Interestingness of Discovered Association Rules in terms of Neighborhood-Based Unexpectedness *

Guozhu Dong and Jinyan Li

The University of Melbourne, Department of Computer Science

Parkville, Vic 3052, Australia. Email: {dong,jyli}@cs.mu.oz.au.

Technical Report 97/24

October, 1997

Abstract

One of the central problems in knowledge discovery is the development of good measures of interestingness of discovered patterns. With such measures, a user needs to manually examine only the more interesting rules, instead of each of a large number of mined rules. Previous proposals of such measures include rule templates, minimal rule cover, actionability, and unexpectedness in the statistical sense or against user beliefs.

In this paper we will introduce neighborhood-based interestingness by considering unexpectedness in terms of neighborhood-based parameters. We first present some novel notions of distance between rules and of neighborhoods of rules. The neighborhood-based interestingness of a rule is then defined in terms of the pattern of the fluctuation of confidences or the density of mined rules in some of its neighborhoods. Such interestingness can also be defined for sets of rules (e.g. plateaus and ridges) when their neighborhoods have certain properties. We can rank the interesting rules by combining some neighborhood-based characteristics, the support and confidence of the rules, and users' feedback. We discuss how to implement the proposed ideas and compare our work with related ones. We also give a few expected tendencies of changes due to rule structures, which should be taken into account when considering unexpectedness. We concentrate on association rules and briefly discuss generalization to other types of rules.

*This work was supported in part by a research grant from the Australian Research Council.

1 Introduction

Data mining is concerned with the extraction of previously unknown and potentially useful high-level knowledge in the form of patterns from a huge mass of data. This area has received extensive attention from the research community and industry recently. When the amount of such high level knowledge is large, which is typically the case for association rules [1], the selection of interesting patterns becomes a serious problem for the human user; we will call this problem the post-mining rule analysis problem. Thus one of the central problems in the field of data mining is the development of good measures of interestingness of discovered patterns. With such measures, we can develop algorithms to help find the interesting rules from the mined rules. The results of mining can then be made more usable, without the need of going through all the mined rules manually. Previous proposals of interestingness measures include: rule templates [5, 4] for limiting attention to only those rules that match the templates, minimal rule covers [12] where rules implied by those presented to the user are eliminated, actionability of rules [8, 10] (some benefit can be obtained by doing something), and unexpectedness of rules [6, 11]. Unexpectedness has been interpreted either in the statistical sense, as having higher chance than that under the independence assumption or as having higher chance than some threshold, or against user beliefs.

We believe that one should take the neighborhood of the rule into account when considering unexpectedness. Using mountains as an analogy, normally one would not say that all peaks of the Himalayas Range of height > 4000 meters are more interesting than the highest mountain in North America and Japan, although these peaks are higher than the highest mountain in North America and Japan. Indeed, the interestingness of a mountain depends on its height as well as on its position in its neighborhoods; indeed, Mount Fuji of Japan is famous because there are no comparable peaks in its neighborhood. In the terminology of association rules, the interestingness of a rule should depend on its confidence as well as on the degree of the confidence fluctuation in its neighborhoods and the density of mined rules there. The purpose of this paper is to introduce the neighborhood-based unexpectedness and to examine interestingness in terms of such unexpectedness.

We first present some novel notions of distance between rules and of neighborhoods of rules. The interestingness of a rule is then defined in terms of the pattern of the fluctuation of confidences or the density of mined rules in some of its neighborhoods. Such interestingness can also be defined for sets of rules (e.g. plateaus and ridges) when their neighborhoods have certain properties. We rank the interesting rules by combining some neighborhood-based characteristics, the support and confidence of these rules, and users' feedback. We discuss how to implement the proposed ideas and compare our work with related works. We also give a few expected tendencies of changes due to rule structures, which should be taken into account when considering unexpectedness. We will mainly concentrate on association rules and will briefly discuss generalization to other types of rules (by defining appropriate distance functions).

For the case of association rules, one might argue that it is possible to solve the post-mining rule analysis problem by increasing the thresholds – the number of mined rules will then decrease. Continuing with the world map analogy, one can easily see that as a result of such an approach only those global peaks will be shown to the user, thus missing the useful information conveyed by those local peaks over vast plains.

Our neighborhood-based interestingness is proposed as a complementary measure to those proposed previously in the literature, including those cited at the beginning of the introduction and the following. The issue of interestingness of general discoveried knowledge was discussed in [7]. Interestingness can also be measured in terms of the statistical strength of a pattern such as confidence and support [2], and ancestor rules' confidences [9]. The problem of how to find patterns satisfying multiple criteria of interestingness has been investigated in [13]. A comprehensive survey about measurements of interestingness can be found in [11].

The rest of this paper is organized as follows. In Section 2, we present the preliminaries of association rules. In Section 3 we introduce the notion of distance among rules. In Section 4 we define neighborhoods of rules. In Section 5 we define interestingness of rules in terms of neighborhood-based unexpectness, and similarly interestingness of sets of rules. In Section 6 we present some expected changes. In Section 7 we discuss how to rank interesting rules. In Section 8 we cover some implementation issues. In Section 9 we compare our work with related ones. In Section 10 some concluding remarks are given.

2 Preliminaries on association rules

Let $I = \{i_1, i_2, \cdots, i_n\}$ be a set of literals, called *items*. Let $\mathcal{D}$ be a set of transactions, where each transaction[1] is a subset T of I. Given a set of items X from I, we say a transaction T *contains*, or *matches*, X if $X \subseteq T$; let $m(X)$ denote the set of *transactions* in $\mathcal{D}$ which match X.

An association rule R is an implication of the form $X \rightarrow Y$, where $X \subseteq I$, $Y \subseteq I$, and $X \cap Y = \emptyset$. The *support* of R, denoted as *support*(R), is $\frac{|m(XY)|}{|\mathcal{D}|}$: the percentage of

[1]A transaction is also referred to as a basket in the literature.

transactions in $\mathcal{D}$ which contain[2] XY. The *confidence* of R, denoted as $conf(R)$, is $\frac{|m(XY)|}{|m(X)|}$: the percentage of transactions containing X which also contain Y. (So confidence is defined only if $m(X) > 0$.)

The task of mining association rules is to find all association rules whose supports and confidences are larger than, respectively, some given minimum support threshold, *min_support*, and some minimum confidence threshold, *min_confidence*.

Given the item set I, there is a unique set of potential association rules. Given different thresholds for support and confidence, different sets of rules can be mined from a given set of transactions.

3 Distance definition

The central theme of this paper is to consider interestingness of rules in terms of neighborhood-based unexpectedness. To this end, we need to have *some* distance functions between rules. Such distance can be defined in more than one way, including semantics-based ways and syntax-based ways; we will review a semantics-based distance and introduce a syntax-based one. The syntax-based definition has several parameters which can be adjusted to suit the need of particular applications.

3.1 A semantics-based distance

A semantics-based distance between association rules was given in [12]. It measures the difference between rules in terms of their sets of matching rows. More specifically, the *matching-set distance* between two rules $X_1 \rightarrow Y_1$ and $X_2 \rightarrow Y_2$ is defined as $|m(X_1Y_1)| + |m(X_2Y_2)| - 2 * |m(X_1Y_1X_2Y_2)|$. We will use $Dist_{mset}$ to denote this function.

Using this definition, there can be different rules R_1 and R_2, namely $A \rightarrow B$ and $B \rightarrow A$, such that $Dist_{mset}(R_1, R_2) = 0$. Thus $Dist_{mset}$ is not a metric distance over the set of all potential rules. This may lead to "more crowded" neighborhoods, and may have some effect on neighborhood-based unexpectedness.

[2]Following the tradition of database literature and for the sake of clarity, we write XY for $X \cup Y$ where X and Y are sets of items. We use $|S|$ to denote the cardinality of a set S.

3.2 A syntax-based distance

We now introduce a syntax-based distance, which is intended to measure the item-set difference between rules. This turns out to be a metric distance.

Our distance function is defined in such a way that one can give different scales of importance to differences for different parts of rules. Item-set differences are divided into three parts[3]: (i) the symmetric difference of all items in the two rules, (ii) the symmetric difference of the left-hand sides of the two rules, (iii) the symmetric difference of the right-hand sides.

Definition 3.1 Given three non negative real numbers $\delta_1, \delta_2, \delta_3$, define an *item-set distance* between two rules $R_1 : X_1 \to Y_1$ and $R_2 : X_2 \to Y_2$ as[4]

$$Dist_{iset}(R_1, R_2) = \delta_1 * |(X_1Y_1) \ominus (X_2Y_2)|$$

$$+\delta_2 * |X_1 \ominus X_2| + \delta_3 * |Y_1 \ominus Y_2|.$$

Example 3.2 To illustrate this definition, consider the following rules:

$$R_1 : D \to BC$$
$$R_2 : AD \to BC$$
$$R_3 : BC \to D$$
$$R_4 : BC \to AD$$

Then $Dist_{iset}(R_1, R_2) = \delta_1 + \delta_2$ and $Dist_{iset}(R_3, R_4) = \delta_1 + \delta_3$. Both of $\delta_1 + \delta_2$ and $\delta_1 + \delta_3$ are contributed by A, as items B, C, and D make no contribution to the two distances.

To illustrate the point that different "positional" differences make different contributions to the distance, consider the following rules:

$$R_5 : AB \to CD$$
$$R_6 : ADF \to CE$$

Then $Dist_{iset}(R_5, R_6) = 3 * \delta_1 + 3 * \delta_2 + 2 * \delta_3$. The different δ's are contributed as follows.

- Item A occurs in both rules and occurs on the same sides of the two rules. So A makes no contribution to the distance. The same happens with C.

[3]We can define distance functions using the difference of (i) only. However, there can be different rules, namely $A \to B$ and $B \to A$, such that the distance between them is 0.

[4]Given two sets X and Y, $X \ominus Y$ denotes the symmetric difference between X and Y, i.e., $X - Y \cup Y - X$.

- Item D occurs in both rules but occurs on different sides of the two rules. So D makes a contribution of $\delta_2 + \delta_3$ to the distance.

- Item B occurs in one rule and on its left-hand side; it does not occur in the other rule. So B makes a contribution of $\delta_1 + \delta_2$ to the distance. The same happens with F.

- Item E occurs in one rule, and on its right-hand side; it does not occur in the other rule. So E makes a contribution of $\delta_1 + \delta_3$ to the distance.

∎

Different choice of values for $\delta_1, \delta_2, \delta_3$ can be used to reflect users' preferences. For most of the paper we will set $\delta_1 = 1, \delta_2 = \frac{n-1}{n^2}$ and $\delta_3 = \frac{1}{n^2}$, where $n = |I|$. However, the approach works for other distance functions.

Having $\delta_1 > \delta_2 > \delta_3$ reflects our belief that the three kinds of item-set differences should contribute differently to the distance: the whole difference $(X_1Y_1) \ominus (X_2Y_2)$ is more important than the left-hand side difference $X_1 \ominus X_2$, which in turn is more important than the right-hand side difference $Y_1 \ominus Y_2$.

We set $\delta_2 = \frac{n-1}{n^2}$ and $\delta_3 = \frac{1}{n^2}$ to ensure that (*) rules with identical set of items are closer to each other than to rules with different sets of items. Suppose $R_1 : X_1 \rightarrow Y_1$ and $R_2 : X_2 \rightarrow Y_2$ are two rules such that $X_1Y_1 = X_2Y_2$. Then $Dist_{iset}(R_1, R_2) = \delta_2 * |X_1 \ominus X_2| + \delta_3 * |Y_1 \ominus Y_2|$. We wish to ensure that $Dist_{iset}(R_1, R_2) < Dist_{iset}(R_1, R_3)$, for every rule $R_3 : X_3 \rightarrow Y_3$ where $|X_1Y_1 \ominus X_3Y_3| > 0$ (that is R_3's item set is different from that of R_1's). We use the following example to further illustrate why (*) holds.

Example 3.3 Consider the following rules:

$$R_1 : ABC \rightarrow DE$$
$$R_2' : DE \rightarrow ABC$$
$$R_3' : ABC \rightarrow D.$$

Then $Dist_{iset}(R_1, R_2') = 0 * \delta_1 + 5 * \delta_2 + 5 * \delta_3 = \frac{5(n-1)}{n^2} + \frac{5}{n^2} = \frac{5}{n} \leq 1 < Dist_{iset}(R_1, R_3') = 1 * \delta_1 + 1 * \delta_3 = 1 + \frac{1}{n^2} = \frac{n^2+1}{n^2}$. It is easy to see that R_2' is a rule farthest away from R_1, among all R_2 containing the same set of items as R_1. Furthermore, R_3' is a rule closest to R_1, among all R_3 containing an item set different from that of R_1's. Therefore $Dist_{iset}(R_1, R_2) < Dist_{iset}(R_1, R_3)$ for all R_2 and R_3 satisfying the conditions given above. ∎

The $Dist_{iset}$ distance behaves like the usual distances we know:

Proposition 3.4 The $Dist_{iset}$ distance is a metric distance over the set of all potential rules. That is, the following properties hold for all rules R_1, R_2 and R_3:

1. $Dist_{iset}(R_1, R_1) = 0$,
2. $Dist_{iset}(R_1, R_2) > 0$ if $R_1 \neq R_2$,
3. $Dist_{iset}(R_1, R_2) = Dist_{iset}(R_2, R_1)$,
4. $Dist_{iset}(R_1, R_3) \leq$
   $Dist_{iset}(R_1, R_2) + Dist_{iset}(R_2, R_3)$.

The proofs of (1–3) are easy, and that of (4) can be given by utilizing a vector representation of item sets (after fixing an ordering on elements of I).

As an aside, observe that the potential rule space may not be dense in the sense that every "realizable" distance is the sum of two "realizable" distances. That is, there can be rules R_1 and R_2 such that $Dist_{iset}(R_1, R_2) < Dist_{iset}(R_1, R_3) + Dist_{iset}(R_2, R_3)$ for every rule R_3; this happens for $R_1 : AB \to C$ and $R_2 : AB \to CD$ and for $R_1 : AB \to C$ and $R_2 : A \to BC$. Sometimes, though, a "realizable" distance is the sum of two "realizable" distances: There can be rules R_1, R_2 and R_3 such that $Dist_{iset}(R_1, R_2) = Dist_{iset}(R_1, R_3) + Dist_{iset}(R_2, R_3)$; for example, $R_1 : AB \to CDE$, $R_2 : B \to CD$ and $R_3 : AB \to CD$.

3.3 Other variants of $Dist_{iset}$

We now suggest some variants of the syntax-based distance by setting $\delta_1, \delta_2, \delta_3$ differently; these can be useful for different user preferences, though not used in the sequel.

If we want to emphasize the changes on the left-hand side of rules, we can set $\delta_1 = 0$, $\delta_2 = 1$, and $\delta_3 = \frac{1}{n+1}$. For any two rules $R_1 : X_1 \to Y_1$ and $R_2 : X_2 \to Y_2$ we have

$$Dist_{iset}(R_1, R_2) = |X_1 \ominus X_2| + \frac{1}{n+1}|Y_1 \ominus Y_2|.$$

If we want to emphasize the changes on the right-hand side of rules, we can set $\delta_1 = 0$, $\delta_2 = \frac{1}{n+1}$, and $\delta_3 = 1$. For any two rules $R_1 : X_1 \to Y_1$ and $R_2 : X_2 \to Y_2$ we have

$$Dist_{iset}(R_1, R_2) = \frac{1}{n+1}|X_1 \ominus X_2| + |Y_1 \ominus Y_2|.$$

3.4 Distance on other types of rules

We can also define distances for rules of other types. For Datalog (or Horn clauses), we can define distance between two rules in terms of their largest unifiable parts. One can use area of overlap for defining distances between interval-based rules such as $< Age : 20..30 > \Rightarrow < car : 1..2 >$ (8% $support, 70\%$ $confidence$).

4 Neighborhoods

In this section, we introduce the notion of neighborhoods of rules. These will be used to define interestingness in the next section.

Definition 4.1 An *r-neighborhood* of a rule R_0 ($r > 0$), denoted as $N(R_0, r)$, is the following set

$$\{R : Dist_{iset}(R, R_0) \leq r, R \text{ a potential rule}\}.$$

Although other types of neighborhoods are possible, in this paper we will only be concerned with neighborhoods defined by circles.

One example type of neighborhoods is the 1-neighborhoods such as $N_{R_0,1}$.

Example 4.2 Suppose $I = \{A, B, C, D, E\}$. Then the 1-neighborhood of the rule $AB \rightarrow CD$ consists of the following rules:

$A \rightarrow BCD$	$B \rightarrow ACD$	$C \rightarrow ABD$
$D \rightarrow ABC$	$AB \rightarrow CD$	$AC \rightarrow BD$
$AD \rightarrow BC$	$BC \rightarrow AD$	$BD \rightarrow AC$
$CD \rightarrow AB$	$ABD \rightarrow C$	$ABC \rightarrow D$
$ACD \rightarrow B$	$BCD \rightarrow A$	

In general, as a consequence of the way $Dist_{iset}$ is defined, all rules in a 1-neighborhood have the same item set as the center and consequently they all have the same support. ∎

We will also talk about interestingness of collections of rules in terms of their neighborhoods. For example, we can consider the interestingness of the collection of rules in $N_{R_0,1}$ in terms of the set of rules in $N_{R_0,2} - N_{R_0,1} = \{R : 1 < Dist_{iset}(R, R_0) \leq 2, R \text{ a potential rule}\}$. One can also view this set as the union of 1-neighborhoods of all rules whose item set differ from that of the center by exactly one item.

5 Interestingness of rules

In this section we introduce several neighborhood-based interestingness. One of these is in terms of unexpected confidence, and the other is in terms of unexpected density. We then define neighborhood-based interestingness of sets of rules. We also discuss the need for distinguishing the true unexpected changes from the inherent changes in confidence and support due to rule structures.

Similarly one can consider interestingness in terms of unexpected support, especially for collections of rules within some 1-neighborhoods of rules. The details of these are omitted.

5.1 Interesting rules with unexpected confidence

To capture "unexpected confidence", we need to introduce two measures of the fluctuation of the confidences of mined rules in a neighborhood: *average confidence* and *standard deviation* of confidence.

Suppose M is a set of mined rules for given minimum support and confidence thresholds $min\_support$ and $min\_confidence$, R_0 is a mined rule in M and $r > 0$.

- The *average confidence* of the r-neighborhood of R_0 is defined as the average of the confidences of rules in the set $M \cap N(R_0, r) - \{R_0\}$; we use $avg\_conf(R_0, r)$ to denote this value.
- The *standard deviation* of the r-neighborhood of R_0 is defined as the standard deviation of the confidences of rules in the set $M \cap N(R_0, r) - \{R_0\}$; we will use $std\_conf(R_0, r)$ to denote this value.

When the set $M \cap N(R_0, r) - \{R_0\}$ is empty, we choose to define these two values as zero, although other choices are possible.

Observe that the value $std\_conf(R_0, r)$ gives the average fluctuation of confidences in the r-neighborhood of R_0.

We choose to identify unexpected confidence of a rule R_0 in its r-neighborhood with the condition that $|conf(R_0) - avg\_conf(R_0, r)|$ is much larger than $std\_conf(R_0, r)$. This is the basis of our first interestingness in terms of neighborhood-based unexpectedness.

Definition 5.1 A rule R_0 is said to be *interesting*, of the *unexpected confidence* type, in its r-neighborhood if $||conf(R_0) - avg\_conf(R_0, r)| - std\_conf(R_0, r)|$ is large[5]; in other

[5] The meaning of large can be specified by a threshold.

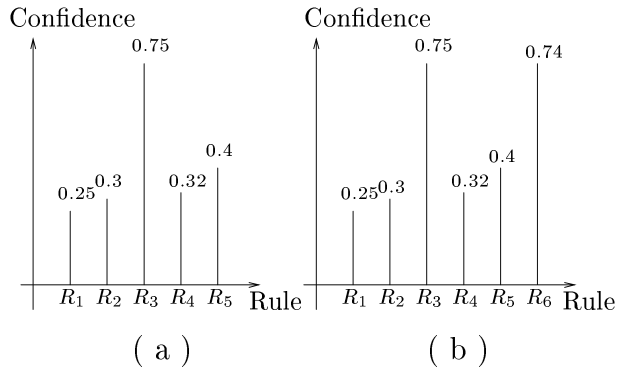

Figure 1: Unexpected Confidence

words, if the confidence of R_0 deviates from the $avg\_conf(R_0, r)$ much more than the average deviation.

Example 5.2 Figure 1(a) shows that this definition indeed captures rules with unexpected confidence in a neighborhood. Suppose there are only five mined rules $R_1, \cdots, R_5$ in some r-neighborhood, whose confidences are 0.25, 0.3, 0.75, 0.32, and 0.4 respectively. Then $avg\_conf(R_3, r) = 0.3175$ and $std\_conf(R_3, r) = (((0.25 - 0.3175)^2 + (0.3 - 0.3175)^2 + (0.32 - 0.3175)^2 + (0.4 - 0.3175)^2)/4)^{\frac{1}{2}} = 0.026$. Thus, $conf(R_3) - avg\_conf(R_3, r) = 0.4325$, a difference which is about 16 times as large as $std\_conf(R_3, r)$. So, the confidence of R_3 is unexpected in its r-neighborhood. ∎

The value of $||conf(R_0) - avg\_conf(R_0, r)| - std\_conf(R_0, r)|$ was chosen because it avoids some deficiencies of two other possible alternatives. (i) Defining a rule R_0 as having unexpected confidence if it achieves near maximum difference of confidence between pairs of rules in their r-neighborhood – if there exists $R' \in M \cap N(R_0, r)$ such that $|conf(R_0) - conf(R')| \approx max\{|conf(R_1) - conf(R_2)| : R_1, R_2 \in M \cap N(R_0, r)\}$. This suffers from the inability to differentiate between the unexpected minority and the prevailing majority. For example, in Figure 1(a), R_1, R_2, R_3, R_4 (the prevailing majority) will then all be considered as having unexpected confidence. (ii) Defining a rule R_0 as having unexpected confidence if it has no competitors in confidence among rules in a neighborhood – if $max\{|conf(R_0) - conf(R_1)| : R_1 \in M \cap N(R_0, r)\}$ is large. This definition is not able to capture rules with unexpected confidence when there are two or more rules with outstanding and equal confidence. For example, in Figure 1(b), neither R_3 not R_6 will then be considered as having unexpected confidence.

The values used in the above definition can be used for ranking the interestingness of rules. For example, the larger the neighborhood the more interesting the rule is, and the larger $||conf(R_0) - avg\_conf(R_0, r)| - std\_conf(R_0, r)|$ the more interesting the rule is. This will be the topic of Section 6.

Rules with unexpected confidence can have around the highest confidence among rules in the neighborhood or the lowest. For the latter, however, caution is needed when the confidence of the rule is not sufficiently larger than the minimum confidence threshold: there might be many potential rules in the neighborhood whose confidences are just below the threshold.

5.2 Interesting rules with sparse neighborhoods

A second kind of rules are considered interesting because there are many potential rules in their neighborhoods but there are very few mined rules there.

Definition 5.3 A rule R_0 is *interesting*, of the *isolated* type, if it has an unexpectedly sparse r-neighborhood: if the number of potential rules in $N(R_0, r)$ is large but and the number of mined rules there, i.e. $|M \cap N(R_0, r)|$, is relatively small.

If we call $\frac{|M \cap N(R_0,r)|}{|N(R_0,r)|}$ the *density* of the r-neighborhood of R_0, then the condition in the above definition can be reworded as "if the number of potential rules in $N(R_0, r)$ is large but the density of the r-neighborhood of R_0 is relatively small."

Consider Example 4.2. There are 14 potential rules in the 1-neighborhood of $ABC \rightarrow D$. If $ABD \rightarrow C$ is the only other mined rule in this neighborhood, then the density of this neighborhood is $\frac{2}{14} \approx 14.3\%$. If no rule other than $ABC \rightarrow D$ is mined in this neighborhood, then the density is $\frac{1}{14} \approx 7\%$.

We choose to use the number of potential rules as well as the density $\frac{|M \cap N(R_0,r)|}{|N(R_0,r)|}$ in the definition, because we believe that the first number can help determine the degree of interestingness of isolated rules. For example, as shown in Figure 2, the r_0-neighborhoods of R_1 and R_2 can both be sparse neighborhoods; the r_0-neighborhood R_1 is more sparse than that of R_2 if the number of potential rules in $N(R_1, r_0)$ is equal to that in $N(R_2, r_0)$.

The meaning of "the number of mined rules is relatively small" can be specified by some threshold, either given by the user or calculated from the application (e.g. $\frac{|M|}{N}$, where N is the total number of potential rules).

An isolated rule in a neighborhood may not be a rule with the highest confidence in the same neighborhood. Isolated rules with unexpected confidence are clearly more interesting than isolated rules which do not have unexpected confidence.

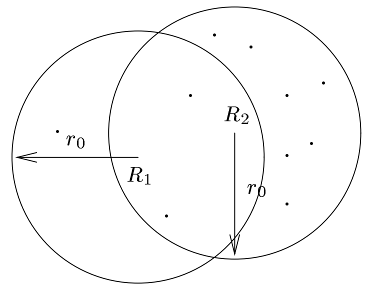

Figure 2: Sparse neighborhoods

Whether a rule is interesting or not is dependent on the *min_support* and *min_confidence* thresholds. For example, when these thresholds are increased, rules which were not isolated may become isolated.

5.3 Interestingness of collections of rules

Neighborhood-based unexpectedness can also be used to define interestingness of collections of rules. We now briefly describe two such collections, including plateau-like, ridge-like rule groups.

Definition 5.4 Let M be a set of mined rules, R_0 a rule in M, and $r_0 < r_1$ be two positive numbers. We say the r_0-neighborhood of R_0 has unexpected confidence in its r_1-neighborhood if the following hold:

- The standard deviation of confidences of the rules in $M \cap N(R_0, r_0)$ is small.
- The average confidence of rules in $N(R_0, r_0)$ is much larger or smaller than the average confidence of the rules in $M \cap (N(R_0, r_1) - N(R_0, r_0))$.

While the interesting group of rules defined above is about a region with a very regular border, it is also possible to consider regions of rules with irregular borders or even curves. For example, one can define an *interesting ridge* as a sequence of rules with very high confidence. Such a ridge can be obtained by iteratively choosing rules with the highest confidence, within some small neighborhood of the rule chosen last.

6 Expected change due to rule structure

There are certain expected changes of supports and confidences of rules implied by the structure of the rules. Such expected changes should be taken into account when considering the (un)expectedness of changes.

We first note two such expected changes.

(1) Given two rules R_1 and R_2, if R_1's item set is a subset of that of R_2, then $support(R_1) \geq support(R_2)$. For example, $support(A \rightarrow C) \geq support(AB \rightarrow C)$.

(2) If R_1's left-hand side is a subset of that of R_2 and R_1 and R_2 have the same item sets, then $conf(R_1) \leq conf(R_2)$. For example, $conf(A \rightarrow CDB) \leq conf(AB \rightarrow CD) \leq conf(ABC \rightarrow D)$; this is because $|m(A)| \geq |m(AB)| \geq |m(ABC)|$.

Such expected changes can happen in a larger scale.

Proposition 6.1 Let U be a fixed item set and PGroup_j^U be the subset of the potential rules whose item sets are U and which have exactly j items on their left-hand sides. Then the average confidence of rules in PGroup_i^U is less than or equal to the average confidence of rules in PGroup_{i+1}^U for any given set of transactions.

Proof: Let k be the number of elements in U. Let $R_1, \cdots, R_m$ be an enumeration of the rules in PGroup_i and $R_1', \cdots, R_n'$ an enumeration of the rules in PGroup_{i+1}. Then $m = \binom{k}{i}$ and $n = \binom{k}{i+1}$; observe that $m = \frac{n(i+1)}{k-i}$.

For each R_j, let S_j be the set of rules in PGroup_{i+1} whose left-hand sides contain the left-hand side of R_j. Then the average of confidences of rules in PGroup_{i+1} is

$$\begin{aligned}
\frac{\sum_{j=1}^{n} conf(R_j')}{n} &= \frac{(i+1)\sum_{j=1}^{n} conf(R_j')}{(i+1)n} \\
&= \frac{\sum_{j=1}^{m}\sum_{R' \in S_j} conf(R')}{(i+1)n} \\
&\geq \frac{(k-i)\sum_{j=1}^{m} conf(R_j)}{(i+1)n} \\
&= \frac{\sum_{j=1}^{m} conf(R_j)}{m}
\end{aligned}$$

which is equal to the average of confidences of rules in PGroup_i. ∎

One might tend to believe that such expected changes can happen for the rules satisfying given support and confidence thresholds as well – that the average confidence of rules in MGroup_i^U is less than or equal to the average confidence of rules in MGroup_{i+1}^U, where

U is a fixed item set and MGroup_j^U is the subset of the rules (i) satisfying given support and confidence thresholds, (ii) whose item sets are U and (iii) which have exactly j items on their left-hand sides. Interestingly, this is false. Indeed, suppose $U = \{A, B, C\}$ and the following set of transactions is given: the transaction ABC occurs 100 times, the transaction AB 50 times, the transaction AC once, the transaction BC 2 times, the transaction A 9 times, the transaction B 8 times, and the transaction C once. Therefore, $|m(ABC)| = 100$, $|m(AB)| = 150$, $|m(AC)| = 101$, $|m(BC)| = 102$, $|m(A)| = 160$, $|m(B)| = 160$, and $|m(C)| = 104$. If $min\_confidence$ is set as 0.65, then $\text{Group}_1 = \{C \rightarrow AB(0.9615)\}$ and $\text{Group}_2 = \{AB \rightarrow C(0.6667), AC \rightarrow B(0.9901), BC \rightarrow A(0.9804)\}$. So, the average confidence of Group_1 is greater than that of Group_2.

7 Ranking of interesting rules

The mined rules can be ranked according to their degree of interestingness and then given to the user in the ranked order. Ranking of rules can be done using some primitive characteristics, including support and confidence and some neighborhood-based characteristics. In this section we first list the important primitive characteristics, and then discuss how to rank the rules using these characteristics.

7.1 Primitive characteristics

We associate with each rule some primitive characteristics, each of them being a function.

Given a rule R and a positive number r, we consider important the following primitive characteristics: $support(R)$, $conf(R)$, $avg\_conf(R, r)$, $std\_conf(R, r)$, $avg\_supp(R, r)$, $std\_supp(R, r)$, $potent\_size(R, r)$, $density(R, r)$. The meaning of the first four items have been explained earlier; $avg\_supp(R, r)$ and $std\_supp(R, r)$ can be defined in manners similar to $avg\_conf(R, r)$ and $std\_conf(R, r)$; $potent\_size(R, r)$ is defined as the number of potential rules of R in its r-neighborhood (which can be calculated from R's item set, r and I). Recall that $density(R, r)$ is defined as $\frac{|M \cap N(R_0, r)|}{|N(R_0, r)|}$.

Given a radius r, the primitive characteristics can be obtained quite efficiently using the partitioning method discussed later.

7.2 Ranking

The above primitive characteristics can be combined to form combined characteristics.

One example of such combined characteristics is the function we used earlier in defining interesting rules with unexpected confidence; that function is defined by $f(R, r) = ||conf(R) - avg\_conf(R, r)| - std\_conf(R, r)|$. Another example of such combined characteristics is $g(R, r)$ defined as $g(R) = 1$ if $|N(R, r)| < 50$ and $g(R, r) = \frac{|M \cap N(R,r)|}{|N(R,r)|}$ otherwise. This function can be viewed as a way of specifying what rules are interesting because they have large and sparse r-neighborhoods.

In general, ranking can be specified by weighted sum of several of such combined characteristics. The ways the combined characteristics are formed and the weighting together specify what are important in an application (or what the users' preferences are).

Ranking can also be augmented by reacting to user feedback. At any moment of time, the user should have looked at the "most interesting rules" produced by the system. After examining a rule R, the user may indicate the relative interestingness of this rule compared with the others, whether he/she is interested in seeing more interesting rules in the neighborhood of the rule just seen, and whether she/he is interested in seeing more rules having the same left-hand side as R, etc. The system can accumulate such feedback and try to adapt to the user's preferences, by perhaps adjusting the functions for the combined characteristics or the weightings. Techniques from neural networks might be helpful here.

8 Implementation issues and a detailed example

In this section, we discuss some implementation issues for finding interesting rules from a set of mined rules. We then give a detailed example to illustrate our ideas.

8.1 First partition then find

To find the interesting rules efficiently from a set M of mined rules, we will first partition M into a number of 1-neighborhoods. We need to have one bucket for each nonempty 1-neighborhood. Recall that rules with the same item sets have identical 1-neighborhoods. Consequently, we can identify buckets with item sets whose corresponding 1-neighborhoods are not empty.

To be able to find a bucket for an item set fast, we have a(n ordered) tree to manage the correspondence between an item set and the physical address of the corresponding bucket.

For each node of the tree we have a pair (U, Ad) where U is an item set and Ad is the address of the bucket.

The parent node of a node (U, Ad) is the node whose item set is V such that V is the smallest item set containing U as a subset. (We fix an order on the items. Then we use the lexical order on item sets when we talk about order between item sets.) The tree can be maintained efficiently – we only need to consider insertions.

The partition based on 1-neighborhoods can be directly used to find interesting rules for radius $r \leq 1$. For radius $r > 1$ we can find all $\lceil r \rceil$-neighborhoods using the tree, by brute force in time $O(p^2)$ where p is the number of nonempty 1-neighborhoods. Observe that the $\lceil r \rceil$-neighborhoods can be formed by merging the pointer sets for the constituent 1-neighborhoods.

After the proper partitioning is done, we can then find those rules which have unexpected confidence or which are isolated from the proper buckets. For each r and each bucket, this can be done in roughly $O(k^2)$, where k is the number of rules in the bucket.

One might wish to find all radius r and rule R such that R has unexpected confidence (or isolated) in its r-neighborhood. When there are too many of such radius, we can get approximate answers by considering only, say, those radius of the form $i + 0.25 * |R|(\delta_2 + \delta_3)$, $i + 0.5 * |R|(\delta_2 + \delta_3)$, $i + 0.75 * |R|(\delta_2 + \delta_3)$, and $i + |R|(\delta_2 + \delta_3)$, where i is a non negative integer.

8.2 A detailed example

A synthetic example is given below to demonstrate the procedures of finding interesting rules. Suppose the total item set is $I = \{A, B, C, D, E, F\}$. Suppose the thresholds for the confidence and support are set as 0.205 and 0.05, respectively, and the following rules are mined.

$CDE \rightarrow F(.78125)$	$CDF \rightarrow E(.7143)$
$CEF \rightarrow D(.625)$	$DEF \rightarrow C(.833)$
$CD \rightarrow EF(.2451)$	$CE \rightarrow DF(.4167)$
$CF \rightarrow DE(.3205)$	$DE \rightarrow CF(.2551)$
$DF \rightarrow CE(.333)$	$EF \rightarrow CD(.3125)$
$E \rightarrow CDF(.2083)$	$CD \rightarrow E(.3137)$
$CE \rightarrow D(.5333)$	$DE \rightarrow C(.3265)$
$D \rightarrow CE(.256)$	$E \rightarrow CD(.267)$
$CD \rightarrow F(.3431)$	$CF \rightarrow D(.4487)$
$DF \rightarrow C(.4667)$	$D \rightarrow CF(.28)$
$F \rightarrow CD(.233)$	$CE \rightarrow F(.6667)$

$CF \rightarrow E(.5128)$	$EF \rightarrow C(.5)$
$E \rightarrow CF(.333)$	$F \rightarrow CE(.2667)$
$DE \rightarrow F(.3061)$	$DF \rightarrow E(.4)$
$EF \rightarrow D(.375)$	$D \rightarrow EF(.24)$
$E \rightarrow DF(.25)$	$C \rightarrow D(.51)$
$D \rightarrow C(.816)$	$C \rightarrow E(.3)$
$E \rightarrow C(.5)$	$C \rightarrow F(.39)$
$F \rightarrow C(.52)$	$D \rightarrow E(.784)$
$E \rightarrow D(.8167)$	$D \rightarrow F(.6)$
$F \rightarrow D(.5)$	$E \rightarrow F(.67)$
$F \rightarrow E(.5333)$	

The percentage following each rule represents its confidence.

The above 43 mined rules can be partitioned into 11 clusters, each of these being a 1-neighborhood. The biggest one is for the 1-neighborhood centered at the rule of $CDE \rightarrow F$ and contains all rules whose item set is $\{C, E, D, F\}$; the other ten 1-neighborhoods are much smaller and are for rules whose item sets are $\{C, D, E\}$, $\{D, E, F\}$, $\{C, D, F\}$, $\{C, E, F\}$, $\{C, D\}$, $\{C, E\}$, $\{C, F\}$, $\{D, E\}$, $\{D, F\}$, and $\{E, F\}$ respectively.

Observe that the 2-neighborhood of $CDE \rightarrow F$ is the union of the 1-neighborhoods for $\{C, D, E, F\}$, $\{C, D, E\}$, $\{D, E, F\}$, $\{C, D, F\}$, $\{C, E, F\}$, $\{A, C, D, E, F\}$ and $\{B, C, D, E, F\}$.

For the rule $R_1 : CDE \rightarrow F$, the following table shows the neighborhood-based parameters for the different radius values of $\rho_j = i + \frac{|R_1|}{4} * j * (\delta_2 + \delta_3)$, where $i = 0$ and $j = 1, 2, 3, 4$. Observe that R_1 is a relatively interesting rule with unexpected confidence in its ρ_1-neighborhood where $\rho_1 = 0.17$. There are no isolated rules.

r	ρ_1	ρ_2	ρ_3	ρ_4
avg_conf	0.3056	0.4711	0.4264	0.4264
std_conf	0.0786	0.2310	0.2075	0.2075
*	0.4757	0.3102	0.3549	0.3549
$Density$	100%	80%	85%	79%
**	4	10	13	14

Table 1: The confidence fluctuation in ρ_j-neighborhoods of R_1, where $\rho_j = \frac{|R_1|}{4} * j * (\delta_2 + \delta_3)$. * stands for $|conf(R_1) - avg\_conf(R_1, r)|$, ** for number of potential rules.

9 More discussion on related works

Typical measures of interestingness can be divided into two classes: the objective ones and the subjective ones. The objective ones, such as rule template and rule cover, focus on the importance of rules' structures. The subjective ones, in contrast, depend not only on the structure of a rule and the data, but also on the user who examines the rules.

Two useful subjective interestingnesses are *actionability* and *unexpectedness*. The notion of *actionability* [8, 10] of association rules is based on the usefulness of the rules to user – whether the users can do something because of the rules to their advantage. Actionability is an important subjective measure of interestingness because users are mostly interested in the knowledge that permits them to do their jobs better by taking some specific actions in response to the newly discovered knowledge. However, it is not an easy matter to decide what rules are actionable; the answer might be obtained only after a period of practical validation.

Unexpectedness can be either subjective or objective. Apparently, if a newly discovered pattern is surprising to the user, then it is certainly interesting. For the subjective ones [6, 11], "surprising" means the discovered knowledge contradicts the user's beliefs. Therefore, unexpectedness is closely related to beliefs or general impressions. Beliefs can be classified into two types: hard beliefs and soft beliefs. The hard beliefs are the constraints that cannot be changed with new evidence, whereas the soft ones are those that the user is willing to change with new evidence.

The objective unexpectedness can be specified in statistical terms. For example, having support and confidence larger than their corresponding thresholds is one such specification; having a higher chance than that under the independence assumption is another.

Our neighborhood-based interestingness belongs to the class of objective measures of interestingness, because the neighborhoods are determined by the rules' structures. Clearly, useful interestingness measures should help identify those rules that are surprising to the user. We believe that our neighborhood-based unexpectedness is very useful in this regard, and can be used in complement the other measures.

Rule template was also used to help find interesting rules [5] and it is an objective measure for interestingness. A template is an expression

$$A_1, \ldots, A_k \Rightarrow A_{k+1},$$

where, each A_i is either an item name, a class name, or an expression $C+$ or $C*$ (C is a class name). Here $C+$ and $C*$ correspond to one or more and zero or more instances of the class C, respectively.

10 Concluding remarks

We have proposed neighborhood-based unexpectedness as a way of identifying interesting rules. In this approach, the interestingness of a rule depends not only on its own support and confidence but also on the support and confidence of rules in its neighborhood. This idea has not been used by previous interestingness measures, including unexpectedness, actionability, rule cover, rule template.

Neighborhood-based interesting rules proposed in this paper include those with unexpected confidence and those with sparse neighborhood. Similar ideas have been used for identifying interesting sets of rules such as plateaus and ridges. The neighborhood-based parameters have been combined with other parameters to rank the interesting rules. We have also addressed some implementation issues for finding neighborhood-based interesting rules.

We gave a few expected tendencies of changes due to rule structures, which should be taken into account when considering unexpectedness. There might be other similar useful properties.

There are also some issues requiring further research, including: How to use users' feedback to adjust the functions used in the ranking of interesting rules? How to adjust the values of δ_1, δ_2, and δ_3 to best fit the application? It is also possible to find other types of neighborhood-based interesting rules.

Acknowledgement

The authors benefited from discussions with Ramamohanarao Kotagiri on data mining.

References

[1] R. Agrawal, T. Imielinski, and A. Swami. Data mining: a performance perspective. *IEEE Transactions on Knowledge and Data Engineering*, 5(6): 914-925, 1993.

[2] R. Agrawal, T. Imielinski, and A. Swami. Mining association rules between sets of items in large databases. In *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, pages 207-216, Washington, D. C., 1993.

[3] M.-S. Chen, J. Han, and P. S. Yu. Data mining: an overview from a database perspective. *IEEE Transactions on Knowledge and Data Engineering*, 8(6): 866-882, 1996.

[4] Y. Fu and J. Han, Meta-rule-guided mining of association rules in relational databases. *Proc. 1995 Int'l Workshop. on Knowledge Discovery and Deductive and Object-Oriented Databases (KDOOD'95)*, Singapore, December 1995, pp. 39-46.

[5] M. Klemettinen, H. Mannila, P. Ronkainen, H. Toivonen, and A. I. Verkamo. Finding interesting rules from large sets of discovered association rules. In *Proceedings of the Third International Conference on Information and Knowledge Management*, pages 401-407, Gaithersburg, Maryland, 1994.

[6] B. Liu and W. Hsu. Post-analysis of learned rules. In *Proceedings of AAAI*, pages 828-834, 1996.

[7] G. Piatetsky-Shapiro. Discovery, analysis, and presentation of strong rules. In *Knowledge Discovery in Databases*, G. Piatetsky-Shapiro and W. J. Frawleay, Eds, pages 229-248. AAAI Press/The MIT Press, Menlo Park, CA, 1991.

[8] G. Piatetsky-Shapiro and C. J. Matheus. The interestingness of deviations. In *Proceedings of the AAAI-94 Workshop on Knowledge Discovery in Databases*, pages 25-36, 1994.

[9] R. Srikant and R. Agrawal. Mining generalized association rules. *IBM Research Report RJ 9963*, 1996.

[10] A. Silberschatz and A. Tuzhilin. On subjective measures of interestingness in knowledge discovery. In *Proceedings of the First International Conference on Knowledge Discovery and Data Mining*, pages 275-281, Montreal, Canada, August, 1995

[11] A. Silberschatz and A. Tuzhilin. What makes patterns interesting in Knowledge discovery systems. *IEEE Transactions on Knowledge and Data Engineering*, 8(6):970-974, December, 1996.

[12] H. Toivonen, M. Klemettinen, P. Ronkainen, K. Hätönen and H. Mannila. Pruning and grouping discovered association rules. In *MLnet Workshop on Statistics, Machine Learning, and Discovery in Databases*, Crete, Greece, April 1995.

[13] A. Tuzhilin. A pattern discovery algebra. In *Proceedings of the 1997 Workshop on Research Issues in Data Mining and Knowledge Discovery*, 1997.