Post–Mining of Association Rules:
Techniques for Effective Knowledge Extraction

Yanchang Zhao
University of Technology, Sydney, Australia

Chengqi Zhang
University of Technology, Sydney, Australia

Longbing Cao
University of Technology, Sydney, Australia

Information Science REFERENCE
INFORMATION SCIENCE REFERENCE
Hershey · New York

Section II
Identifying Interesting Rules

Chapter II
From Change Mining to Relevance Feedback: A Unified View on Assessing Rule Interestingness

Mirko Boettcher
University of Magdeburg, Germany

Georg Ruß
University of Magdeburg, Germany

Detlef Nauck
BT Group plc, UK

Rudolf Kruse
University of Magdeburg, Germany

ABSTRACT

Association rule mining typically produces large numbers of rules, thereby creating a second-order data mining problem: which of the generated rules are the most interesting? And: should interestingness be measured objectively or subjectively? To tackle the amount of rules that are created during the mining step, the authors propose the combination of two novel ideas: first, there is rule change mining, which is a novel extension to standard association rule mining which generates potentially interesting time-dependent features for an association rule. It does not require changes in the existing rule mining algorithms and can therefore be applied during post-mining of association rules. Second, the authors make use of the existing textual description of a rule and those newly derived objective features and combine them with a novel approach towards subjective interestingness by using relevance feedback methods from information retrieval. The combination of these two new approaches yields a powerful, intuitive way of exploring the typically vast set of association rules. It is able to combine objective and subjective measures of interestingness and will incorporate user feedback. Hence, it increases the probability of finding the most interesting rules given a large set of association rules.

INTRODUCTION

Nowadays, the discovery of association rules is a relatively mature and well-researched topic. Many algorithms have been proposed to ever faster discover and maintain association rules. However, one of the biggest problems of association rules still remains unresolved. Usually, the number of discovered associations will be immense, easily in the thousands or even tens of thousands. Clearly, the large numbers make rules difficult to examine by a user. Moreover, many of the discovered rules will be obvious, already known, or not relevant.

For this reason a considerable amount of methods have been proposed to assist a user in detecting the most interesting or relevant ones. Studies about interestingness measures can roughly be divided into two classes: objective and subjective measures. Objective (data-driven) measures are usually derived from statistics, information theory or machine learning and assess numerical or structural properties of a rule and the data to produce a ranking. In contrast to objective measures, subjective (user-driven) measures incorporate a user's background knowledge and mostly rank rules based on some notion of actionability and unexpectedness.

In spite of a multitude of available publications the problem of interestingness assessment still is regarded as one of the unsolved problems in data mining and still experiencing slow progress (Piatetsky-Shapiro, 2000). The search for a general solution is one of the big challenges of today's data mining research (Fayyad et al., 2003). Existing approaches for interestingness assessment have several shortcomings which render them inadequate for many real-world applications.

Nonetheless, objective and subjective measures both have their justification to be used within the process of interestingness assessment. Objective measures help a user to get a first impression at what has been discovered and to obtain a starting point for further exploration of the rule set. This exploration step can then be accomplished by methods for subjective interestingness assessment. Ideally, the interestingness assessment of association rules should therefore be seen as a two step process. It is clear that for this process to be optimal it is necessary that both, the calculus used for the objective and the subjective rating, are based on the same notion of interestingness. Nevertheless, most approaches for objective and subjective ratings have been developed independently from each other with no interaction in mind such that the information utilized for the objective is neglected for the subjective rating. In fact, approaches rarely do fit together.

In this article we discuss a framework which combines objective and subjective interestingness measures to a powerful tool for interestingness assessment and addresses the problems mentioned above. Our framework incorporates several concepts which only recently have been introduced to the area of interestingness assessment: rule change mining and user dynamics. In particular, we show how to analyse association rules for *changes* and how information about change can be used to derive meaningful and interpretable objective interestingness measures. Based on the notion of change, we discuss a novel *relevance feedback* approach for association rules. We relate the problem of subjective interestingness to the field of Information Retrieval where relevance estimation is a rather mature and well-researched field. By using a vector-based representation of rules and by utilizing concepts from information retrieval we provide the necessary tool set to incorporate the knowledge about change into the relevance feedback process.

BACKGROUND

Significant research has been conducted into methods which assess the relevance, or interestingness, of a rule. Studies concerning interestingness assessment can roughly be divided into two classes. The first class are objective measures. These are usually derived from statistics, information theory or machine learning and assess numerical or structural properties of a rule and the data to produce a ranking (Tan et al., 2004). Objective measures do not take any background information into account and are therefore suitable if an unbiased ranking is required, e.g. in off-the-shelf data mining tools. Examples of such measures are lift, conviction, odds ratio and information gain. Overviews can be found in (Tan and Kumar, 2000), (McGarry, 2005), and (Geng and Hamilton, 2006). In (Tan et al., 2004) it is empirically shown that some measures produce similar rankings while others almost reverse the order. This poses the problem of choosing the right measure for a given scenario. One solution is to discover all rules that are interesting to any measure out of a predefined set (Bayardo, Jr. and Agrawal, 1999). A different approach is presented by (Tan et al., 2004). They developed two preprocessing methods which–integrated into a mining algorithm–render many measures consistent with each other. The same publication also presents an algorithm which finds a measure that best fits the requirements of a domain expert. This is accomplished by an interactive interestingness rating of a small set of patterns.

The second class are subjective measures which incorporate a user's background knowledge. In this class a rule is considered interesting if it is either *actionable* or *unexpected*. Actionability of a rule means that the user "can act upon it to his advantage" (Silberschatz and Tuzhilin, 1996). Their focal point is on rules that are advantageous for the user's goals. The actionability approach needs detailed knowledge about the current goals and also about the cost and risks of possible actions. Systems that utilise it are hence very domain specific, like the *KEFIR* system described in (Piatetsky-Shapiro and Matheus, 1994).

A rule is unexpected if it contradicts the user's knowledge about the domain. Systems that build upon this approach require the user to express his domain knowledge–a sometimes difficult, long and tedious task. The methods are usually based on pairwise comparison of a discovered rule with rules representing the user knowledge. This comparison can be logic-based, as in (Padmanabhan and Tuzhilin, 1999), (Padmanabhan and Tuzhilin, 2000) or (Padmanabhan and Tuzhilin, 2002) or syntax-based (Liu et al., 1997). In logic-based systems a contradiction is determined by means of a logical calculus, whereas in syntax-based systems a rule contradicts if it has a similar antecedent but a dissimilar consequent.

In (Padmanabhan and Tuzhilin, 1999), (Padmanabhan and Tuzhilin, 2000) and (Padmanabhan and Tuzhilin, 2002) the authors connect belief models with association rules. In particular, they assume that a belief system has been provided by the user whereby beliefs are defined as association rules. Based on this definition they provide a set of conditions to verify whether a rule is *unexpected* with respect to the belief on the rule database *D*. They propose an algorithm *ZoomUR* which discovers the set of unexpected rules regarding a specified set of beliefs. The algorithm itself consists of two different discovery strategies: *ZoominUR* discovers all unexpected rules that are refinements (or specialisations). On the other hand, *ZoomoutUR* discovers all unexpected rules that are more general.

In (Liu et al., 1997) the authors address the insufficiency of objective interestingness measures by focusing on the unexpectedness of generalised association rules. They assume that taxonomies exist among association rules' attributes. In subsequent work (Liu et al., 2000), human knowledge is recognised to have different degrees of certainty or preciseness. Their system allows for three degrees, notably *general impressions*, *reasonably precise concepts* and *precise knowledge*. The approach they propose accounts for these degrees and uses the gathered knowledge to find rules which are unexpected in regard to the expressed knowledge. The approach works iteratively: first, the user specifies his knowledge or modifies previously specified knowledge, supported by the specification language; second, the system analyses the association rules according to conformity and unexpectedness; and third, the user inspects the analysis results (aided by visualisation), saves interesting rules and discards uninteresting rules.

How to incorporate user dynamics into the relevance assessment has been studied in (Wang et al., 2003). They propose an approach based on two models which a user has to specify prior to any analysis: a model of his existing knowledge and a model of how he likes to apply this knowledge. The degree of unexpectedness of each discovered rule is calculated with respect to these two models. Their approach is based on what they call the See-and-Know assumption. Once a user has seen a rule, the rule itself and similar rules are not of interest anymore. Our approach, in contrast, uses two classes of seen rules, relevant and non-relevant ones. The ranking is calculated by aggregating the (dis-)similarity of a rule with respect to rules in both classes. Our approach also does not require a user to specify any kind of prior model of his knowledge.

In the area of rule change mining the discovery of interesting changes in histories for association rules has been studied by several authors. In (Agrawal and Psaila, 1995) a query language for history shapes is introduced. In (Dong and Li, 1999) and (Zhang et al., 2000) efficient algorithms which detect emerging itemsets are proposed. A fuzzy approach to reveal the regularities in how measures for rules change and to predict future changes was presented by (Au and Chan, 2005). In (Chakrabarti et al., 1998) an algorithm that ranks itemsets based on a change measure derived from the minimum description length principle is presented. (Liu et al., 2001b) proposes a statistical approach to distinguish trend, semi-stable and stable rules with respect to their histories of confidence and support. In (Liu et al., 2001a) a method to detect so-called fundamental rule changes is presented.

PROBLEMS

Data mining aims at discovering patterns in data which are novel, potentially useful and understandable (Fayyad et al., 1996). While being understandable is an inherent property of association rules which largely contributes to their popularity, it is the task of interestingness assessment to decide which of the many rules discovered are novel and useful. In practise, nevertheless, existing methods often perform poorly in reaching this goal.

Objective measures rely on a user's ability to choose the right measure for a given scenario out of a huge set of available ones. In (Tan et al., 2004) it is empirically shown that some measures produce similar rankings while others almost reverse the order. This poses the problem of choosing the right measure for a given scenario. Moreover, due to their rather mathematical foundations most measures lack interpretability and meaningfulness because the rule properties they measure rarely reflect the practical considerations of a user. For a user it is often unclear which measure to choose and how to link its results to his application scenario. Consequently many rules deemed interesting will not be very useful. Because objective measures do not memorize the past they are unable to identify patterns which have already been discovered multiple times in the past, which are diminishing or emerging. This ability, in turn, is crucial for distinguishing novel patterns from prevalent ones which often represent domain knowledge and thus are of less interest.

Subjective measures, on the other hand, require a user to be aware what he knows, to have a rough idea what he is looking for, and to be able to specify this knowledge in advance. A lot of effort is necessary to collect, organise and finally incorporate domain knowledge into a knowledge base against which association rules will be compared. Moreover, domain experts often forget certain key aspects or may not remember others which come into play under rarer circumstances. This problem can be termed ‘expert dilemma’ and has already been observed by designers of expert systems in the 1980s (Fogel, 1997). Building a knowledge base can also become a task never to be finished. During the knowledge acquisition process domain knowledge may become outdated, invalid, or loose its relevance. On the other hand, new knowledge may evolve. Users almost always have only partial, if any, awareness about this knowledge ageing process. Because of these knowledge dynamics it is often difficult to obtain a complete knowledge base. In contrast, subjective approaches treat domain knowledge as something static that never changes. Hence, they do not account for the ageing of knowledge nor do they support the user in maintaining it. Consequently, there is a risk that patterns are regarded as interesting based on outdated knowledge while a user is being left uninformed about the outdatedness itself.

ON THE INTERESTINGNESS OF CHANGE

While it is very challenging to design an algorithmic method to assess the interestingness of a rule, it is astonishingly simple for us humans to decide what is relevant to us and what is not. One of the clues to how humans judge the interestingness of an object is that they take its past and how it changes into account. When investing in stocks or buying expensive consumer goods one does not only look at the current price but also how it developed other the last couple of months. When we like to place a bet we do not only look at how a team scored last weekend but during the whole season. When we drive a car we could see many objects in our field of vision but we focus only on those which change, for example visually, like warning signs, or spatially, like pedestrians and cars.

For a business change can mean a risk (like a shrinking subgroup of target customers) or opportunity (like an evolving market niche). In either case, the business has to detect the change in order to survive or to win. In some business domains the value of information about change as a key enabler for anticipating events and conditions that may arise has been known for a long time. For example, stock traders aim to optimize buy and sell decisions by analyzing stock price behaviour over time. Moreover, many data collected are already time-stamped. In fact, the time dimension is the one dimension which is present in every data warehouse (Kimball, 1996). Due to its temporal nature business data reflect external influences like management decisions, economic and market trends and thus capture the changes a business is interested in.

Change, therefore, has some inherent interestingness for a user, and is likewise a concept that is easy to understand and can directly lead to business actions. For this reason it provides a basis for assessing a rule’s interestingness and has already proven to be successful in a variety of applications, e.g. retail marketing (Chen et al., 2005), exception detection (Baron et al., 2003) and customer segmentation (Boettcher et al., 2007). Rules which change hint at unknown or surprising changes in the underlying data-generating process which may require intervening action (Chakrabarti et al., 1998). A downward trend in a rule’s confidence indicates that it may disappear from the discovered rule set in the future, while an upward trend may hint at a rule which has emerged recently. On the other hand, rules which are stable over time often represent invariant properties of the data-generating process and thus are either already known, or, if discovered once, should not be displayed to the user a second time. Nonetheless, information about stability can be useful if a domain is only poorly understood. In the context of change, the information whether a rule has a high confidence is of less interest than the information that the confidence has a trend or other regular, or surprising characteristics.

HISTORIES OF ASSOCIATION RULE MEASURES

The underlying idea of our framework is to detect interesting association rules by analysing their support and confidence along the time axis. The starting point of such a *rule change mining* approach is as follows: a timestamped data set is partitioned into intervals along the time axis. Association rule discovery is then applied to each of these subsets. This yields sequences—or *histories*—of support and confidence for each rule, which can be analysed further. Of particular interest are regularities in the histories which we call *change patterns*. They allow us to make statements about the future development of a rule and thus provide a basis for proactive decision making.

In the following we will denote an association rule r as $X \rightarrow Y$ where X and Y are itemsets, $|Y| > 0$ and $X \cap Y = \emptyset$. If for two rules $r : X \rightarrow y$ and $r' : X' \rightarrow y$, $X \subset X'$ holds, then it is said that r is a *generalisation* of r'. This is denoted by $r' \prec r$. As usual, the reliability of a rule $r : X \rightarrow y$ is measured by its *confidence* conf(r) and the statistical significance measured by its *support* supp(r). We also use the support of an itemset X denoted by supp(X).

Further, let D be a time-stamped data set and $[t_0,t_n]$ the minimum time span that covers all its tuples. The interval $[t_0,t_n]$ is divided into n>1 non-overlapping periods $T_i := [t_{i-1}, t_i]$, such that the corresponding subsets $D(T_i) \subset D$ each have a size $|D(T_i)| \gg 1$. Let $\hat{T} := \{T_1, \ldots, T_n\}$ be the set of all periods, then for each $T_i \in \hat{T}$ association rule mining is applied to the transaction set $D(T_i)$ to derive rule sets $R(D(T_i))$.

Because the measures, like confidence and support, of every rule $r : X \rightarrow y$ are now related to a specific transaction set $D(T_i)$ and thus to a certain time period T_i we need to extend their notation. This is done straightforwardly and yields $supp(r,T_i)$ and $conf(r,T_i)$.

Each rule

$$r \in \hat{R}(D) := \bigcap_{i=1}^{n} R(D(T_i))$$

is therefore described by n values for each measure. Imposed by the order of time the values form sequences called *confidence history*

$$H_{conf}(r) := (conf(r, T_1), \ldots conf(r, T_n))$$

and *support history*

$$H_{supp}(r) := (supp(r, T_1), \ldots supp(r, T_n))$$

of the rule r. These histories are the input to most rule change mining approaches, which then detect interesting change patterns.

FRAMEWORK FOR RULE INTERESTINGNESS ASSESSMENT

In order to analyse rules for change *rule histories* need to be derived which in turn serve as the basis for both objective and subjective interestingness assessment. To derive a history, data sets collected during many consecutive periods have to be analysed for association rules. After each analysis session the discovered rules have to be compared to those discovered in previous periods and their histories have to be extended. On the other hand, history values may be discarded if their age exceeds an application-dependent threshold. Therefore, rules and histories have to be stored on a long-term basis. Taking all of the aforesaid into account the first task of our framework is:

1. Association rules have to be *discovered* and their histories efficiently stored, managed and maintained.

If histories with a sufficient length are available, the next task is straightforward and constitutes the core component of rule change mining:

2. Histories that exhibit specific change patterns have to be reliably *detected*.

Association rule discovery is generally connected with two problems. In the first place, a vast number of rules will be detected, which is also referred to as the *rule quantity problem*. Secondly, rules may be obvious, already known or not relevant, which is also referred to as the *rule quality problem* (Tan and Kumar, 2000).

Since a history is derived for each rule, the rule quantity problem also affects rule change mining: it has to deal with a vast number of histories and thus it is likely that many change patterns will be detected. Moreover, as we will briefly discuss in the following section methods that were developed to deal with this problem for association rules cannot be used in rule change mining. Furthermore, there is also a quality problem: not all of the detected change patterns are equally interesting to a user and the most interesting are hidden among many irrelevant ones. Overall, the third task is:

3. Histories with a change pattern have to be analysed for redundancies and *evaluated* according to their interestingness.

Such an initial interestingness ranking for association rules proves to be helpful in providing a user with a first overview over the discovered rules and their changes. Still, it is also clear that the user starts to build his own notion of interestingness as soon as he starts browsing the rules and histories. Our framework should support this dynamics of interestingness and therefore the fourth task is:

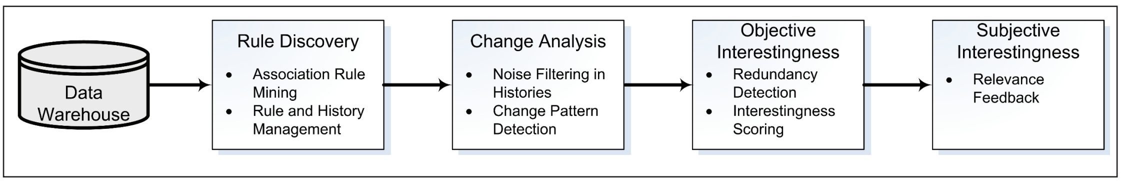
4. A user's feedback about the rules and histories seen thus far should be collected, analysed and used to obtain a new interestingness ranking.

Because the aforementioned tasks build upon each other, they can be seen as layers of a processing framework termed *Rule Discovery*, *Change Analysis*, *Objective Interestingness*, and *Subjective Interestingness*. The framework itself has first been described in (Boettcher et al, 2006) and evaluated in a real-life scenario in (Boettcher et al, 2007). Figure 1 illustrates and summarises the workflow.

RULE DISCOVERY

Given a timestamped data set collected during a certain period, the task of the Rule Discovery layer is to discover and store the association rules hidden in it. Therefore, the first component of this layer is an association rule mining system, its second component is a database that stores and manages rules and their histories. Both components, but also the choice of the time periods, will be explained in the following.

Figure 1. Detailed design of each layer



In order to obtain the data set, the period length has to be chosen. Two aspects have to be considered. Long periods lead to many transactions in the individual data sets for the different periods and thus can enhance the reliability of the metrics used. However, due to this coarse-grainedness interesting short-duration patterns may be missed. Short periods allow for measuring a rule's statistics more frequently. The reduced robustness in the model estimation makes it then more difficult to distinguish true change patterns from incidental worthless patterns. The choice of period length should therefore depend on the application. Data often is collected in regular intervals, for instance survey may be conducted weekly or bulk updates to a database may be carried out daily. In practise, these application-specific intervals can then be used to guide the choice of the time period.

After the data set is available, association rule mining is applied to it. A typical system for association rule mining may not only consist of the rule miner itself, but also of methods for pruning, constrained mining and interestingness assessment. Such methods have been developed to cope with the problem of a vast amount of discovered rules in each period. A huge number of histories has to be processed and consequently far too many change patterns will be reported. In order to cope with this problem, pruning methods are used to constrain the set of generated rules. From the perspective of rule change mining such pruning methods treat rule sets independently from each other. However, in rule change mining we have many, temporally ordered rule sets. Thus the rule property utilized for pruning—in general a measure based on rule statistics—may vary for some rules over time, but still match the pruning criterion in each rule set. Although these variations may render rules interesting, they are discarded by approaches for association rule pruning. Consequently, conventional pruning approaches should not directly be used in conjunction with rule change mining.

CHANGE ANALYSIS

The task of change analysis is to discover change patterns in rule histories. In this article, however, we only discuss how histories are detected that are stable or exhibit a trend. The Change Analysis layer fulfils its task by a two-step approach. In the first step a filter is applied to the histories to reduce the noise contained in them. In a second step statistical tests for trend and stability are conducted.

Rule histories inherently may contain random *noise*. Random noise may influence subsequent analysis steps in such a way that wrong and misleading results are produced. To reduce this effect we use *double exponential smoothing* (Chatfield, 2003) in order to reveal more clearly any trend or stability. It is a simple and fast, yet effective method, which can easily be automated. Nevertheless, it has to be considered that after smoothing association rule measures may be inconsistent with each other. For example, the confidence of a rule can in general not be obtained anymore by dividing the rule's support by the support of its antecedent itemset.

Detection of Trends

A trend is present if a sequence exhibits steady upward growth or a downward decline over its whole length. This definition is rather loose, but in fact there exists no fully satisfactory definition for trend (Chatfield, 2003). From a data mining perspective a trend describes the pattern that each value is likely to be larger or smaller than all its predecessors within a sequence, depending on whether the trend is upward or downward. Hence it is a qualitative statement about the current and likely future development of a sequence. However, taking aspects of interpretability and usefulness into account, such a statement is sufficient in the case of rule change mining. When faced with a vast number of rules and their histories, a user often has a basic expectation whether they should exhibit a trend and of what kind. By comparing his expectations with reality he will mostly be able to roughly assess the implications for its business. On the other hand, a user will rarely know in advance how trends should look like quantitatively, e.g., their shape or target values. Thus he may be unable to exploit the advantages of more sophisticated trend descriptions, like regression models.

To choose a method for trend detection, it has to be taken into account that the number of sequences to examine is huge. Whenever a trend is reported the user is basically forced to rely on the correctness of this statement, because it is infeasible for him to verify each trend manually. In addition to the requirement of reliable detection, the method should incorporate no assumptions about any underlying model, because it is very unlikely that it will hold for all or at least most sequences. Therefore non-parametric statistical tests are the appropriate choice for trend detection.

Within our framework we provide two statistical tests for trend, the *Mann-Kendall test* (Mann, 1945) and the *Cox-Stuart test* (Cox and Stuart, 1955). The Cox-Stuart test exploits fewer features of the sequence, leading to a computational effort that increases linearly with the sequence length. Although this may render the Cox-Stuart test susceptible to noise, because the influence of artefacts on the test result is stronger, it is considerably faster for long sequences. In contrast to this, the Mann-Kendall test is much more robust, but its computational effort increases quadratically with the sequence length. Therefore it has to be determined which of the two issues—speed or robustness—is more important depending on the actual application scenario.

Detection of Stability

Roughly speaking, a history is considered stable if its mean level and variance are constant over time and the variance is reasonably small. Similar to trends, a clear definition of stability is difficult. For example, a sequence may exhibit a cyclical variation, but may nevertheless be stable on a long term scale. Depending on the problem domain, either the one or the other may have to be emphasised. From a data mining perspective, stability describes the pattern that each value is likely to be close to a constant value, estimated by the mean of its predecessors. Thus it is, like a trend, a qualitative statement about the future development of a sequence. However, in contrast to a trend, it can easily be modelled in an interpretable and useful way, e.g., by the sequence's sample mean and variance. Generally, stable rules are more reliable and can be trusted—an eminently useful and desirable property for long-term business planning (Liu et al., 2001b).

To test for stability we use a method based on the well-known chi-square test which was proposed in (Liu et al., 2001b). However, since the test does not take the inherent order of a history's values into account, the single use of this method may infrequently also classify histories as stable which actually exhibit a trend. Therefore, we chose to perform the stability test as the last one in our sequence of tests for change patterns.

OBJECTIVE INTERESTINGNESS

Since usually a vast number of change patterns will be detected, it is essential to provide methods which reduce their number and identify potentially interesting ones. This is the task of objective interestingness assessment. To reduce the number of change patterns to assess we use a redundancy detection approach, based on so-called derivable histories.

Non-Derivable Rules Filter

Generally, most changes captured in a history and consequently also change patterns are simply the snowball effect of the changes of other rules (Liu et al., 2001a). Suppose we are looking at churn prevention and our framework would discover that the support of the rule

$r_1 : \text{Age} > 50 \rightarrow \text{Complain} = \text{Yes}$

shows an upward trend. That is, the fraction of customers over 50 who complain increased. However, if the fraction of males among all over 50 year old complaining customers is stable over time, the history of

$r_2 : \text{Age} > 50, \text{Gender} = \text{Male} \rightarrow \text{Complain} = \text{Yes}$

shows qualitatively the same trend. In fact, the history of rule r_2 can be *derived* from the one of r_1 by multiplying it with a gender-related constant factor. For this reason, the rule r_2 is *temporally redundant* with respect to its history of support.

It is reasonable to assume that a user will generally be interested in rules with non-derivable and thus non-redundant histories, because they are likely key drivers for changes. Moreover, derivable rules may lead to wrong business decisions. In the above example a decision based on the change in rule would account for the gender as one significant factor for the observed trend. In fact, the gender is completely irrelevant. Therefore, the aim is to find rules that are non-redundant in the sense that their history is not a derivative of related rules' histories. In a way, the approach is to search for and discard rules that are not the root cause of a change pattern which, in turn, can be seen as a form of pruning. In order to find derivable rules we have to answer the following questions. First, what is meant by *related* rules, and second, what makes a history a *derivative* of other histories. Regarding the first question, a natural relation between association rules is *generalisation*. We therefore define that a rule r' is *related to a rule* r iff r' is more general than r, i.e. $r \prec r'$.

The following definition for derivable measure histories includes those of itemsets as a generalisation from rules. Thereby, the superset relation is used to define *related itemsets*: an itemset Y is related to an itemset X iff $X \prec Y := X \supset Y$. As before, XY is written for $X \cup Y$.

Definition 1: Let s, $s_1, s_2 \ldots s_p$ be rules or itemsets with $s \prec s_i$ for all i and p>0. In case of rules, let the antecedent itemsets of the s_i be pairwise disjoint, in case of itemsets let the s_i be pairwise disjoint. Let m be a measure like support or confidence, $m(T) := m(s,T)$ and $m_i(T) := m(s_i, T)$ its functions over time and $\mathcal{M} := \{g : \mathbb{R} \rightarrow \mathbb{R}\}$ be the set of real-valued functions over time. The history $H_m(s)$ regarding the measure m is called derivable iff a function $f : \mathcal{M}^p \rightarrow \mathcal{M}$ exists such that for all $T \in \hat{T}$

$$m(T) = f(m_1, m_2, \ldots, m_p)(T) \qquad (1)$$

For simplicity, we call a rule or itemset *derivable with respect to a measure* m iff its history of m is derivable. The temporal redundancy of a rule therefore depends on the measure under consideration, e.g. a rule can be redundant (derivable) with respect to its support history, but not redundant (not derivable) with respect to its confidence history. This in turn is consistent with existing rule change mining approaches, because they typically process histories of different measures independently from another.

The main idea behind the above definition is that the history of a rule (itemset) is derivable, if it can be constructed as a mapping of the histories of more general rules (itemsets). To compute the value m(s,T) the values $m(s_i,T)$ are thereby considered. The definition above does not allow for a pointwise definition of f on just the $T \in \hat{T}$ but instead states a general relationship between the measures of the rules independent from the point in time. It can therefore be used to predict the value of, for example, supp(s) given future values of the $\text{supp}(s_i)$. A simple example we will see below is $m = f(m_1) = cm_1$, i.e. the history of a rule can be obtained by multiplying the history of a more general rule with a constant c.

In the following we introduce three criteria for detecting derivable histories which can be used in combination or independently from another. The first two criteria deal with itemsets and can therefore be directly applied to the support of rules as well. The last criterion is related to histories of rule confidences. The functions f are quite simple and we make sure that they are intuitive.

The first criterion checks if the support of an itemset can be explained with the support of exactly one less specific itemset.

Criterion 1: The term supp(XY,T)/supp(Y,T) is constant over $T \in \hat{T}$ given disjoint itemsets X and Y.

When being rewritten as

$$c = \text{supp}(XY, T) / \text{supp}(Y, T) = P(XY|T) / P(Y|T) = P(X|YT)$$

with a constant c the meaning of the criterion becomes clear. The probability of X is required to be constant over time given Y, so the fraction of transactions containing X additionally to Y constantly grows in the same proportion as Y. Due to

$$\text{supp}(XY, T) = c \cdot \text{supp}(Y, T) \tag{2}$$

with c=supp(XY,T)/supp(Y,T) for any $T \in \hat{T}$, XY is obviously a derivative of Y with respect to support history as defined in Definition 1.

Figure 2 and Figure 3 show an example of a derivable support history of a rule. The histories have been generated from a customer survey dataset which is described in (Boettcher et al, 2006). Figure 2 shows the support histories of the less specific rule at the top and the more specific rule underneath over 20 time periods. The shape of the two curves is obviously very similar and it turns out that the history of the more specific rule can be approximately reconstructed using the less specific one based on (2). As shown in Figure 3 the reconstruction is not exact due to noise.

Opposed to the criterion above, the following is based on the idea of explaining the support of an itemset with the support values of two subsets.

Criterion 2: The term

$$\frac{\text{supp}(XY, T)}{\text{supp}(X, T)\text{supp}(Y, T)}$$

is constant over $T \in \hat{T}$ given disjoint itemsets X and Y.

supp(XY,T) measures the probability of the itemset XY in period T which is P(XY | T). The term

$$\frac{\text{supp}(XY, T)}{\text{supp}(X, T), \text{supp}(Y, T)} = \frac{P(XY|T)}{P(X|T)P(Y|T)}$$

Figure 2. Histories of the rule $X \rightarrow z$ and its derivable rule $Xy \rightarrow z$

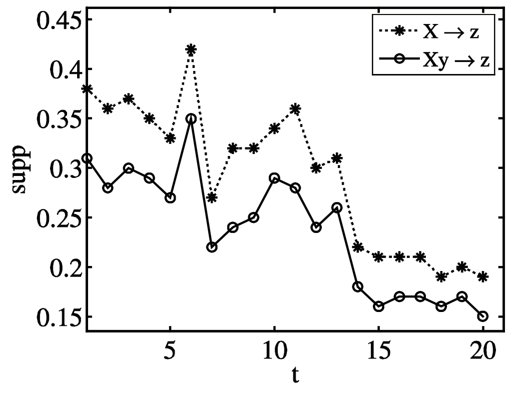
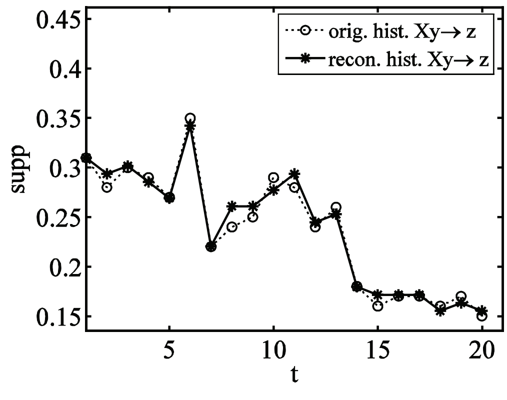


Figure 3. Reconstructed history of $Xy \rightarrow z$ using the history of $X \rightarrow z$



is quite extensively used in data mining to measure the degree of dependence of X and Y at time T. Particularly in association rule mining this measure is also known as *lift* (Webb, 2000). The criterion therefore expresses that the degree of dependence between both itemsets is constant over time. The support history of XY can then be constructed using

$$\mathrm{supp}(XY,T) = c{\cdot}\mathrm{supp}(X,T)\mathrm{supp}(Y,T) \qquad (3)$$

with c=supp(XY,T)/(supp(X,T)supp(Y,T)) for any $T \in \hat{T}$, that is, the individual support values of the less specific itemsets are used corrected with the constant degree of dependence on another. According to Definition 1 the support history of XY is therefore derivable.

Overall, an itemset is considered derivable with respect to support if more general itemsets can be found, such that at least one of the Criteria 1 or 2 holds.

Finally, the last criterion deals with derivable confidence histories of rules.

Criterion 3: The term

$$\frac{\mathrm{conf}(r,T)}{\mathrm{conf}(r',T)}$$

is constant over $T \in \hat{T}$ given two rules r and r' with $r \prec r'$.

Assuming the rules $r: XY \to z$ and $r': Y \to z$ with disjoint itemsets X and Y, the criterion translates to

$$\frac{P(z|XYT)}{P(z|YT)}$$

being constant over time. This basically means that the contribution of X in addition to Y to predict z relative to the predictive power of Y remains stable over time and can therefore be neglected. The confidence history of r is derivable because of the following. Be $c = \mathrm{conf}(r,T)/\mathrm{conf}(r',T)$ for any $T \in \hat{T}$, then for all $T \in \hat{T}$

$$\mathrm{conf}(r,T) = c{\cdot}\mathrm{conf}(r',T) \qquad (4)$$

Suitable statistical tests for these criterions have been proposed in (Boettcher et al, 2005).

Objective Interestingness Measures

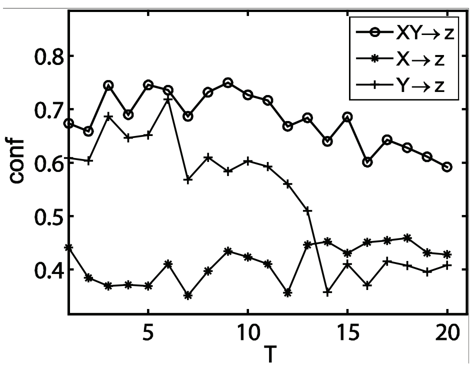
To assess the interestingness of detected trends and stabilities it has to be considered that each history is linked to a rule, which, prior to rule change mining, has a certain relevance to a user. However, the detection of a specific change pattern may significantly influence this prior relevance. In this sense a rule can have different degrees of interestingness, each related to another history. However, there is no broadly accepted and reliable way of measuring a rule's interestingness up to now (Tan et al., 2004). Therefore we consider any statement about the interestingness of a history also as a statement about the interestingness of its related rule.

To assess stable histories two things should be considered. Firstly, association rule discovery typically assumes that the domain under consideration is stable over time. Secondly, measures like support and confidence are interestingness measures for rules themselves. It is summarised by the mean of its values, which in turn can then be treated as an objective interestingness measure. Here the variance of the history can be neglected, since it is constrained by the stability detection method.

Developing objective interestingness measures for trends is more complex due to their richness of features. For identifying salient features of a given trend, it is essential to provide reference points for comparison. As such we chose the assumptions a user naively makes in the absence of any knowledge about the changes in rule histories. From a psychological perspective they can be seen as the anchors relative to which histories with a trend are assessed: a trend becomes more interesting with increasing inconsistency between its features and the user's naive assumptions. We identified three such assumptions and defined heuristic measures for the discrepancy between a history and an assumption:

- **Stability:** Unless other information is provided, a user assumes that histories are stable over time. This assumption does not mean that he expects no trends at all, but expresses his naive expectations in the absence of precise knowledge about a trend. It should be noted that this is consistent with conventional association rule mining, which implicitly assumes that the associations hidden in the data are stable over time. The confidence histories of the rule $XY \rightarrow z$ in Figure 4 would violate the stability assumption because its trend is very clear.
- **Non-rapid change:** Since a user shapes his business, he will be aware that the domain under consideration changes over time. However, he will assume that any change is continuous in its direction and moderate in its value. For example, if a business starts a new campaign, it will probably assume that the desired effect evolves moderately, because, for instance, not all people will see a commercial immediately. On the other hand, a rapid change in this context attracts more attention, because it may hint at an overwhelming success or an undesired side effect. For example, the history of the rule $Y \rightarrow z$ in Figure 4 would be very interesting according to the non-rapid change assumption because the depicted trend is very pronounced and steep.
- **Homogeneous change:** If the support of a rule (itemset) changes over time, it is assumed that the rate and direction of changes in the support of all its specialisations are the same. This basically means that the observed change in the rule (itemset) does not depend on further items. For example, a user may know that the fraction of satisfied customers increases. The homogeneous change assumption states that the observed change in satisfaction affects all customers and not only selected subpopulations, e.g. females over 40. If, on the other hand, the confidence of a rule changes over time, it is assumed that the confidence of all more specialised rules changes at the same rate. For example, the history of the rule $XY \rightarrow z$ in Figure 4 would be very interesting because its shape is completely different from those of its more general rules.

Figure 4 . Examples of interesting histories which exhibit a trend



SUBJECTIVE INTERESTINGNESS

Conservative approaches that employ mostly objective measures of interestingness only insufficiently reflect the way a user searches for relevant rules because a user's perception of relevance is not a static but rather a dynamic process due to several reasons: firstly, when a user starts to explore a set of discovered association rules he only has a very vague notion about which rules might be relevant to him. Secondly, while seeing more rules his knowledge about the domain of interest changes, some aspects might gain while others might lose importance. His notion of relevance depends on these changes and thus changes too, almost always becoming clearer. The more rules a user examines, the more knowledge he gathers about the domain of interest. This knowledge then helps him to decide for newly encountered rules whether they are (non-)relevant for him, for example, because they are kind-of similar to previously seen (non-)relevant ones.

The importance of user dynamics and incremental knowledge gathering in assessing the relevance of data mining results only recently gained attention in the research community (Wang et al., 2003). However, it is a rather well-researched topic in the field of information retrieval where it has been known for a long time that a user cannot express his information need from scratch. For example, when using an internet search engine to search documents about a non-trivial topic most users start with a rather simple query. By analyzing the search results they gain more knowledge about what they actually look for and thus are able to further refine their initial query, i.e. to express their notion of relevance more clearly. To support a user in this process techniques like relevance feedback based on document similarities have been developed.

In fact, the way a user builds up his internal notion of relevance when searching for the most relevant association rules described above is very similar to the models of user behaviour used in information retrieval (cf. (Baeza-Yates and Ribeiro-Neto, 1999)). Based on these similarities we present a new approach to the problem of finding the most relevant rules out of a large set of association rules which is inspired by ideas from information retrieval. Our approach, as described in, e.g., (Ruß, 2007),or (Ruß, 2008) , uses relevance feedback to acquire users' preferences and to build a knowledge base of what he considers to be relevant and non-relevant, respectively. By calculating the (dis-)similarity of each unexamined rule with the rules in the knowledge base and aggregating the scores we obtain a relevance score which—with each feedback provided—better reflects the user's notion of relevance.

Using Concepts from Information Retrieval

Existing approaches to assess the relevance of association rules strongly require a user to explicitly specify his existing knowledge in advance. This leads to two major drawbacks. In the first place, when specifying their existing knowledge, domain experts often forget certain key aspects or may not remember others which come into play under rarer circumstances. This problem can be termed 'expert dilemma' and has already been observed by designers of expert systems in the 1980s (Fogel, 1997). Secondly, at the beginning of an analysis session a user can only very vaguely specify what he considers to be relevant. His notion of relevance only becomes clearer the more rules he examines. This problem, that a user is incapable of specifying his information need from scratch, is very well-known in the field of information retrieval (Baeza-Yates and Ribeiro-Neto, 1999) where it lead to the development of relevance feedback methods.

Relevance feedback is an intuitive technique that has been introduced to information retrieval in the mid-1960s (Salton, 1971). In information retrieval it is a controlled, semi-automatic, iterative process for query reformulation that can greatly improve the usability of an information retrieval system (Jaakkola and Siegelmann, 2001). Relevance feedback allows a user to express what he considers to be relevant by marking rules as relevant and non-relevant, respectively. Whenever a rule has been marked as relevant, it is added to the set of relevant rules R_r. Whenever a rule is marked as non-relevant, it is added to the set of non-relevant rules R_n. For simplicity, we will assume that in each feedback cycle exactly one rule is marked.

After each feedback cycle the remaining rules are compared to the set of annotated rules and a new relevance score is calculated. The set of annotated rules, in turn, can be seen as a representation of the user's notion of relevance. Hence it also provides a solution to the first of the above-mentioned drawbacks by supporting an iterative, easy way for a user to specify his knowledge about a domain. For example, he may annotate rules that are already known as non-relevant and some novel rules as relevant.

In order to develop a feedback system for association rules the following questions need to be answered:

- How do we represent association rules for the purpose of relevance feedback?
- How do we score the likely relevance of a rule in relation to a rule already marked as (non-)relevant?
- How do we aggregate those scores to an overall relevance score?

We will provide answers to these questions in the subsequent sections. In particular we are aiming at adapting established methods from information retrieval.

Rule Representation

To be the core building block of a relevance feedback approach it is necessary to transform the rules into an equivalent representation. In particular, such a representation should have a couple of properties. Firstly, rather than relying on generalisation and specialisation relationships among rules as a key to rule similarity it should support a less crisp and thus more flexible definition. For example, rules that have the same consequent and share items in their antecedent should be regarded as similar to a certain degree. Secondly, items have a different importance to a user. For example, an item that is contained in almost every rule does not contribute much towards a user's understanding of the domain, whereas an item that is only contained in a few rules can contribute considerably. This importance should be reflected in the rule representation. Thirdly, it should be easy to extend the rule representation by further numeric properties of a rule. As we pointed out in the first part of this article, there has been increasing interest into the change of a rule's support and confidence values (cf. (Liu et al., 2001b; Boettcher et al., 2006)) as a key to rule interestingness. In this scenario the rule representation should account for the change of rules and allow for change information, which can be histories of support or confidence, or higher order features derived thereupon, to be incorporated in order to enable similarity calculations based on rule change. To illustrate the usage of further information about rules for relevance feedback we will use the example of rule change throughout this article.

As a representation that fulfils all of the above requirements we define a *feature vector of an association rule* r whose elements are numerical values and which consists of three components: a representation of the rule's antecedent, a representation of the rule's consequent and a rule's time series. The latter component can easily be replaced by other numeric features of a rule or completely omitted. The different components can be seen as a projection of $\vec{r}$ and will be referred to as follows:

$$\vec{r}_{body} = (r_1, \ldots, r_b) \tag{5}$$

$$\vec{r}_{head} = (r_{b+1}, \ldots, r_{b+h}) \tag{6}$$

$$\vec{r}_{sym} = (r_1, \ldots, r_{b+h}) \tag{7}$$

$$\vec{r}_{time} = (r_{b+h+1}, \ldots, r_{b+h+t}) \tag{8}$$

To calculate the *item weights* r_i we adapted the well-known TF-IDF approach (Salton and Buckley, 1987) from information retrieval. The TF-IDF approach weights terms according to their appearance in a document and in the overall document collection. A high term weight, which is correlated with a high importance of that particular term, is achieved if the term appears frequently in the document (term frequency, TF) but much less frequently in the document collection (inverse document frequency, IDF). This approach filters out commonly used terms and tries to capture the perceived relevance of certain terms.

This method, carried over to association rules, means that items that appear in the vast majority of rules will get a very low weight whereas items that are rather infrequent will get a rather high weight. Since item appearance in rules is linked to item appearance in a data set this also means that infrequent attribute values in the data set will receive a high weight.

The term frequency tf of an item x in an association rule r is calculated as follows:

$$tf(x,r) = \begin{cases} 1 & x \in r \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

The inverse document frequency idf of an item x in an association rule r and in regard to a rule set R is calculated as follows:

$$idf(x,R) = 1 - \frac{\ln|r : r \in R \wedge x \in r|}{\ln|R|} \tag{10}$$

To generate $\vec{r}_{body}$ and $\vec{r}_{head}$ a series of steps has to be performed. For antecedent and consequent separately, a set of items is generated: $I_{body}=\{x_1,...,x_b\}$ and $I_{head}=\{x_1,...,x_h\}$ where the x_i are the items that occur in antecedent or consequent of the association rules in R, respectively. Each item of these sets is assigned exactly one vector dimension in $\vec{r}_{body}$ or $\vec{r}_{head}$, respectively. Hence, the values for *b* and *h* in - are the cardinalities of the respective itemsets: $b = |I_{body}|$ and $h = |I_{head}|$

The part of the feature vector of an association rule r which covers antecedent and consequent consists of TF-IDF values. Let x_i the *i*-th item of the alphabetically ordered set I_{body} and let r_i be the *i*-th component of $\vec{r}_{body}$. Then, $\vec{r}_{body}$ is defined as follows:

$$r_i = tf(x_i, r) \cdot idf(x_i, R), \quad i = 1, \ldots, b \tag{11}$$

$\vec{r}_{head}$ is treated in the same way, except that x_j is the j-th item of the alphabetically ordered set $\vec{r}_{head}$

$$r_{b+j} = tf(x_j, r) \cdot idf(x_j, R), \quad j = 1, \ldots, h \tag{12}$$

Pairwise Similarity

A relevance feedback system must have the ability to compare unrated rules, or features of those, with rules previously rated as (non-)relevant. Instead of utilizing the generalisation and specialisation relationships among rules we choose a more flexible approach based on a notion of similarity among rules. As a similarity measure we have chosen the cosine similarity. It calculates the cosine of the angle between two n-dimensional vectors r and s as follows:

$$sim(\vec{r}, \vec{s}) = \frac{\sum_{i=1}^{n} r_i s_i}{\sqrt{r_i^2}\sqrt{s_i^2}} \tag{13}$$

Since the cosine measure yields values in [0,1], the corresponding dissimilarity measure is:

$$dissim(\vec{r}, \vec{s}) = 1 - sim(\vec{r}, \vec{s}) \tag{14}$$

The cosine similarity compared to other similarity measures, like ones based on the Euclidean distance, has the advantage that it does not take missing items in a rule into account. For example, when measuring the similarity between a rule $Xy \rightarrow z$ and its more general rule $X \rightarrow z$ only the item weights contained in both rules (i.e. X and z) contribute towards the similarity measure. This property of the cosine measure is also the reason why it is frequently used in information retrieval systems. When comparing, for example, a query with a document it is desirable only to take the actual words contained in the query into account and not each of the many words the user did not specify.

The similarity between rules' antecedents or rules' consequents can be calculated straightforwardly using the cosine measure, yielding $\mathrm{sim}(\vec{r}_{head}, \vec{s}_{head})$ and $\mathrm{sim}(\vec{r}_{body}, \vec{s}_{body})$, respectively. We aim to emphasize antecedent and consequent equally, so by averaging both we obtain the similarity of a rule $\vec{r}_{sym}$ with regard to a rule $\vec{s}_{sym}$:

$$\mathrm{sim}(\vec{r}_{sym}, \vec{s}_{sym}) = 0.5\mathrm{sim}(\vec{r}_{body}, \vec{s}_{body}) + 0.5\mathrm{sim}(\vec{r}_{head}, \vec{s}_{head}) \tag{15}$$

The cosine measure is also suitable as a measure of similarity $\mathrm{sim}(\vec{r}_{time}, \vec{s}_{time})$of a time series which we use in this article as an example of further information about rules embedded into the rule vector. For time series the cosine measure has the advantage only to reflect the magnitude of the angle between two vectors but—compared with other distance measures (e.g. Euclidean distance)—to ignore the magnitude difference between the two vectors. This means, it is robust w.r.t. different variation ranges of the time series. It is, however, not robust w.r.t. shifts of the time series' mean value. Nevertheless, robustness can be achieved by subtracting from both time series their respective mean value prior to similarity calculation.

Similarity Aggregation

So far, we have discussed how to calculate pairwise similarities between vectors which represent certain features of a rule like its consequent, antecedent or a time series of rule measures. For the purpose of relevance feedback it is necessary to measure the similarity of a feature of an unrated rule r relative to the features contained in the elements of a rule set R which may represent relevant and non-relevant rules. Generally, we define the similarity of a vector $\vec{r}$ relative to a set $R = \{\vec{s}_1, \dots, \vec{s}_m\}$ as

$$\mathrm{sim}_{rs}(\vec{r}, R) = \Omega(\{\mathrm{sim}(\vec{r}, \vec{s}_1), \dots, \mathrm{sim}(\vec{r}, \vec{s}_m)\}) \tag{16}$$

whereby Ω denotes a suitable aggregation operator which we will describe in the next section. As in the previous section the dissimilarity of a vector relative to a set is defined as

$$\mathrm{dissim}_{rs}(\vec{r}, R) = 1 - \mathrm{sim}_{rs}(\vec{r}, R) \tag{17}$$

The OWA Operator

Our choice of the aggregation operator Ω is guided by two requirements: firstly, the user should be able to influence the aggregation operator, either implicitly or explicitly. Secondly, to obtain comparable results, the aggregation operator should be able to represent also simple aggregation operators like min, max or median. These two requirements are met by the family of OWA operators, which originate in the Fuzzy Domain and have been introduced by (Yager, 1988). An OWA operator Ω is a mapping $\Omega : S \to R$, where *S* is a set of numerical values with $S \neq \emptyset$ and $|S| = n$. The OWA operator Ω has an associated weighting vector $W = (w_1, w_2, \dots, w_n)^T$ with $w_j \in [0,1]$ and

$$\sum_{j=1}^{n} w_j = 1.$$

It is defined as:

$$\Omega(\{s_1, s_2, \dots, s_n\}) = \sum_{j=1}^{n} w_j b_j \tag{18}$$

with b_j being the j-th largest of the s_i.

The most important feature of this operator is the ordering of the arguments by value. The OWA operator is in a way very general in that it allows different conventional aggregation operators. This is achieved by appropriately setting the weights in W–different arguments can be emphasised based upon their position in the ordering.

Min, *max*, *mean*, and *median* are special cases for the OWA operator and were described by (Yager, 1997). They illustrate the generality and flexibility of the OWA operator. By setting the weights accordingly, the user can influence the relevance score to suit the needs of his particular application scenario. For example, $(1/n,1/n,...1/n)^T$ yields the mean, whereas $(1,0,...,0)^T$ yields the maximum operator.

Furthermore, the OWA operator is strongly related to the concept of linguistic quantifiers, such as *many, a few, most*. In (Yager, 1988) the connection to linguistic quantifiers is presented by explaining how the weights of the OWA expression can be obtained by using the membership function of any linguistic quantifier.

Relative Importance of Recent Relevance Choices

The retrieval of relevant association rules is a consecutive, iterative process. The user's knowledge, his beliefs and assumptions change during the relevance feedback cycle as he sees more rules. Therefore, the user's latest choices should be considered as having a higher priority over the first, relatively uninformed ones. This concept can be captured as the *decay of a relevant or non-relevant* rule's importance over time. The similarity aggregation should account for this and thus should weight recently selected rules higher than older ones.

Let t(r) be the *age* of a relevant or non-relevant association rule r. This means, t(r) is the number of feedback cycles that have been performed since the rule r was marked as being (non-)relevant, thereby a newly selected rule receives t=0. Two possibilities to model such relevance decay are:

$$\tau_{exp}(r) = (1-\delta)^{t(r)} \tag{19}$$

$$\tau_{lin}(r) = \max(1 - t(r)\cdot\delta, 0) \tag{20}$$

with (19) for an exponential type of decay and (20) for a linear decay down to a minimum of zero, whereby $\delta \in [0,1]$ is a decay constant that controls the speed of decay.

This concept can also be described as a kind of *memory* of the relevance feedback engine. The higher the decay factor δ, the faster the system forgets what has been chosen in an earlier step. If we set δ=1 then our approach would only consider the user's latest relevance decision in its relevance score calculation. The value of δ=0 would deactivate the decay completely. Values of δ in between those bounds activate a gradual decay. Using the time-weighted importance we refine our definition of the similarity of a vector to a set R and yield

$$\begin{aligned}&\mathrm{sim}_{rs}(\vec{r}, R) = \Omega(\{\tau(\vec{s}_1)\mathrm{sim}(\vec{r}, \vec{s}_1),\\ &\ldots, \tau(\vec{s}_m)\mathrm{sim}(\vec{r}, \vec{s}_m)\})\end{aligned} \tag{21}$$

Relevance Scoring

Based on the similarity measure we defined in the last section we can develop a notion of a rule's pairwise score, i.e. its relevance score with respect to a certain rule that was marked as relevant. While in information retrieval it is mostly assumed that those documents which are similar to (non-)relevant ones are (non-)relevant too, we use a slightly different approach.

For rules marked as relevant we assume that once a user has seen such a rule rather than being interested in similar ones his attention is attracted by those which are similar in certain features but dissimilar in others. This means, a user aims for rules which have an element of surprise. For example, a rule could have a very similar antecedent, but a rather dissimilar consequent when compared to a relevant one. It would therefore be surprising to a user because it is an exception to his previous knowledge. This approach also captures the case of rule contradiction employed by other authors (Liu et al., 1997; Padmanabhan and Tuzhilin, 2002), albeit in a fuzzy, less crisp way.

Table 1 shows three of such interesting combinations of rule features. The case discussed above is named C_1 in this table. Another example is C_2. It assigns a high score to those rules that are very different in their symbolic representation, but exhibit a similar time series. Such a combina-

Table 1. Interestingness matrix

		dissimilar		
		consequent	time series	symbolic
similar	antecedent	C_1	-	-
	time series	-	-	C_2
	symbolic	-	C_3	-

tion can hint at an unknown hidden cause for the observed changes, which in turn are of interest to a user who typically will assume that only similar rules change similarly. The remaining entry C_3 is basically the inversion of the last one. A rule is considered interesting if it is similar to a relevant one, but has a very dissimilar time series.

For rules marked as non-relevant we use an approach similar to the one used in information retrieval, i.e. rules that are similar to non-relevant ones are also considered non-relevant.

Based on these considerations our calculation of the overall relevance score is split into two parts: one each for the relevant and non-relevant rules, respectively.

Our definition of the relevance of a rule with regard to the set of relevant rules is rather straightforward and shown in , and for the three cases mentioned above. To pick up on our examples from the previous section, using C_1 a rule receives a high relevance score if its antecedent is similar to the rule antecedents in R_r and its consequent dissimilar to the rule consequents in R_r. Likewise, the score for C_2 is calculated by multiplying the similarity of the rule/rule set combination for the time series with the dissimilarity of the rule/rule set combination for the symbolic representation.

$$C_1 \ : \ \Phi(\vec{r}, R_r) = \mathrm{sim}_{rs}(\vec{r}_{body}, R_r)\mathrm{dissim}_{rs}(\vec{r}_{head}, R_r) \tag{22}$$

$$C_2 \ : \ \Phi(\vec{r}, R_r) = \mathrm{sim}_{rs}(\vec{r}_{time}, R_r)\mathrm{dissim}_{rs}(\vec{r}_{sym}, R_r) \tag{23}$$

$$C_3 \ : \ \Phi(\vec{r}, R_r) = \mathrm{sim}_{rs}(\vec{r}_{sym}, R_r)\mathrm{dissim}_{rs}(\vec{r}_{time}, R_r) \tag{24}$$

For the non-relevant rules we assume that rules in R_n specify a subspace of the rule space where more non-relevant rules are located. To direct the user away from this subspace, rules that are far away from it will receive a higher score, whereas those in the vicinity will receive a low score. An unrated rule r should therefore receive a high interestingness score the more dissimilar it is from the set of non-relevant rules, i.e.

$$\Psi(\vec{r}, R_n) = \mathrm{dissim}(\vec{r}, R_n) \tag{25}$$

Our final relevance score of an unrated rule r under consideration of the set of relevant and (non-) relevant rules consists of two parts, $\Phi(\vec{r}, R_r)$ and $\Psi(\vec{r}, R_n)$, which are both weighted to give the user more influence on the scoring.

$$F(\vec{r}, R_r, R_n) = w_{rel}\Phi(\vec{r}, R_r) + w_{nrel}\Psi(\vec{r}, R_n) \tag{26}$$

After every feedback cycle, i.e. after every update of R_r or R_n, each unrated rule r is being re-evaluated whereby a new score $F(\vec{r}, R_r, R_n)$ is assigned. Rules which previously have been ranked as rather non-relevant can now receive a higher score whereas others may lose their relevance.

EVALUATION

To evaluate the proposed framework we applied it to a data set from the customer relationship management domain in the context of a telecommunication company (Boettcher et al 2006; Boettcher et al, 2007). In particular we were looking into using association rule mining for detecting interesting changes of customer segments in data. Customer segmentation is the process of dividing customers into homogeneous groups on the basis of common attributes. Here, we define a customer segment as a set of customers who have certain features or attributes in common. Given a data set which describes customers any attribute value combination of each subset of its attributes therefore qualifies as a candidate customer segment. Thus, an association rule's antecedent can be seen as a customer segment whereas its consequent can be seen as one of its properties. Picking up our earlier example suppose that the following association rule has been discovered:

$$r\ : \mathrm{Age} > 50, \mathrm{Gender} = \mathrm{Male} \rightarrow \mathrm{Complain} = \mathrm{Yes}$$

The antecedent of this rule describes the segment of customers which are over 50 years old and male. The support of this rule is the relative frequency of customers, who fall within this segment, i.e., it describes the relative size of a customer group. The confidence of this rule, in contrast, can be interpreted as the relative frequency of customers within the group of over 50 year old, male customers who did complain about something, i.e., it describes the frequency a certain property has within a group.

To evaluate our framework we extracted a representative dataset from the company's data warehouse. The dataset contains answers of customers to a survey collected over a period of 40 weeks. Each tuple is described by 33 nominal attributes with a domain size between 2 and 39. We transformed the dataset into a transaction set by recoding every (attribute, attribute value) combination as an item. Then we split the transaction set into 20 subsets, each corresponding to a period of two weeks. The subsets contain between 1480 and 2936 transactions. To each subset we applied the well-known apriori algorithm (Agrawal et al, 1993). From the obtained 20 rule sets we created a compound rule set by intersecting them. Its size is 77401 for the parameters $supp_{min} = 0.05$ and $conf_{min} = 0.2$, respectively. Subsequently we applied the proposed framework. Thereby we will first focus on two objectives within our evaluation. First, the number of trends and stabilities contained in histories has to be determined. Second, the number of derivable rule histories has to be determined.

The results for trend are shown in Table 2 whereby we only show the results for the Mann-Kendall test. Furthermore, the results for stability detection are included, since they depend on the outcome of a prior test for trend. Roughly 50% of support histories exhibit a trend, whereas the number of confidence histories with a trend is considerably smaller. On the other hand, around 30% of confidence histories are stable, compared to fewer than 3% for support. The significant difference can be explained with the density of the data. Since some items are highly correlated, it is very likely that many rules have a stable history of high confidence values. The support history of such rules, nonetheless, may exhibit a trend.

Only histories which exhibit change patterns were tested if they are a derivative of another history. The first row of Table 3 shows the obtained results for trends separately for support and confidence histories. As it can be seen between 40.7% (for confidence) and 66.3% (for support) of the histories are derivable. The second row shows that these numbers are considerably smaller for stable histories; ranging from 26.7% (for confidence) to 39.6% (for support).

Table 2. Fraction of histories with trend and stability

	Trend(%)			Stability(%)
	Down	Up	All	
conf	21.2	18	39.3	28.1
supp	37.5	17.1	54.7	2.6

Table 3. Fraction of derivable histories among all histories which have a trend or are stable

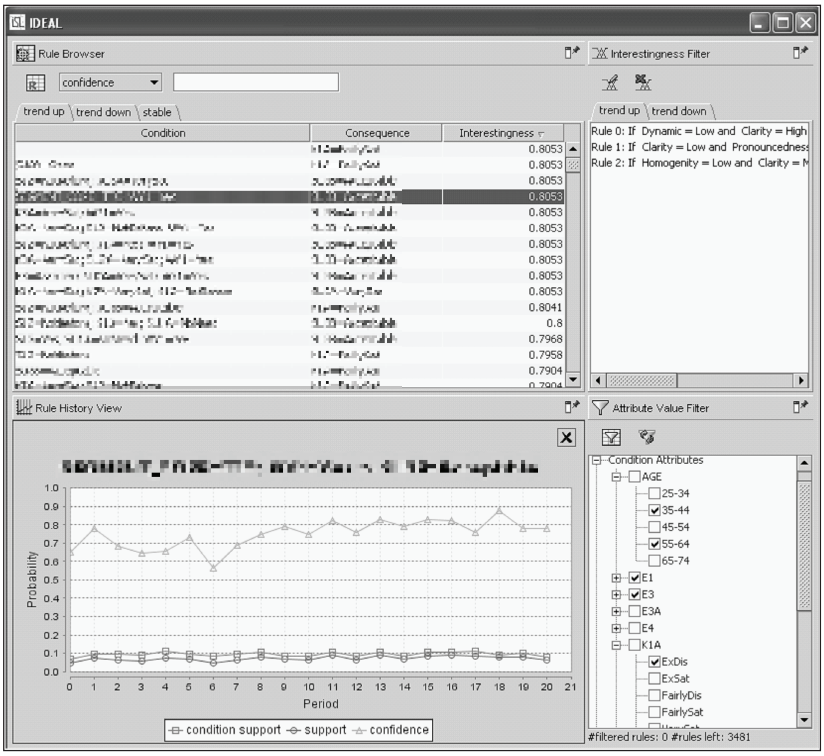
	Support		*Confidence*	
Pattern	*#Histories*	*Derivable(%)*	*#Histories*	*Derivable(%)*
Trend	*42307*	*66.3*	*30387*	*40.7*
Stable	*2019*	*39.6*	*21753*	*26.7*

Proving the utility of methods for interestingness assessment is rather difficult due to a lack of suitable public benchmark dataset's which in our case must contain time-stamped data. Moreover, interestingness is a highly subjective matter influenced by the role, experience and task of the person who does the actual evaluation which renders it problematic to use a precision-recall-based approach on the basis of predetermined interesting patterns. For these reasons we decided to trial our methods for interestingness assessment within a real business scenario of a telecommunication company with experienced users as the test persons in order to measure their acceptance and receive feedback.

We developed a user interface to display the ranked rules to a user, to allow for rule browsing, and to gather relevance feedback. The user interface is shown in Figure 5. Due to reasons of data protection those parts which reveal discovered rules are obfuscated. Its main part consists of a list of rules which are sorted by interestingness. The user can choose whether the rules are being ranked by a change-based objective measure or by the relevance feedback received so far. Here, the underlying assumption is that the user starts with an objective interestingness rating to get a first overview about the discovered associations and to have a starting point for building up his notion of what is relevant and what is not. The user can filter the list by change pattern and rule measure. For instance, he could choose to display only rules which exhibit an upward trend in their confidence history, or rule which are stable in their support history. At any time he can select a rule as relevant or non-relevant using a context menu. The user can access the rules rated so far through another window which is not shown in Figure 5. He can also use this view to revise earlier rating decisions. Double-clicking on a rule opens a chart in the lower left part of the user interface which displays support and confidence histories. On the right hand side of the user interface there are several filters which allows for restricting the displayed rules based on the items contained in them.

The users with whom we trialed our interestingness framework are experienced analysts. Part of their role is to analyze the customer survey data we also used for our experiments on a regular basis by off-the-shelf business intelligence and statistics tools. Therefore the acceptance criterion for our framework and the prototype was whether they would discover anything completely novel or unexpected patterns, and the ease of relating the obtained knowledge to current business opera-

Figure 5. User Interface used for the Trial. For reasons of data protection rules are obfuscated



tions. The trial is still going on but we will the some feedback we received so far is very promising: The users reported that they did discover novel patterns and they also pointed out that they could easily reproduce why certain rules had been judged as interesting by the system. Most of them discovered interesting and unexpected rules within a few feedback cycles. There was very positive feedback that the change information is readily available and that the objective interestingness measures rank the most dramatic changes rather high. With other tools they first had to identify certain patterns and then manually trace their change over time. Since this approach is purely hypothesis driven it is clear that interesting trends might be missed if the analyst does not expect that these trends occur. For example, several rules could be identified whose change was rather unexpected. Further investigation showed that the trend in those rules was linked to a campaign (the trend started with the beginning of the campaign) and represented a positive, yet unexpected side-effect.

FUTURE TRENDS

The idea of assessing the interestingness of an association rule by analyzing histories of support and confidence can be traced back to the early days of association rule mining itself (Agrawal and Psaila, 1995; Chakrabarti et al., 1998). Still, it only recently received increasing attention in research publications (Liu et al., 2001b; Baron et al, 2003; Boettcher et al., 2006) and large scale business applications (Boettcher et al., 2007). Because it is an evolving field there are many challenges which still need to be addressed and solved. We have identified two areas that we believe merit future work which would significantly enhance the discussed approaches.

Incremental algorithms. At the moment for each association rule at least one, often more histories have to be processed to detect, for example, a trend. Currently, each time the history is extended the same processing is repeated without taking advantage of prior results. Here, it would be advantageous to investigate or adapt incremental algorithms to reduce the computational complexity while speeding up, e.g., the discovery of trends.

Business alignment. Businesses often do need knowledge about change to monitor how their decisions impact their business. For example, when starting a new marketing campaign a business wants to know how it impacts its customers both in terms of desired (like increasing sales figures) and undesired (like decreasing satisfaction in certain customer groups) effects. Clearly, an association rule is particularly interesting if its change can be related to recent business decisions. Such a *business-aligned* interestingness assessment would be objective in the sense that the required data about decisions and campaigns has not to be collected by a user but is often electronically available in corporate databases and document management systems. Still, change mining for association rules is only a first step in this direction; it also involves fields as diverse as data mining across heterogeneous sources, time series analysis, and maybe semantic technologies.

CONCLUSION

This article dealt with the cornerstones of a comprehensive interestingness assessment framework for association rules which provides a unified handling of objective and subjective interestingness measures based on a notion of rule change. In the first part we introduced our idea of change mining of association rules and showed how it can be used to derive objective interestingness measures which are meaningful to a user and can be justified from a business perspective. These measures assign rules high ranks which most urgently require intervening or supporting actions to be taken. Having provided a user with a first impression of the rules discovered we introduced our idea of relevance feedback on association rules. This approach accounts for the fact that a user's perception of relevance during the exploration process is a dynamic rather than a static process. Our approach is inspired by well-known methods from the area of Information Retrieval. In particular, we processed the rules to yield a vector notation that unifies a rule's symbolic representation with its (numeric) change information. Based on this representation we proposed our relevance feedback method which can be configured to account for the element of surprise when exploring a rule set. This notion of surprise was defined as the dissimilarity of a newly encountered rule with the set of previously seen ones. Overall, our unified approach to interestingness assessment can greatly improve on the usability and the practicability of any association rule mining process by post-processing the rules accordingly and incorporating user feedback.

REFERENCES

Agrawal, R., Imieliński, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. ACM *SIGMOD Record* 22(2),207-216

Agrawal, R., & Psaila, G. (1995). Active data mining. In U. M. Fayyad, & R. Uthurusamy, (Eds.), *Proceedings of the 1st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (pp. 3–8), Montreal, Quebec, Canada. AAAI Press, Menlo Park, CA, USA.

Au, W.-H., & Chan, K. (2005). Mining changes in association rules: A fuzzy approach. *Fuzzy Sets and Systems*, *149*(1), 87–104.

Baeza-Yates, R. A., & Ribeiro-Neto, B. A. (1999). *Modern Information Retrieval*. ACM Press / Addison-Wesley.

Baron, S., Spiliopoulou, M., & Günther, O. (2003). Efficient monitoring of patterns in data mining environments. In *Proc. of 7th East-European Conf. on Advances in Databases and Inf. Sys. (ADBIS'03)*, LNCS, (pp. 253–265). Springer.

Bayardo, Jr., R. J., & Agrawal, R. (1999). Mining the most interesting rules. *In Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (pp. 145–154).

Boettcher, M., Nauck, D., Ruta, D., & Spott, M. (2006). Towards a framework for change detection in datasets. In *Proceedings of the 26th SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence*, (pp. 115–128). Springer.

Boettcher, M., Nauck, D., & Spott, M. (2005). Detecting Temporally Redundant Association Rules. In *Proceedings of the 4th International Conference on Machine Learning and Applications*, (pp. 397-403). IEEE Computer Science Press.

Boettcher, M., Spott, M., & Nauck, D. (2007). A framework for discovering and analyzing changing customer segments. In *Proceedings of the 7th Industrial Conference on Data Mining (ICDM'2007)*, LNAI 4597, (pp. 255–268). Springer.

Chakrabarti, S., Sarawagi, S., & Dom, B. (1998). Mining surprising patterns using temporal description length. In *Proceedings of the 24th International Conference on Very Large Databases*, (pp. 606–617). Morgan Kaufmann Publishers Inc.

Chatfield, C. (2003). *The Analysis of Time Series: An Introduction, Sixth Edition (Texts in Statistical Science).* Chapman & Hall/CRC.

Chen, M.-C., Chiu, A.-L., & Chang, H.-H. (2005). Mining changes in customer behavior in retail marketing. *Expert Systems with Applications*, *28*(4), 773–781.

Cox, D., & Stuart, A. (1955). Some quick sign tests for trend in location and dispersion. *Biometrika*, *42*, 80–95.

Dong, G., & Li, J. (1999). Efficient mining of emerging patterns: Discovering trends and differences. *In Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (pp. 43–52).

Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., & Uthurusamy, R. (1996). *Advances in Knowledge Discovery and Data Mining.* AAAI Press and MIT Press, Menlo Park and Cambridge, MA, USA.

Fayyad, U. M., Piatetsky-Shapiro, G., & Uthurusamy, R. (2003). Summary from the KDD-03 panel: Data mining–the next 10 years. *SIGKDD Explorations Newsletter*, *5*(2), 191–196.

Fogel, D. B. (1997). The advantages of evolutionary computation. In D. Lundh, B. Olsson, & A. Narayanan, (Eds.), *Bio-Computing and Emergent Computation*. World Scientific Press, Singapore.

Geng, L., & Hamilton, H. J. (2006). Interestingness measures for data mining: A survey. *ACM Computing Surveys, 38*(3), 9.

Jaakkola, T., & Siegelmann, H. (2001). Active information retrieval. In *Advances in Neural Information Processing Systems, 14*, 777–784. MIT Press.

Kimball, R. (1996). *Data Warehouse Toolkit: Practical Techniques for Building High Dimensional Data Warehouses*. John Wiley & Sons.

Liu, B., Hsu, W., & Chen, S. (1997). Using general impressions to analyze discovered classification rules. In *Proceedings of the 3rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (pp. 31–36).

Liu, B., Hsu, W., Chen, S., & Ma, Y. (2000). Analyzing the subjective interestingness of association rules. *IEEE Intelligent Systems*, *15*(5), 47–55.

Liu, B., Hsu, W., & Ma, Y. (2001a). Discovering the set of fundamental rule changes. In *Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (pp. 335–340).

Liu, B., Ma, Y., & Lee, R. (2001b). Analyzing the interestingness of association rules from the temporal dimension. In *Proceedings of the IEEE International Conference on Data Mining*, (pp. 377–384). IEEE Computer Society.

Mann, H. (1945). Nonparametric tests against trend. *Econometrica*, *13*, 245–259.

McGarry, K. (2005). A survey of interestingness measures for knowledge discovery. *The Knowledge Engineering Review, 20*(1), 39-61. Cambridge University Press.

Padmanabhan, B. & Tuzhilin, A. (1999). Unexpectedness as a measure of interestingness in knowledge discovery. *Decision Support Systems*, *27*.

Padmanabhan, B., & Tuzhilin, A. (2000). Small is beautiful: discovering the minimal set of unexpected patterns. In *Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (pp. 54–63).

Padmanabhan, B., & Tuzhilin, A. (2002). Knowledge refinement based on the discovery of unexpected patterns in data mining. *Decision Support Systems*, *33*(3), 309–321.

Piatetsky-Shapiro, G. (2000). Knowledge discovery in databases: 10 years after. *SIGKDD Explorations Newsletter*, *1*(2), 59–61.

Piatetsky-Shapiro, G., & Matheus, C. J. (1994). The interestingness of deviations. In *Proceedings of the AAAI Workshop on Knowledge Discovery in Databases*, (pp. 25–36).

Ruß, G., Boettcher, M., & Kruse, R. (2007). Relevance feedback for association rules using fuzzy score aggregation. In *Proc. Conf. North American Fuzzy Information Processing Society (NAFIPS 2007)*, (pp. 54–59).

Ruß, G., Kruse, R., Nauck, D., & Boettcher, M. (2008). Relevance feedback for association rules by leveraging concepts from information retrieval. In M. Bramer, (Ed,), *Research and Development in Intelligent Systems*, *24* of *Proceedings of AI-2007*, (pp. 253–266). BCS SGAI, Springer.

Salton, G. (1971). *The SMART Information Retrieval System*. Englewood Cliffs, NJ: Prentice Hall.

Salton, G., & Buckley, C. (1987). Term weighting approaches in automatic text retrieval. *Information Processing and Management*, *5*(24), 513– 523.

Silberschatz, A., & Tuzhilin, A. (1996). What makes patterns interesting in knowledge discovery systems. *IEEE Transactions on Knowledge and Data Engineering*, 8(6), 970–974.

Tan, P., & Kumar, V. (2000). *Interestingness measures for association patterns: A perspective*. Technical Report 00-036, Department of Computer Science, University of Minnesota.

Tan, P.-N., Kumar, V., & Srivastava, J. (2004). Selecting the right objective measure for association analysis. *Information Systems*, *29*(4), 293–313.

Wang, K., Jiang, Y., & Lakshmanan, L. V. S. (2003). Mining unexpected rules by pushing user dynamics. In *Proceedings of the 9th ACM*

SIGKDD International Conference on Knowledge Discovery and Data Mining, (pp. 246–255).

Webb, G. I. (2000). Efficient search for association rules. In *Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (pp. 99–107).

Yager, R. R. (1988). On ordered weighted averaging aggregation operators in multicriteria decisionmaking. *IEEE Trans. Syst. Man Cybern., 18*(1), 183–190.

Yager, R. R. (1997). On the inclusion of importances in owa aggregations. In *The ordered weighted averaging operators: theory and applications*, (pp. 41–59), Norwell, MA, USA. Kluwer Academic Publishers.

Zhang, X., Dong, G., & Kotagiri, R. (2000). Exploring constraints to efficiently mine emerging patterns from large high-dimensional datasets. In *Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (pp. 310–314).