac{n_X - n_{X\bar{Y}}}{n_X}$$
20. $$\frac{n + n_X - n_Y - 4n_{X\bar{Y}}}{n}$$
21. $$\frac{n_Y + n_{X\bar{Y}} - n_X}{n - n_X}$$
22. $$\frac{(n_X - n_{X\bar{Y}})(n - n_X)}{n(n_X - n_{X\bar{Y}}) + n_X(n_Y - n_X)}$$
23. $$\frac{n_X - n_{X\bar{Y}}}{n_X}(\frac{n_X - n_{X\bar{Y}}}{n_X} - \frac{n_Y + n_{X\bar{Y}} - n_X}{n - n_X})$$
24. $$\frac{n_X - n_{X\bar{Y}}}{n_Y}$$
25. $$\frac{n_X(n - n_Y)}{nn_{X\bar{Y}}}$$
26. $$\frac{n(n_X - n_{X\bar{Y}}) - n_X n_Y}{\sqrt{n_X n_Y(n - n_X)(n - n_Y)}}$$

27. $$\frac{n_X - n_{X\bar{Y}}}{\sqrt{n_X n_Y}}$$

28. $$\frac{n_X}{n}$$

29. $$\frac{2(n_X - n_{X\bar{Y}})}{n_X + n_Y}$$

30. $$1 - \frac{2n_{X\bar{Y}}}{n_X}$$

31. $$\frac{n_X - 2n_{X\bar{Y}}}{n}$$

32. $$\left(n^2 n_X n_Y (n-n_X)(n-n_Y) \Big/ \left(\min\begin{pmatrix}\min(n_X, n-n_X),\\ \min(n_Y, n-n_Y)\end{pmatrix}^2 \min\begin{pmatrix}\max(n_X, n-n_X),\\ \max(n_Y, n-n_Y)\end{pmatrix}\right) \right)^{\alpha} \chi^2$$

33. $$\frac{n_X - \frac{n_X n_Y}{n} - n_{X\bar{Y}}}{\sqrt{\frac{n_X n_Y}{n}}}$$

34. $\sqrt{IIM\left(\left(1-H_{Y|X}^{\propto}\right)\left(1-H_{\bar{X}|\bar{Y}}^{\propto}\right)\right)^{\frac{1}{2\propto}}}$ with (α=1) and

$$H_{Y|X} = -\frac{n_X - n_{X\bar{Y}}}{n_X}\log_2\frac{n_X - n_{X\bar{Y}}}{n_X} - \frac{n_{X\bar{Y}}}{n_X}\log_2\frac{n_{X\bar{Y}}}{n_X}$$

$$H_{\bar{X}|\bar{Y}} = -\frac{n - n_Y - n_{X\bar{Y}}}{n - n_Y}\log_2\frac{n - n_Y - n_{X\bar{Y}}}{n - n_Y} - \frac{n_{X\bar{Y}}}{n - n_Y}\log_2\frac{n_{X\bar{Y}}}{n - n_Y}$$

where IIM is Inplication Intensity

35. $\sqrt{IIM\left(\left(1-H_{Y|X}^{\propto}\right)\left(1-H_{\bar{X}|\bar{Y}}^{\propto}\right)\right)^{\frac{1}{2\propto}}}$ with (α=2) and

$$H_{Y|X} = -\frac{n_X - n_{X\bar{Y}}}{n_X}\log_2\frac{n_X - n_{X\bar{Y}}}{n_X} - \frac{n_{X\bar{Y}}}{n_X}\log_2\frac{n_{X\bar{Y}}}{n_X}$$

$$H_{\bar{X}|\bar{Y}} = -\frac{n - n_Y - n_{X\bar{Y}}}{n - n_Y}\log_2\frac{n - n_Y - n_{X\bar{Y}}}{n - n_Y} - \frac{n_{X\bar{Y}}}{n - n_Y}\log_2\frac{n_{X\bar{Y}}}{n - n_Y}$$

where IIM is Inplication Intensity

36. $$\frac{n_X - 2n_{X\bar{Y}}}{n_X - n_{X\bar{Y}}}$$

37. $$\frac{n_X + n_Y}{n}$$

38. $$\frac{n_X(1-\theta) - n_{X\bar{Y}}}{n}$$

39. $$\frac{(n_X - n_{X\bar{Y}})^2 + n_{X\bar{Y}}{}^2}{n n_X} + \frac{(n_Y - n_X + n_{X\bar{Y}})^2 + (n - n_Y - n_{X\bar{Y}})^2}{n(n-n_X)} - \frac{n_Y{}^2 + (n-n_Y)^2}{n^2}$$

40. $\frac{\alpha}{\beta}$ where

$$\alpha = max\left(\frac{n_X - n_{X\bar{Y}}}{n}, \frac{n_{X\bar{Y}}}{n}\right) + max\left(\frac{n_Y - n_X + n_{X\bar{Y}}}{n}, \frac{n - n_Y - n_{X\bar{Y}}}{n}\right) + max\left(\frac{n_X - n_{X\bar{Y}}}{n}, \frac{n_Y - n_X + n_{X\bar{Y}}}{n}\right) + max\left(\frac{n_{X\bar{Y}}}{n}, \frac{n - n_Y - n_{X\bar{Y}}}{n}\right) - max\left(\frac{n_X}{n}, \frac{n - n_X}{n}\right) - max\left(\frac{n_Y}{n}, \frac{n - n_Y}{n}\right)$$

$$\beta = 2 - max\left(\frac{n_X}{n}, \frac{n - n_X}{n}\right) - \left(\frac{n_Y}{n}, \frac{n - n_Y}{n}\right)$$

41. $$\frac{n_{X\bar{Y}} - \frac{n_X(n-n_Y)}{n}}{\sqrt{\frac{n_X(n-n_Y)}{n}}}$$

42. $$1 - \sum_{k=\max(0, n_X - n_Y)}^{n_{X\bar{Y}}} \frac{C_{n_Y}^{n_X - k} C_{(n-n_Y)}^{k}}{C_n^{n_X}}$$

43. $$\left|1 - \frac{n}{n_X}\right|$$

44. $$1 - \frac{2}{2^{n_X}} \sum_{k=0}^{n_{X\bar{Y}}} C_{n_X}^{k}$$

45. $$\log_2 \frac{n(n_X - n_{X\bar{Y}})}{n_X n_Y}$$

46. $$\sqrt{DIR(X => Y) DIR(\bar{Y} => X)}$$

47. $\left(\left(\frac{n(n_X - n_{X\bar{Y}})}{n_X n_Y}\right)^l - 1\right)\left(\frac{n_X - n_{X\bar{Y}}}{n}\right)^m$ with l, m are weights.
Give l = m = 1

48. $$\begin{cases} 1 - \frac{n n_{X\bar{Y}}}{n_X(n-n_Y)} & if \ \frac{n - n_{X\bar{Y}}}{n_X} > \frac{n_Y}{n} \\ \frac{n(n_X - n_{X\bar{Y}})}{n_X n_Y} - 1 & otherwise \end{cases}$$

49. $$\frac{n_X - n_{X\bar{Y}}}{n}\log_2\frac{n(n_X - n_{X\bar{Y}})}{n_X n_Y} + \frac{n_{X\bar{Y}}}{n}\log_2\frac{n n_{X\bar{Y}}}{n_X(n-n_Y)}$$

50. $$\frac{n_X - n_{X\bar{Y}}}{n}\log\frac{n(n_X - n_{X\bar{Y}})}{n_X n_Y}$$

51. $$\frac{n_X - n_{X\bar{Y}}}{n_Y + n_{X\bar{Y}}}\left(\frac{n_X - n_{X\bar{Y}}}{n_Y + n_{X\bar{Y}}} - \frac{n_X(n_Y - n_X + n_{X\bar{Y}})}{n(n_X + n_Y) - n_X(2n_Y + n_{X\bar{Y}})}\right)$$

52. $$\frac{n_X(n_Y - n_X + n_{X\bar{Y}})}{n(n_X + n_Y) - n_X(2n_Y + n_{X\bar{Y}})}$$

53. $$\frac{nn_X - n_X n_Y - nn_{X\bar{Y}}}{n_X(n - n_Y) - n_{X\bar{Y}}(n - 2n_Y)}$$

54. $$\sqrt{\frac{n_X - n_{X\bar{Y}}}{n}\left(\frac{n - n_Y}{n} - \frac{n_{X\bar{Y}}}{n_X}\right)}$$

55. $$\left(\frac{n_X - n_{X\bar{Y}}}{n_X} - \frac{n - n_Y - n_{X\bar{Y}}}{n - n_X}\right)(\log_2(n - n_Y) - \log_2 n_Y)$$

56. $$\frac{n_X - n_{X\bar{Y}}}{n_Y - n_X + 2n_{X\bar{Y}}}$$

57. $$(n_X - n_{X\bar{Y}})\left(\frac{1}{n_X} + \frac{1}{n_Y}\right)$$

58. $$\frac{(n_X - n_{X\bar{Y}})}{2}\left(\frac{1}{n_X} + \frac{1}{n_Y}\right)$$

59. $$\frac{n_X - n_{X\bar{Y}} + 1}{n_X + 2}$$

60. $$\frac{n_X - 2n_{X\bar{Y}}}{n_Y}$$

61. $$1 - \frac{n_{X\bar{Y}}}{n_X} - \frac{n_X n_Y}{n^2}$$

62. $$\frac{n(n_X - n_{X\bar{Y}})}{n_X n_Y}$$

63. $$\frac{nn_Y - nn_X + nn_{X\bar{Y}} - {n_Y}^2 + n_X n_Y}{nn_Y - n_Y n_{X\bar{Y}} - {n_Y}^2}$$

64. $$\log\frac{(n - n_Y)(n_X - n_{X\bar{Y}})}{n_Y n_{X\bar{Y}}}$$

65. $$\log\frac{n(n_X - n_{X\bar{Y}})}{n_X n_Y}$$

66. $$\max\left(\frac{n_X - n_{X\bar{Y}}}{n_X}, \frac{n_X - n_{X\bar{Y}}}{n_Y}\right)$$

67. $$\frac{2(n_X - n_{X\bar{Y}})^2}{n(n_X + n_Y)}$$

68. $$\frac{n_X - n_{X\bar{Y}}}{n}\log_2\frac{n(n - n_{X\bar{Y}})}{n_X n_Y} + \frac{n_{X\bar{Y}}}{n}\log_2\frac{nn_{X\bar{Y}}}{n_X(n - n_Y)} + \frac{n_Y - n_X + n_{X\bar{Y}}}{n}\log_2\frac{n(n_Y - n_X + n_{X\bar{Y}})}{(n - n_X)n_Y} + \frac{n - n_Y - n_{X\bar{Y}}}{n}\log_2\frac{n(n - n_Y - n_{X\bar{Y}})}{(n - n_X)(n - n_Y)}$$

69. $$\frac{n(n_X - n_{X\bar{Y}}) - n_X n_Y}{n_Y(n - n_X)}$$

70. $$\frac{MI}{\frac{n_Y}{n}\log n_Y - \log n - \frac{n - n_Y}{n}\log(n - n_Y)}$$

71. $$\frac{n(n_X - n_{X\bar{Y}}) - n_X n_Y}{n_X(n - n_X)}$$

72. $$\frac{2(n_X - n_{X\bar{Y}})}{n_X + n_Y}$$

73. $$\frac{MI}{-\frac{n_X}{n}\log_2\frac{n_X}{n} - \frac{n - n_X}{n}\log_2\frac{n - n_X}{n}}$$

74. $$\frac{n(n_X - n_{X\bar{Y}}) - n_X n_Y}{\sqrt{n_X n_Y(n - n_X)(n - n_Y)}}\min\left(\frac{n_X - n_{X\bar{Y}}}{n_X}, \frac{n_X - n_{X\bar{Y}}}{n_Y}\right)$$

75. $$\frac{n(n_X - n_{X\bar{Y}}) - n_X n_Y}{\sqrt{n_X n_Y(n - n_X)(n - n_Y)}}\frac{(n_X - n_{X\bar{Y}})}{n_X}$$

76. $$\frac{n(n_X - n_{X\bar{Y}}) - n_X n_Y}{\sqrt{n_X n_Y(n - n_X)(n - n_Y)}}\frac{(n_X - n_{X\bar{Y}})}{n_Y + n_{X\bar{Y}}}$$

77. $$\frac{n(n_X - n_{X\bar{Y}}) - n_X n_Y}{\sqrt{n_X n_Y(n - n_X)(n - n_Y)}}\frac{n_X - n_{X\bar{Y}}}{n_Y + n_{X\bar{Y}}}\left(\frac{n_X - n_{X\bar{Y}}}{n_Y + n_{X\bar{Y}}} - \frac{n_X(n_Y - n_X + n_{X\bar{Y}})}{n(n_X + n_Y) - n_X(2n_Y + n_{X\bar{Y}})}\right)$$

78. $$\frac{n(n_X - n_{X\bar{Y}}) - n_X n_Y}{\sqrt{n_X n_Y(n - n_X)(n - n_Y)}}\frac{2(nn_X - n_X n_Y - nn_{X\bar{Y}})}{nn_X + nn_Y - 2n_X n_Y}$$

79. $$\frac{(n_X - n_{X\bar{Y}})(n - n_Y - n_{X\bar{Y}})}{n_{X\bar{Y}}(n_{X\bar{Y}} + n_Y - n_X)}$$

80. $$\frac{n_X(n - n_Y)^2}{nn_{X\bar{Y}}}$$

81. $$\frac{n(n_X - n_{X\bar{Y}}) - n_X n_Y}{n^2}$$

82. $$\log\left(\frac{n(n_X - n_{X\bar{Y}})}{n_X n_Y}\right)$$

83. $$\frac{n_Y}{n}$$

84. $$\frac{3}{2} + \frac{4n_X - 3n_Y}{2n} - \left(\frac{3}{2n_X} + \frac{2}{n - n_Y}\right)n_{X\bar{Y}}$$

85. $$\frac{2n_{X\bar{Y}}(n_X - n_{X\bar{Y}})}{{n_X}^2}$$

86. $$\begin{cases} 1 - \frac{n_X}{n} & if\ \frac{n - n_{X\bar{Y}}}{n_X} > \frac{n - n_{X\bar{Y}}}{n_Y} \\ 1 - \frac{n}{n_X} & otherwise \end{cases}$$

87. $$\log(2 - \frac{n_{X\bar{Y}}}{n_X})\ \log(1 + \frac{n_X}{n_Y})$$

88. $$\frac{n_{X\bar{Y}}}{n}$$

89. $$\frac{n(n_X - n_{X\bar{Y}})}{n_X n_Y} - 1$$

90. $$\frac{n_X - n_{X\bar{Y}}}{n_Y}$$

91. $$\frac{n_X - n_{X\bar{Y}}}{n}\left||n_{X\bar{Y}} + n_Y|\right| - \left(\frac{n_X - n_{X\bar{Y}}}{n} + \left||n_{X\bar{Y}} + n_Y|\right|\right)$$

92. $$\frac{n + n_X - n_Y - 2n_{X\bar{Y}}}{n - n_X + n_Y + 2n_{X\bar{Y}}}$$

93. $$\frac{n(n_X - n_{X\bar{Y}}) - n_X n_Y}{n}$$

94. $$\log\left(\frac{1}{\sqrt{1 + \frac{\min(n_{X\bar{Y}}, n_Y - n_X + n_{X\bar{Y}})}{1 + n_X - n_{X\bar{Y}}}}}\right)$$

95. $$\frac{n_X}{n_{X\bar{Y}}} - 1$$

96. $$\log_2 n_X - \frac{n_{X\bar{Y}}}{n_X}\log_2 n_{X\bar{Y}} - \frac{n_X - n_{X\bar{Y}}}{n_X}\log_2(n_X - n_{X\bar{Y}})$$

97. $$\frac{n - n_Y - n_{X\bar{Y}}}{n - n_X}$$

98. $$\frac{n - n_Y - n_{X\bar{Y}}}{n - n_Y}$$

99. $$\frac{n_X - n_{X\bar{Y}}}{n}$$

100. $$\frac{n_X + n_Y - nn_X + nn_{X\bar{Y}}}{n^2}$$

101. $$\frac{n_X(n_Y - n_X + n_{X\bar{Y}})}{n(n - n_X)}$$

102. $$\frac{\frac{n_X - n_{X\bar{Y}}}{n}}{\frac{n_X - n_{X\bar{Y}}}{n} + \frac{n_X(n_Y - n_X + n_{X\bar{Y}})}{n(n - n_X)}}$$

103. $\sqrt{U\ x\ S\ x\ R}$ where U: U Cost, S: S Cost and R: R Cost

104. $$\frac{MI}{-\frac{n_Y}{n}\log_2\frac{n_Y}{n} - \frac{n - n_Y}{n}\log_2\frac{n - n_Y}{n}}$$

105. $$\log\left(1 + \frac{\min(n_{X\bar{Y}}, n_Y - n_X + n_{X\bar{Y}}) + (n_X - n_{X\bar{Y}})}{\max(n_{X\bar{Y}}, n_Y - n_X + n_{X\bar{Y}}) + (n_X - n_{X\bar{Y}})}\right)$$

106. $$\frac{n_X - n_{X\bar{Y}}}{n} - \propto$$

107. $$\frac{nn_X - n_X n_Y - nn_{X\bar{Y}}}{nn_X - n_X n_Y - nn_{X\bar{Y}} - 2n_{X\bar{Y}}(n_X - n_Y - n_{X\bar{Y}})}$$

108. $\frac{1-\sqrt{k}}{1+\sqrt{k}}$ where $k = \frac{n_{X\bar{Y}}(n_Y - n_X + n_{X\bar{Y}})}{(n_X - n_{X\bar{Y}})(n - n_Y - n_{X\bar{Y}})}$

109. $$\frac{nn_X - n_X n_Y - nn_{X\bar{Y}}}{\max((n_X - n_{X\bar{Y}})(n - n_Y), n_Y n_{X\bar{Y}})}$$

Acknowledgements

This research is funded by UK - ASEAN Research Hub Project, VNUK Institute for Research and Executive Education, the University of Danang, Grant number 07/HĐ-UARH).

References

[1] Al-Hegami, AS., Bhatnagar, V., Kumar, N. (2004) “Novelty as a Measure of Interestingness in Knowledge Discovery”, *Data Warehousing and Knowledge Discovery*, Springer-Verlag Berlin Heidelberg, pp. 48-57.

[2] Anandhavalli, M., Ghose, M.K., Gauthaman, K. (2010) “Interestingness Measure for Mining Spatial Gene Expression Data using Association Rule”, *Journal of Computing*, **Volume** 2, Issue 1, pp. 110-114.

[3] Baena-Garcı´a, M., Morales-Bueno, R. (2012) “Mining interestingness measures for string pattern mining”, *Knowledge- Based Systems*, **Volume** 25, Issue 1, pp. 45-50, Elsevier.

[4] Bayardo, R. J. Jr., Agrawal, R. (1999) “Mining the most interestingness rules”, KDD'99 - Proceedings of the 5th ACM SIGKDD International Confeference on Knowledge Discovery and Data Mining, pp. 145-154.

[5] Belohlavek, R., Grissa, D., Guillaume, S., Nguifo, E.M., Outrata, J. (2014) “Boolean factors a s a means of clustering of interestingness measures of association rules”, *Annals of Mathematics and Artificial Intelligence*, **Volume** 7, Issue 1-2, Springer International Publishing, pp. 151-184.

[6] Benites, F., Sapozhnikova, E. (2014) “Evaluation of Hierarchical Interestingness Measures for Mining Pairwise

Generalized Association Rules", *IEEE Transactions on Knowledge & Data Engineering*, **Volume** 26, Issue 12, pp. 3012-3025.

[7] Blanchard, J., Guillet, F., Briand, H., Gras, R. (2005) "Assessing rule interestingness with a probabilistic measure of deviation from equilibrium", *Proceedings of the 11th international symposium on Applied Stochastic Models and Data Analysis ASMDA-2005*, pp. 191-200.

[8] Brijs, T., Vanhoof, K., Wets, G. (2003) "Defining interestingness for association rules", *International Journal "Information Theories & Applications"*, **Volume** 10, pp. 370-375.

[9] Carvalho, D.R., Freitas, A.A., Ebecken, N. (2005) "Evaluating the Correlation between Objective Rule Interestingness Measures and Real Human Interest", *Knowledge Discovery in Databases: PKDD 2005*, Springer-Verlag Berlin Heidelberg, pp. 453-461.

[10] Deora, C.S., Arora, S., Makani, Z. (2013) "Comparison of Interestingness Measures: Support-Confidence Framework versus Lift-Irule Framework", *International Journal of Engineering Research and Applications (IJERA)*, **Volume** 3, Issue 2, pp. 208-215.

[11] Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P. (1996) "From data mining to knowledge discovery", *Advances in Knowledge Discovery and Data Mining*, pp. 1-34.

[12] Geng, L., Hamilton, H. J. (2006) "Interestingness Measures for Data Mining: A Survey", *ACM Computing Surveys*, Volume 38, Issue 3, Article 9.

[13] Glass, D.H. (2013) "Confirmation measures of association rule interestingness", *Knowledge-Based Systems 24*, **Volume 44**, pp. 65-77.

[14] Grissa, D., Guillaume, S., Nguifo, E.M. (2012) "Combining Clustering techniques and FCA to characterize Interestingness Measures", Research Report LIMOS/RR-12-05.

[15] Guillaume, S., Grissa, D., Nguifo, E.M. (2012) "Categorization of interestingness measures for knowledge extraction", *CoRR* abs/1206.6741.

[16] Guillet, F., Hamilton, H.Z. (2007) *Quality Measures in Data Mining* - Series in Computational Intelligence (43), Springer-Verlag.

[17] Gyenesei, A., Teuhola, J. (2001) "Interestingness Measures for Fuzzy Association Rules", *Principles of Data Mining and Knowledge Discovery*, Springer-Verlag Berlin Heidelberg, pp. 152–164.

[18] H.X. Huynh, F. Guillet, T.Q. Le, H. Briand, "Ranking objective interestingness measures with sensitivity values", *VNU Journal of Science, Natural Sciences and Technology 24 (2008)*, pp. 122-132, 2008.

[19] Heravi, M.J., Zaïane, O.R. (2010) "A Study on Interestingness Measures for Associative Classifiers", *SAC'10- Proceedings of the 2010 ACM Symposium on Applied Computing*, pp. 1039-1046.

[20] Hilderman, R. J., Hamilton, H. J. (2001) *Knowledge Discovery and Measures of Interestingness*, Kluwer Academic Publishers.

[21] Hills, J., Davis, L. M., Bagnall, A. (2012) "Interestingness Measures for Fixed Consequent Rules", *Intelligent Data Engineering and Automated Learning - IDEAL 2012*, Springer-Verlag Berlin Heidelberg, pp. 68–75.

[22] Huynh, H. X., Guillet, F., Blanchard, J., Kuntz, P., Gras, R., Briand, H. (2007) "A graph-based clustering approach to evaluate interestingness measures: a tool and a comparative study (Chapter 2)", *Quality Measures in Data Mining*, Springer-Verlag, pp. 25-50.

[23] Huynh, H. X., Lam, N. C., Guillet, F. (2008) "On interestingness interaction", *RIVF'08 - IEEE International Conference on Research, Innovation and Vision for the Future (2)*, pp.161-166.

[24] Huynh, X.H., Guillet, F., Briand, H. (2006) "Extracting representative measures for the post-processing of association rules", Fourth IEEE International Conference on Computer Sciences dedicated to Research, Innovation and Vision for the Future, Ho Chi Minh, Vietnam: IEEE Computer Society Press, pp. 99–106.

[25] Huynh, X.H., Guillet, F., Briand, H. (2006) "A Graph-based Approach for Comparing Interestingness Measures", *Engineering of Intelligent Systems, 2006 IEEE International Conference*, pp. 1-6.

[26] Huynh, X.H., Guillet, F., Briand, H. (2006) "Evaluating Interestingness Measures with Linear Correlation Graph", *Advances in Applied Artificial Intelligence*, Springer-Verlag Berlin Heidelberg, pp. 312-321.

[27] Huynh, X.H., Guillet, F., Briand, H. (2005) "A Data Analysis Approach for Evaluating the Behavior of Interestingness Measures", *Discovery Science*, Springer-Verlag Berlin Heidelberg, pp. 330-337.

[28] I.N.M. Shaharanee and F. Hadzic and T.S. Dillon, "Interestingness measures for associattion rules based on statistical validity", *Knowledge-Based Systems 24*, pp.386-392, 2011.

[29] Immanuel, T., Dwarakanath, B. (2013) "Mine Frequent Patterns using Valid Measures", *International Journal of Research in Engineering & Advanced Technology*, **Volume** 1, Issue 1.

[30] Jaiswal, V., Agarwal, J. (2012) "The Evolution of the Association Rules", *International Journal of Modelling and Optimization*, **Volume** 2, Issue 6, pp. 726-729.

[31] Jalalvand, A., Minaei, B., Atabaki, G., Jalalvand, S. (2008) "A new interestingness measure for associative rules based on the geometric context", *The Third 2008 International Conference on Convergence and Hybrid Information Technology*, pp.199-203.

[32] K. McGarry, "A survey of interestingness measures for knowledge discovery", *Knowledge Engineering Review Journal 20(1)*, pp. 39-61, 2005.

[33] Lallich, S., Vaillant, B., Lenca, P. (2005) "Parametrised measures for the evaluation of association rule interestingnes", *ASMDA'05 - Proceedings of the 11th International Symposium on Applied Stochastic Models and Data Analysis*, pp. 220-229.

[34] Lallich, S., Teytaud, O. (2004) "Evaluation et validation de l'intérêt des règles d'association", *Mesures de Qualité pour la Fouille de Données, RNTI-E-1*, pp. 193-217.

[35] Le, T.T.N., Huynh, H.X., Guillet, F. (2009) "Finding the Most Interesting Association Rules by Aggregating Objective Interestingness Measures", *Knowledge Acquisition: Approaches, Algorithms and Applications*, Springer-Verlag Berlin Heidelberg, pp. 40-49.

[36] Le Bras, Y., Lenca, P., Lallich, S. (2012) "Optimonotone Measures for Optimal Rule Discovery", *Computational Intelligence*, **Volume** 28, Issue 4, pp. 475-504.

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

[38] Lenca, P., Vaillant, B., Meyer, P., Lallich, S. (2007) "Association rule interestingness measures: experimental and theoretical studies", *Quality Measures in Data Mining,* Springer-Verlag Berlin Heidelberg, pp. 51-76.

[39] Lingyun, W., Sheng, W. (2013) “Research on Improvement of Objective Interestingness Measures for Association Rules”, Mechatronic Sciences, Electric Engineering and Computer (MEC), Proceedings 2013 International Conference, pp. 2642-2647, IEEE.

[40] Liu, B., Hsu, W., Mun, L.-F., Lee, H.-Y. (1999) “Finding interesting patterns using user expectations”, *IEEE Transactions on Knowledge and Data Engineering 11(6)*, pp. 817-832.

[41] Liu, J., Fan, X., Qu, Z. (2008) “A New Interestingness Measure of Association Rules”, *Genetic and Evolutionary Computing, 2008. WGEC '08. Second International Conference*, pp. 393-397, IEEE.

[42] Maddouri, M., Gammoudi, J. (2007) “On Semantic Properties of Interestingness Measures for Extracting Rules from Data”, *Adaptive and Natural Computing Algorithms,* Springer-Verlag Berlin Heidelberg, pp. 148–158.

[43] Nguyen, L.T.T., Vo, B., Hong, T.-P., Thanh, H.C. (2012) “Interestingness Measures for Classification Based on Association Rules”, *Computational Collective Intelligence - Technologies and Applications*, Springer-Verlag Berlin Heidelberg, pp. 383-392.

[44] Omiecinski, A. R. (2003) “Alternative interest measures for mining associations in databases”, *IEEE Transactions on Knowledge and Data Engineering 15(1)*, pp. 57-96.

[45] P.-N. Tan and V. Kumar and J. Srivastava, “Selecting the right objective measure for association analysis”, *Information Systems 29(4)*, pp. 293-313, 2004.

[46] Piatetsky-Shapiro, G., Matheus, C. J. (1994) “The interestingness of deviations”, *AAAI'94 - Knowledge Discovery in Databases Workshop*, pp. 25-36.

[47] Sahar S., Mansour, Y. (1999) “Empirical evaluation of interest-level criteria”, *SPIE'99 - Proceedings of SPIE - DMKD: Theory, Tools and Technology (3695)*, pp. 63-74.

[48] Schaul, T., Pape, L., Glasmachers, T. (2011) “Coherence Progress: A Measure of Interestingness Based on Fixed Compressors”, *Artificial General Intelligence*, Springer-Verlag Berlin Heidelberg, pp. 21-30.

[49] Selvarangam, K., Ramesh Kumar, K. (2014) “Selecting perfect interestingness measures by coefficient of variation based ranking algorithm”, *Journal of Computer Science*, **Volume** 10, Issue 9, pp. 1672-1679, Science Publications.

[50] Shaw, G., Xu, Y., Geva, S. (2014) “Interestingness Measures for Multi-Level Association Rules”, *Innovations in Intelligent Machines-4*, Springer International Publishing, pp. 47-74.

[51] Silberschatz, A., Tuzhilin, A. (1995) “On subjective measures of interestingness in knowledge discovery”, KDD'95 - *Proceedings of the First International Conference on Knowledge Discovery and Data Mining*, pp. 275-281.

[52] Silberschatz, A., Tuzhilin, A. (1996) “What makes patterns interesting in knowledge discovery systems”, *IEEE Transactions on Knowledge and Data Engineering 8(6)*, pp. 970-974.

[53] Tamir, R., Singer, Y. (2006) “On a confidence gain measure for association rule discovery and scoring”, *The International Journal on Very Large Data Bases 15(1)*, pp. 40-52.

[54] Tan, P.-N., Kumar, V. (2000) Interestingness measures for association patterns: a perspective, Tech. Rep. 00-036, Department of Computer Science, University of Minnesota.

[55] Tew, C., Giraud-Carrier, C., Tanner, K., Burton, S. (2014) “Behavior-based clustering and analysis of interestingness measures for association rule mining”, *Data Mining and Knowledge Discovery*, **Volume** 28, Issue 4, Springer US, pp. 1004-1045.

[56] Zhang, Y., Zhang, L., Nie, G., Shi, Y. (2009) “A Survey of Interestingness Measures for Association Rules”, *Business Intelligence and Financial Engineering 2009 - BIFE '09*. International Conference, pp. 460-463, IEEE.