

# A Critical Review of Rule Surprisingness Measures

D. R. Carvalho<sup>(1,3)</sup>, A. A. Freitas<sup>(2)</sup> & N. F. F. Ebecken<sup>(1)</sup>

<sup>1</sup> *COPPE / Universidade Federal do Rio de Janeiro, Brasil.*

<sup>2</sup> *University of Kent at Canterbury, UK.*

<sup>3</sup> *Universidade Tuiuti do Paraná, Brasil.*

## Abstract

In data mining it is usually desirable that discovered knowledge have some characteristics such as being as accurate as possible, comprehensible and surprising to the user. The vast majority of data mining algorithms produce, as part of their results, information of a statistical nature that allows the user to assess how accurate and reliable the discovered knowledge is. However, in many cases this is not enough for the user. Even if discovered knowledge is highly accurate from a statistical point of view, it might not be interesting for the user. Few data mining algorithms produce, as part of their results, a measure of the degree of surprisingness of discovered knowledge. However, these measures can be computed in a post-processing phase, as a form of additional evaluation of the quality of discovered knowledge, complementing (rather than replacing) statistical measures of discovered knowledge accuracy. This paper presents a review of four measures of classification-rule surprisingness, discussing their main characteristics, advantages and disadvantages. Hence, the main contribution of this paper is to improve our understanding of these rule surprisingness measures, which is a step towards solving the very difficult problem of selecting the “best” rule surprisingness measure for a given application domain.

**Keywords** – classification, rule surprisingness, rule interestingness

## 1 Introduction

The process of knowledge discovery in databases (KDD) aims at discovering knowledge that is interesting and useful to the user. Some desirable properties of discovered knowledge are: accuracy, comprehensibility and surprisingness (unexpectedness) (Han & Kamber, 2001).

The vast majority of data mining algorithms produce, as part of their results, information of a statistical nature that allows the user to assess how accurate and reliable the discovered knowledge is. However, in many cases this is not enough

for the user. Even if discovered knowledge is highly accurate from a statistical point of view, it might not be interesting for the user.

For instance, a set of discovered rules might be too large to be analyzed by a user, or contain too much redundancy, which would be assigned a low mark in the comprehensibility criterion of discovered knowledge.

Furthermore, it should be noted that the discovery of surprising, unexpected knowledge is even more difficult than the discovery of accurate and comprehensible knowledge, and the two latter criteria do not imply the former. As a simple example of this point, consider the following hypothetical rule that could be discovered from a medical database: IF (patient is pregnant) THEN (gender is female). This rule is highly accurate and comprehensible, but it is useless because it represents an obvious, previously-known pattern.

Few data mining algorithms produce, as part of their results, a measure of the degree of surprisingness of discovered knowledge. However, these measures can be computed in a post-processing phase, as a form of additional evaluation of the quality of discovered knowledge, complementing (rather than replacing) statistical measures of discovered knowledge accuracy.

This paper presents a review of four measures of classification-rule surprisingness, discussing their main characteristics, advantages and disadvantages. Hence, the main contribution of this paper is to improve our understanding of these rule surprisingness measures, as a step towards solving the very difficult problem of selecting the “best” rule surprisingness measure for a given application domain. This contribution is important because works dedicated specifically to discuss and compare a number of rule “interestingness” measures usually focus on statistical, accuracy-oriented measures, rather than focusing on measures involving the notion of surprisingness to the user – see e.g. (Hilderman & Hamilton, 2001), (Tan et al., 2002).

The remainder of this paper is organized as follows. Section 2 reviews four rule surprisingness measures proposed in the literature. Section 3 presents a comparative discussion of those measures, evaluating them according to several criteria. Finally, section 4 concludes the paper.

## **2 Rule Surprisingness Measures**

There are several surprisingness measures proposed in the literature. In general these measures can be divided into two broad groups, namely user-driven and data-driven measures (Silberchatz & Tuzhilin, 1996), (Freitas, 1999). The basic idea of user-driven measures is that the user specifies his/her believes or previous knowledge about the application domain, and then a rule is considered surprising to the extent that it captures knowledge that is unexpected with respect to the user’s believes or previous knowledge. In contrast, data-driven measures try to estimate how surprising a rule will be to the user in a more automatic and indirect fashion, without requiring the user to specify his/her believes or previous knowledge.

A note on terminology is appropriated here. Most of the literature uses the terms subjective and objective measures, rather than user-driven and data-driven

measures. However, we prefer to use the latter terminology, because the term “subjective measure” is somewhat deceiving. The believes of the user are certainly subjective, but those believes are just an input for the rule surprisingness measure. The rule surprisingness measure ultimately consists of a mathematical formula that will assign a numerical degree of surprisingness to the rule. This degree is usually computed in an objective manner, using a well-defined formula. Hence, the term user-driven seems more appropriate.

User-driven measures have the advantage of directly taking into account the user’s believes or previous knowledge, but have the disadvantages of being strongly domain-dependent and less automatic, requiring a significant effort from the user to make his/her believes explicit. Actually, they are not only domain-dependent, but also user-dependent, since within the same application domain two or more users might have very different believes or previous knowledge. Data-driven measures have the disadvantage of being an indirect estimate of how surprising the rule will be to the user, but have the advantages of being more domain-independent and more automatic, relieving the user from the responsibility of making his/her believes explicit, which could be difficult or very time-consuming for the user. Hence, intuitively user-driven measures are particularly indicated when a specific user is available and has enough time and expertise to produce a good specification of his/her believes and previous knowledge; whereas data-driven measures are particularly indicated when there are several potential users or when the user(s) do not have the time or the expertise necessary to produce a good specification of his/her believes and previous knowledge. In any case, the two kinds of measure are not mutually exclusive, i.e., one can use both in a given system.

This paper reviews data-driven rule surprisingness measures only, since this study is not focused on any particular application domain or user. Hence, we avoid the difficult issues associated with the specification of believes or previous knowledge in the user-driven approach, and focus on comparing the main characteristics of several data-driven rule surprisingness measures. All the measures discussed in this paper are based on the knowledge representation of IF-THEN rules, which has the advantage of being intuitively comprehensible for the user – an important criterion in data mining (Han & Kamber, 2001).

### **2.1 Rule surprisingness based on exception rules and information change**

(Hussain et. al., 2000) proposes a method that identifies, in a set of discovered rules, a subset consisting of exception rules. Figure 1 shows the general structure of exception rules, with respect to a common sense rule and a reference rule. In this figure, A and B are non-empty sets of attribute-value pairs, and X is the class predicted by the rule. The symbol “ $\neg$ ” denotes logical negation. Note that an exception rule is a specialization of a “common sense” rule, and the exception rule predicts a class different from its corresponding (generalized) common sense rule. The method assumes that common sense rules represent patterns that are probably known by the user – since they have a large coverage (number of examples covered by the rule). On the other hand, users are considerably less likely to know the patterns represented by exception rules, since they have a low

coverage. Hence, exception rules tend to be surprising, because they tend to represent novel patterns that contradict the common sense rules. Note that, in order for an exception rule to be considered interesting, both the common sense rule and the exception rule must have a high classification accuracy. Note also that the reference rule helps to explain the cause of the exception rule.

We emphasize that the rule surprisingness measure that we discuss in this section is not the same as proposed by (Hussain et al. 2000), but rather a variation of that proposal adapted for the context of this paper, as follows. First, (Hussain et al. 2000) do not make a clear distinction between association and classification rules. On one hand their method discovers rules predicting a single class attribute, like classification rules. On the other hand it refers to association-rule terminology such as support and confidence, it involves pre-specified minimum support thresholds, and it does *not* try to discover a set of rules that could be used to classify any unseen example in the test set – which are characteristics of association rules. Considering that association rules have a semantics quite different from classification rules – see (Freitas, 2000) for details – and that the focus of this paper is on classification rules, we have simplified the rule surprisingness measure proposed by (Hussain et al. 2000). In essence, the authors’ proposed surprisingness measure takes into account both support and confidence, whereas the simplified surprisingness measure discussed here takes into account only confidence. We also avoid the issue of minimum support and confidence thresholds, not only because this is related to association (rather than classification) rules but also because this is related to the algorithm to discover the rules, which is beyond the scope of the paper. We are interested only in measuring the degree of surprisingness of rules, regardless of how this measure will be used by a given data mining algorithm.

---

$A \rightarrow X$  (common sense rule) (high coverage, high accuracy)  
 $A, B \rightarrow \neg X$  (exception rule) (low coverage, high accuracy)  
 $B \rightarrow \neg X$  (reference rule) (low coverage and/or low accuracy)

---

Figure 1: Structure of exception rules

The measure of surprisingness of an exception rule is based on calculating the amount of change in information relative to common sense rules. More precisely, consider a rule of the form  $AB \rightarrow X$ . The method calculates the difference in the amount of information (number of bits) associated with the description of this rule, denoted  $I^{AB_0}$ , and the amount of information associated with the description of the two rules  $A \rightarrow X$  and  $B \rightarrow X$ , denoted  $I^{AB_1}$ . In other words,  $I^{AB_0}$  denotes the number of bits required to describe the specific rule  $AB \rightarrow X$  in the absence of knowledge represented by the generalized rules  $A \rightarrow X$  and  $B \rightarrow X$ , whereas  $I^{AB_1}$  is the corresponding number of bits when the relationship between  $X$  and  $AB$  is rather described by the two rules  $A \rightarrow X$  and  $B \rightarrow X$ . Mathematically, the

measure of surprisingness of the rule  $AB \rightarrow X$  with respect to the rules  $A \rightarrow X$  and  $B \rightarrow X$ , denoted by  $RI^{AB}$ , is given by – see (Hussain et al. 2000) for details:

$$\text{InfoChange} = I^{AB1} - I^{AB0}$$

$$I^{AB0} = (-\Pr(X|AB) \log_2 \Pr(X|AB) + (-\Pr(\neg X|AB) \log_2 \Pr(\neg X|AB)))$$

$$I^{AB1} = -\Pr(X|AB) [\log_2 \Pr(X|A) + \log_2 \Pr(X|B)] - \Pr(\neg X|AB) [\log_2 \Pr(\neg X|A) + \log_2 \Pr(\neg X|B)]$$

One limitation of this measure, which is inherent to its design, is that it addresses a specific kind of surprising rule, namely an exception rule with respect to a common sense rule. Hence, in an extreme case, if there is no pair of common sense rule and exception rule, no surprising rule will be reported to the user, even though other kinds of surprising rules might be hidden in the data.

## 2.2 Rule surprisingness based on multiple minimum generalizations

(Freitas, 1998) proposed a measure of the degree of surprisingness of a rule that is based on considering several generalizations of that rule and counting how many of those generalized rules predict a class different from the original rule. This measure was originally proposed in the context of small disjuncts, i.e. rules covering a small number of examples, but the issue of whether or not a rule is a small disjunct is not relevant for the context of this paper. Hence, in this paper we will refer to the rule whose surprisingness we want to measure simply as a “specific” rule. Given a specific rule  $r$ , the method first computes the “minimum generalizations” of  $r$ . Let  $m$  be the number of conditions (attribute-value pairs) in the antecedent of rule  $r$ . Then rule  $r$  has  $k$  minimum generalizations, each of them a generalized rule with  $k - 1$  conditions. The  $k$ -th minimum generalization of  $r$ ,  $k=1, \dots, m$ , is obtained by removing the  $k$ -th condition from the specific rule.

(Actually, this procedure for obtaining minimum generalizations is a somewhat simplified version of the procedure described in (Freitas, 1998). In that reference it is made a distinction between conditions with a categorical attribute and conditions with a continuous attribute (such as  $\text{Age} > 25$ ). The latter can be generalized not only by removing the condition but also by modifying the “cut point” in the inequality associated with the condition. However, this is a detail that is not essential to the discussion of the rule surprisingness measure per se, so we abstracted away this detail in our discussion in the following.)

Note that each of the  $m$  generalized rules produced by this procedure cover a superset of the examples covered by the original, specific rule  $r$ . As a result, the distribution of classes in the set of examples covered by each generalized rule can be significantly different from the distribution of classes in the rule  $r$ . Hence, the rule consequent (predicted class) is re-computed for each generalized rule, i.e., each of those  $m$  generalized rules will predict the most frequent class in its set of examples. The rule surprisingness measure can now be defined as follows.

Let  $C$  be the class predicted by the original specific rule  $r$  and  $C_k$  be the class predict by the rule produced by the  $k$ -th minimum generalization. The system compares  $C$  with each  $C_k$  and computes  $N$ , the number of times where  $C$  is different from  $C_k$ . The number  $N$ , in the range  $0 \dots m$ , could be defined as the degree of surprisingness of rule  $r$  – the larger  $N$ , the more surprising rule  $r$  is, in the sense of predicting a class different from its minimum generalizations.

However, that measure would be biased to favour very long rules (with many conditions), i.e., the value of the measure would tend to grow with the value of  $m$ . In order to avoid a potential confusion between the issues of rule length and rule surprisingness, measuring the latter in a way as independent as possible from the former, we use the following normalized version of the rule surprisingness measure, denoted *MinGen*:

$$MinGen = N / m .$$

The larger the value of this measure, the more surprising the rule is.

One disadvantage of this rule surprisingness measure is its relatively high computational cost. Note that, for each specific rule being evaluated, the system needs to compute  $m$  generalized rules. It is quite possible that many or most of those  $m$  rules will not have been generated by the system before, i.e., they have to be generated only for the purpose of measuring the degree of surprisingness of the original specific rule. In this case the computation of the class to be assigned to the consequent of each of those generalized rules will involve scanning the training set to identify the coverage of that generalized rule and its corresponding majority class.

### 2.3 Surprisingness at the level of individual attributes

(Freitas, 1998) proposed another rule surprisingness measure called *AttSurp* (Attribute Surprisingness), based on the degree of surprisingness associated with the individual attributes occurring in a rule antecedent. The basic idea is that the degree of surprisingness of an attribute is estimated as the inverse of its information gain – see (Mitchell, 1997) for a review of the concept of information gain. The rationale for this measure is that the occurrence of an attribute with a high information gain in a rule will not tend to be surprising to the user, since users often know the most relevant attributes for the classification task at hand. On the other hand, the occurrence of an attribute with a low information gain in a rule tends to be more surprising, because this kind of attribute is usually considered little relevant for classification purposes. Note that, although an attribute can have a low information gain individually, it is possible that, when combined with other attributes in the rule antecedent, 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 as:

$$AttSurp = 1 / \sum_{i=1}^K InformationGain(A_i) / K$$

where  $InformationGain(A_i)$  is the information gain of the  $i$ -th attribute in the rule antecedent and  $K$  is the number of attributes in the rule antecedent. In this formula the value of *AttSurp* can be very large when the information gain values are very low, which makes it difficult to compare the value of this formula with other rule surprisingness measures. However, the original formula was later normalized to return values in the range 0..1 (Noda et al., 1999), as follows:

$$AttSurp = 1 - \left( \frac{\sum_{i=1}^K InfoGain(A_i)}{K \log_2(Number\_of\_classes)} \right)$$

It is well-known that the information gain measure has a bias favouring attributes with many values. As the *AttSurp* measure favours attributes with a small information gain, *AttSurp* has a bias favouring attributes with few values.

This measure has the advantage of being very generic. It can be applied to virtually any classification rule. It does not require that the current rule be an exception rule, does not require the existence of other rules more generic than the current rule, etc. Another advantage of this measure is that it can be computed very fast. Actually, the measure only requires the computation of the information gain of each attribute in a *preprocessing* step. When a given rule is being evaluated, the system simply computes the mean of the precomputed information gain values for the attributes occurring in the rule, which does not require any access to the data set.

One potential disadvantage of this measure is that it requires a careful handling of the trade-off between accuracy and surprisingness. Since it favours attributes with a low information gain, it tends to favour rules where accuracy is not so large. Obviously, this measure should not be used alone to evaluate rule quality. It is essential that it be used together with another rule quality criterion favouring more accurate rules. One possibility is to use *AttSurp* just in a post-processing step, in order to select the most surprising rules among all the discovered (and presumably accurate) rules. Another possibility is to use it during the search for rules, using it as one of terms in a weighted formula where at least one of the other terms is a measure of rule accuracy (Noda et al., 1999).

#### 2.4 Rule surprisingness based on exception rules and intensity of implication

(Suzuki & Kodratoff, 1998) proposed a method that searches for surprising rules based, at a high level of abstraction, in the same framework of finding exception rules with respect to a common sense rule and a reference as discussed in section 2.1 in regard to the work of (Hussain et al. 2000). However, there is a great difference between these two references with respect to how the degree of surprisingness of a rule is computed. As discussed earlier, (Hussain et al., 2000) use an information-theoretic measure. In contrast, (Suzuki & Kodratoff, 1998) proposed the use of the intensity of implication measure, as discussed next.

First, let us review the notation used by (Suzuki & Kodratoff, 1998), which will also be used here. Let  $A_\mu \rightarrow c$  be a common sense rule,  $A_\mu \wedge B_\nu \rightarrow c'$  be an exception rule and  $B_\nu \rightarrow c'$  be a reference rule, where  $A_\mu \equiv a_1 \wedge a_2 \wedge \dots \wedge a_\mu$  and  $B_\nu \equiv b_1 \wedge b_2 \wedge \dots \wedge b_\nu$ .

In essence, the intensity of implication represents the degree of surprisingness that a rule  $A_\mu \rightarrow c'$  has so few counter-examples. Let  $U$  and  $V$  be

sets of examples randomly-selected from the data being mined with the restriction that the set cardinalities  $|U|$  and  $|V|$  are equal to the set cardinalities  $|c|$  and  $|A_\mu|$ , respectively, that is:  $|U| = |c|$ ,  $|V| = |A_\mu|$ . The intensity of implication  $I$  for the rule  $A_\mu \rightarrow c'$  is:

$$IntImp = 1 - \Pr(|VU'| \leq |Ac'|)$$

Note that in this formula the term  $|VU'|$  represents the number of counter-examples (false positives) of the randomly-generated rule  $V \rightarrow U$ , so that the term  $\Pr(|VU'| \leq |Ac'|)$  is essentially measuring the probability that the rule  $A_\mu \rightarrow c'$  has at least as many counter-examples as a randomly-generated rule. The smaller the value of this probability, the higher the value of  $I$ , and the better the rule. The value of this probability is computed assuming that  $U$  and  $V$  are independent, so that the random variable  $|VU'|$  follows the hypergeometric law – see (Suzuki & Kodratoff, 1998) for details.

Some desirable properties of the intensity of implication are as follows. First, it takes into account the size of the data set in order to evaluate the reliability of a rule: it increases as the data set size increases. Note that a conventional conditional probability measure –  $\Pr(c|A_\mu)$  – does not have this property. Second, it increases as  $|c|$  decreases, favouring rarer classes, whose prediction tend to be more interesting to the user than the prediction of the majority class.

In order to apply this measure to exception rules, recall that the examples covered by the antecedent of a common sense rule can be considered as the “universe” for the exception rule (since the latter is a specialization of the former). Hence, in the above formula  $A_\mu$  and  $c$  are replaced by  $A_\mu B_v$  and  $A_\mu c'$ , so that the intensity of implication of the rule  $A_\mu \wedge B_v \rightarrow c'$  is given by:

$$IntImp = 1 - \Pr(|VU'| \leq |A_\mu B_v c'|)$$

### 3 A Comparative Study of Rule Surprisingness Measures

Table 1 summarizes the main characteristics of the rule surprisingness measures discussed in this paper, in order to better identify their advantages and disadvantages. The characteristics included in this table are:

**Computational Cost:** can be *low*, *medium* or *high*. A rule surprisingness measure is classified as a *low cost* one if the computation of that measure does not require an access to the training set for every rule being evaluated. This is the case only for the *AttSurp* measure, as discussed earlier. Indeed, that measure requires only a single scan of the data set in a preprocessing phase, whose result (information gain for each attribute) is simply re-used in any rule to be evaluated. MinGen has a high computational cost, since it involves potentially several scans of the training set, as explained in section 2.2.

**Cardinal dilatation:** indicates whether or not the value of the rule surprisingness measure is altered with a cardinal dilatation of the data set – i.e., an increase in the size of the data set with all other things (particularly the probabilities of rule antecedent, rule consequent and rule consequent given rule antecedent) being equal.

**Rule Interaction:** indicates whether or not the evaluation of a rule takes into account the interaction between that rule and other rules. For instance, the



*AttSurp* measure involves no rule interaction, but the other three measures discussed in this paper involve rule interaction. Note that the general structure of rule interaction is quite similar in the measures proposed by (Hussain et al., 2000) and (Suzuki & Kodratoff, 1998), since in both projects there is an interaction between a common sense rule and an exception rule and both projects require that the common sense rule be an accurate rule. In contrast, the measure proposed by (Freitas, 1998) involves a more complex interaction between a rule and all of its minimum generalizations. In particular, many minimum generalizations might be generated only for the specific purpose of evaluating the original specific rule, and there is no quality requirement (such as a high accuracy) associated with the generalized rules.

**Granularity of the evaluation:** indicates the level of granularity of the rule evaluation, classifying it into one of three possible cases: rule antecedent (i.e., the rule surprisingness measure considers the antecedent as a whole, rather than its individual conditions), condition (i.e., the measure considers individual conditions – attribute-value pairs – in the rule antecedent), or attribute (i.e., the measure considers individual attributes in the rule antecedent).

**Range of values:** indicates the range of values that can be taken on by the rule surprisingness measure.

Table 1 – Comparison of four rule surprisingness measures

Characteristics	<i>InfoChange</i>	<i>MinGen</i>	<i>AttSurp</i>	<i>IntImp</i>
Computational cost	Medium	High	Low	Medium
Cardinal Dilatation	Independent	Independent	Independent	Dependent
Rule interaction	Yes	Yes	No	Yes
Granularity of evaluation	condition	condition	attribute	condition
Range of values	-1 ... 1	0 .. 1	0 ...1	>0 ... +∞

## 4 Conclusion and Future Research

This paper discussed four measures of classification-rule surprisingness. It is well-known that no classification algorithm is universally best across all data sets. The same principle applies to rule surprisingness measures. There is no rule surprisingness measure which is universally best across all application domains. Actually, finding the “best” rule surprisingness measure for a given application domain is even more difficult than finding the “best” algorithm (in the sense of maximizing classification accuracy) for a given data set.

The reason is that, in the latter case, we can at least use an objective, purely data-driven measure of performance, such as classification accuracy on the test

set. In the case of rule surprisingness measures, the definition of “best” is problematic, since the notion of surprisingness is ultimately a subjective, user-dependent one. However, there are several data-driven, objective measures of rule surprisingness, such as the ones discussed in this paper, and there is a motivation for their use, as discussed in section 2.

Hence, given the great difficulty of finding the best surprisingness measure for a given application domain, this paper aimed at contributing to a better understanding of the advantages and disadvantages of four of those measures.

Future work will involve computational experiments measuring the value of each of the rule surprisingness measures discussed in this paper in several data sets. This will produce an empirical comparison of the pros and cons of these measures, complementing the analytical comparison presented in this paper.

## References

- Freitas, A.A. (1998) On objective measures of rule surprisingness. *Principles of Data Mining & Knowledge Discovery* (Proc. 2nd European Symp., PKDD'98. Nantes, France, Sep. 1998). *LNAI 1510*, 1-9. Springer-Verlag.
- Freitas, A.A. (1999). On Rule Interestingness Measures. *Knowledge – Based Systems Journal* 12 (5-6), p. 309-315.
- Freitas, A.A. (2000) Understanding the crucial differences between classification and discovery of association rules - a position paper. *ACM SIGKDD Explorations*, 2(1), 65-69. ACM.
- Hilderman, R.J. and Hamilton (2001) H.J. *Knowledge Discovery and Measures of Interest*. Kluwer.
- Han, J. & Kamber (2001) M. *Data Mining: concepts and techniques*. Morgan Kaufmann.
- Hussain, F.; Liu, H.; Lu, H. (2000). Exception Rule Mining with a Relative Interestingness Measure. *PAKDD-2000, LNAI 1805*, p. 86-96.
- Mitchell, T. M. (1997) *Machine Learning*, MacGraw-Hill, USA.
- 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.
- Silberschatz, A. & Tuzhilin, A. (1996) What makes patterns interesting in knowledge discovery systems. *IEEE Trans. Knowledge & Data Eng.* 8(6).
- Suzuki, E.; Kodratoff, Y. (1998) Discovery of Surprising Exception Rules Based on Intensity of Implication. *Proc. PKDD-1998*.
- Tan, P.N.; Kumar, V. And Srivastava, J. (2002) Selecting the right interestingness measure for association patterns. *Proc. ACM SIGKDD-2002*.