*Artificial Intelligence for Engineering Design, Analysis and Manufacturing* (2001), *15*, 401–410. Printed in the USA.
Copyright © 2001 Cambridge University Press 0890-0604/01 $12.50

# Visualizing constraint-based temporal association rules

PANU KORPIPÄÄ

VTT Electronics, Kaitoväylä 1, P.O. Box 1100, FIN-90571 Oulu, Finland

(Received September 11, 2000; Accepted April 29, 2001)

**Abstract**

When dealing with time continuous processes, the discovered association rules may change significantly over time. This often reflects a change in the process as well. Therefore, two questions arise: What kind of deviation occurs in the association rules over time, and how could these temporal rules be presented efficiently? To address this problem of representation, we propose a method of visualizing temporal association rules in a virtual model with interactive exploration. The presentation form is a three-dimensional correlation matrix, and the visualization methods used are brushing and glyphs. Interactive functions used for displaying rule attributes and exploring temporal rules are implemented by utilizing Virtual Reality Modeling Language v2 mechanisms. Furthermore, to give a direction of rule potential for the user, the rule statistical interestingness is evaluated on the basis of combining weighted characteristics of rule and rule matrix. A constraint-based association rule mining tool which creates the virtual model as an output is presented, including the most relevant experiences from the development of the tool. The applicability of the overall approach has been verified by using the developed tool for data mining on a hot strip mill of a steel plant.

**Keywords:** Association Rules; Hot Strip Mill; Interestingness; Virtual Model; Visualization

## 1. INTRODUCTION

Most real world processes, such as the hot strip mill process in steel production, entail massive amounts of time-dependent information. This information may consist of, for example, target settings for the process devices and the product, symbolic anomaly information derived from online-analyzed diagnostic measurements, the actual measured process and product values, and facts on failed or defective products. When mining this kind of time-dependent data, the association between different events concerning process attributes is an issue of particular interest. For example, it would be interesting to find out what kind of product target settings are likely to produce an imperfect product, and this information could then be used to detect the areas on which process improvement actions should be focused. The problem of discovering association rules was introduced in Agrawal et al. (1993), and has been intensively studied since then, for example, in Agrawal et al. (1996). An example of an association rule in the domain of steel making is: "if furnace id = 1, then local temperature deviation is high in the coiling phase with a confidence of 0.72."

In addition to the association rule itself, in many cases it is important to know when the effect of the rule has first appeared or when it has disappeared. The changes in the rules over time are likely to reflect changes in the process itself. For example, if a product flaw is associated with a particular process device value, it is important to find out the initial correlation of the attributes, so as to be able to match it with what has happened in the process to make the device act abnormally. One may also be interested in monitoring the rule confidence changes between maintenance operations to find out which problems persist and which vanish after maintenance. Furthermore, increasing rule confidence may indicate a future failure of a process device. Temporal data mining concerns the analysis of events ordered by the time dimension (Roddick & Spiliopoulou, 1999). Literature gives a somewhat diverse terminology for temporal mining; coarsely, however, three directions can be distinguished. "Sequence mining" focuses on discovering frequent episodes in time series (Mannila et al., 1995). "Pattern discovery" concerns finding items ordered by time, which together satisfy certain (statistical) properties (Berndt & Clifford 1996). "Temporal rule discovery" focuses on the discovery of relationships among events that may be ordered in time and may be causally related (Long et al. 1991). In this article, we use the term "temporal association

Reprint requests to: Panu Korpipää, VTT Elektroniikka, Kaitoväylä 1, P.O. Box 1100, FIN-90571 Oulu, Finland. E-mail: Panu.Korpipaa@vtt.fi



https://doi.org/10.1017/S0890060401155034 Published online by Cambridge University Press

rule" strictly to refer to an association rule observed as a function of time, where time is divided into nonoverlapping user-specified intervals of equal length. This approach of monitoring rule behavior over time adds to the difficulty of presenting rules effectively for the user, since the number of rules is in the worst case multiplied by the number of time intervals.

With large data sets, problems are caused by the time and memory performance of the association rule algorithm, and by the large number of resulting rules. Our approach to these problems is to let the user guide the association rule mining process by imposing different constraints on target data. The basic idea of constraint-based data mining (Han et al., 1999) is to allow computers to perform the tasks they can do best, such as querying, searching, and calculation, and to assign the users to carry out the tasks they can do best, such as specifying the focus of the data mining session by setting constraints, based on, for example, domain knowledge.

Observing rules over intervals of time basically multiplies the number of resulting rules by the number of intervals. The number of rules displayable within the bounds of normal display resolution is small, and therefore usable and efficient methods for an interactive visualization of the rules are needed. The first decision in designing an appropriate visualization is to choose between two-dimensional (2D) and three-dimensional (3D) presentation. The benefits of these two over each other have been debated for a long time, and clearly both have proven their advantages. In this work, 3D paradigm is chosen mainly because it gives a perceptionally natural extra dimension compared to 2D. Literature lists a multitude of successful applications utilizing interactive 3D visualizations. STARLIGHT (Risch et al., 1999) is an example of a 3D visualization-oriented user interface, enabling integrated use of multiple visualization techniques to support the comparison of content and interrelationship information at several levels of abstraction. More applications can be found, for example, in Card et al. (1999). Modeling languages such as Virtual Reality Modeling Language v2 (VRML2; Ames et al., 1997) have been developed for implementing interactive 3D models. The repertoire of specific visualization methods is extensive (Nielson et al., 1997). We have applied glyphs (Beddow, 1990) as a method for presenting multiple variables simultaneously in compact form, and brushing (Becker & Cleveland, 1987) as a method for interactively focusing and providing more information for the user.

Concerning the task of visualizing association rules, one interesting method is a correlation matrix. Three-dimensional association rule correlation matrix has been presented in literature (Brunk et al., 1997) previously, and it has been applied to the Mineset data mining tool. Also 2D association rule correlation matrix has been introduced, for example, by Westphal and Blaxton (1998). Wong et al. (1999) present a 3D model for visualizing association rules consisting of multiple items instead of one antecedent and one consequent. Their method is an excellent and innovative way to visualize association rules in 3D. Compared to their approach, the difference of the method introduced in this article, concerning data mining session, is that it gives a quick overview of the main associations of the user-specified domain context, and provides a flexible way to access and compare the behavior of that context over time. The main visualization design principle has been to first give an overview, then provide a means to interactively focus on more detailed information. In our approach, multiple item associations are discovered separately if necessary, after choosing the appropriate items based on the overview. These focused multiple item rules could then be visualized using the method of Wong et al. (1999), but this possibility has not been implemented so far. It must be noted here that the purpose of this article is by no means to show the superiority of our method compared to some other, since a statement of that kind cannot even be proven without an extensive usability study. Comparing to the previous one-item 3D correlation matrix solutions, we have further developed the concept by introducing interactive visual exploration of temporal rules, the use of automatically generated virtual models, and a method for assessing and visualizing rule interestingness based on combining weighted characteristics of rule and rule matrix.

The interestingness measures which are usable with association rules have been discussed in the literature earlier, by Smyth and Goodman (1992) and Wang et al. (1998), among others. The knowledge viewpoint to interestingness is discussed, for example, by Klemettinen et al. (1994). Unlike these previous approaches, in this work the characteristics of the rule matrix are considered as a component of rule interestingness, resulting in the ability to assess the interestingness of a rule with respect to its neighbors. Another difference is the use of weight factors in order to enable flexibility in emphasizing different interestingness components.

In summary, the main technical contributions of this work are the enhancements to the existing association rule visualization methods, the application of which were considered useful in the association rule mining at the hot strip mill, and which, to our knowledge, have not been published previously; visualizing temporal association rules in a virtual model with interactive exploration; and calculating rule interestingness based on combining weighted characteristics of rule and rule matrix. In addition to this, the most relevant practical experiences of developing a data mining tool for a specific real-world domain are shared.

## 2. ASSOCIATION RULE MINING

This section discusses properties of association rule mining by introducing the prototype tool developed, and the types of constraints used. Heterogeneous data restructuring and the role of time in mining are described shortly, and key

https://doi.org/10.1017/S0890060401155034 Published online by Cambridge University Press

views on user interaction supporting the discovery of rules are presented. Examples from the hot strip mill process data mining are used frequently to illustrate the connection of the abstract concepts with the concrete world of steel production.

### 2.1. Constraints

In this work, we have adopted the constraint-based approach to mining association rules. The user can define a domain-specific context of the mining session by setting three kinds of constraints on the data: data constraints, dimension constraints, and rule constraints (Han et al., 1999). This approach signifies that the user interaction of the tool is best handled by a wizard-style stepwise interface (see Fig. 1). Each step consists of controls containing context-sensitive information about its function for the user. In the case of hot strip mill data, there are five separate domains from which to choose the data: primary data, diagnostic data, production data, exception data, and reject data. These domains represent information in the different phases of the target process. Primary data consists of the target values for the process devices and product dimensions, diagnostic data contains online-analyzed symbolic information describing possible anomalies in a product or process (symptoms), production data holds the actual device settings and product dimensions attained by the process, and exception and reject data consist of codes describing the abnormalities de-

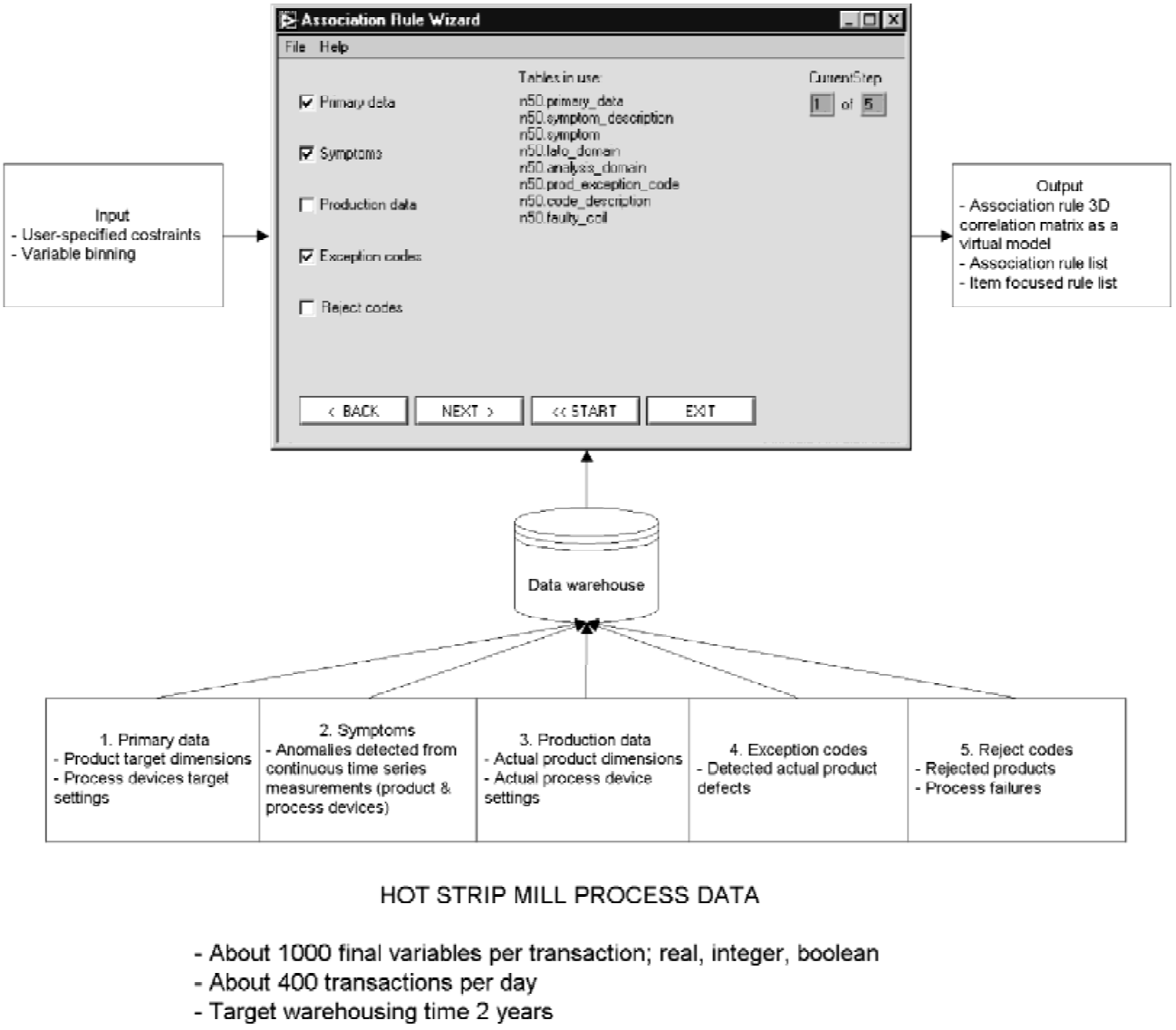



**Fig. 1.** The simplified overall system structure in hot strip mill data mining. The user is currently taking the step of setting data constraints, and the corresponding physical database tables are shown in the middle of the screen. The order of magnitude of the target total (2 year) data amount is tens of gigabytes.

https://doi.org/10.1017/S0890060401155034 Published online by Cambridge University Press

tected by the automation system, and of information about rejected products.

Thus the first choice for the user is to decide which connections between these five overall process domains are sought. Data constraints are specified by selecting any number of these domains; one may for example be interested in finding out which diagnostic symptoms indicate a high probability of subsequent product rejection. Each domain is expressed by physical database tables containing fields, that is, items from which the user can choose to set dimension constraints. Different constraint compositions can be saved for later use. Rule constraints refer to bounding the value ranges of the items; for example, only symptoms with a severity level over 0.5 are counted. The effects of the constraints on algorithm performance have been evaluated earlier (Skrikant et al., 1997), with the results indicating a dramatic enhancement. Figure 1 presents a simplified overall system structure for the case of hot strip mill data mining. The values in the data warehouse are well selected, and preprocessed by the automation and diagnosis systems (Kurki et al., 1998).

### 2.2. Data restructuring

The mining of multiple separate sources reveals a practical problem caused by heterogeneous database table structures. Since the tables are originally structured according to the domain needs, the information has to be rearranged for the association rule mining. Basically, there are two feasible solutions to this problem in the case of real-time online systems: first, to create conversion modules which at the server end continuously convert information into a new table, where the information is homogeneously arranged; and second, to make the preprocessing modules a part of the mining tool to perform the conversion at the client end. Both solutions have their advantages and disadvantages, and the latter alternative was selected as it was estimated to be less laborious and to require less effort in database management. The preprocessing modules constitute a part of the mining tool, their task being to convert data into a structure to be used as an input for the association rule algorithm module. There are two main operations used in this conversion. First, according to the constraints set by the user, join operations are used to gather data from multiple database tables into a table structure in the following form: transaction id (TID), *n* items, where *n* is the number of columns. Second, to achieve the form of TID, item, there is a need to concatenate the *n* items in some cases (*n* may also be 1). For example, if we have a row of TID, symptom, location, the result would be TID, symptom_location.

### 2.3. User-centered data mining

In the development of a data mining tool, the user needs and the user interface are of great importance. It is essential to provide adequate support for the usage processes of the intended users, in this case steel-making domain experts, concerning the interface and functions. For example, usually the domain experts in this case were primarily interested in simple rules, stating the main correlations between items in the chosen context, and after finding an interesting simple rule, possibly expanding it. The developed prototype tool supports this by first giving an overview of a set of simple rules, these having one item in both antecedent and consequent, with matrix visualization and interestingness indicators, and then providing the user with a possibility of focusing on certain items for a more complex rule consisting of multiple items. Thus the main design principle has been to first give an overview, then to offer a way to look at the details. A more detailed view of one rule and the history of the rule can be accessed easily by moving the mouse cursor over the glyphs in the model.

The needs for mining vary: the user may wish to verify certain presumed hypotheses; or if there has been a change in the process some time earlier, the user may want to track down the reasons leading to it, which may require temporal mining. For example, if a product flaw is associated with a particular process device value, it is important to find out the initial correlation of the attributes to reveal the cause. Sometimes the user may be aware of some problems in the process which occur during certain time period(s) and may wish to focus on them by ascertaining which events are connected. Changes in rule confidence between maintenance operations are of potential interest in tracking which problems persist and which vanish after maintenance. Furthermore, increasing rule confidence may indicate a future failure of a process device or a following decrease in the product quality.

In general, time is the most important attribute in modeling the changing behavior of the process, since it is the only common factor for all the dependencies in the heterogeneous large-scale process. Many features of the process, such as wearing of components, product attributes, symptoms, rejected products, and so forth, and associations between them are potentially interesting when they are observed as a function of time and can thus be connected to the different actions made to modify the process, such as maintenance, repairs, and device updates. Therefore, the tool must be able to support effectively the observation of the associations over time, and this is achieved by letting the user specify the time period to be mined and the number of intervals. The association rule discovery algorithm (Agrawal et al., 1996) is then simply repeated by the tool for each interval, resulting in as many sets of rules as there were intervals.

In some cases, the need for mining can be purely discovery driven, but even then there are usually vast amounts of futile domain and item combinations, which can be pruned by adopting a constraint-based user-driven mining process. There are a number of advantages to this kind of user-centered mining of association rules, as opposed to a constraintless searching of all rules. The most important issue

https://doi.org/10.1017/S0890060401155034 Published online by Cambridge University Press

is the ability to concentrate on specific contexts. Furthermore, if no constraints are employed, the number of generated rules will be huge, even with small data sets. Obviously, the partitioning of the data set according to intervals of time is also of benefit for the algorithm performance. The most practical way of carrying out data mining is to do it on personal computers with limited resources. With an appropriate focusing of the mining session, problems with memory and processing time can be avoided.

## 3. VISUALIZATION AND INTERACTION

The result of the data-mining session is a collection of association rules. From the rules, the tool generates a virtual model as an output. This section briefly introduces the visualization and interaction methods used, and describes the visualization model and interestingness measures.

### 3.1. Three-dimensional matrix presentation

The generated rules are shown using an association rule matrix, presented in 3D. In the basic version of this representation, rows represent rule antecedents, columns are rule consequents, and the height of the object at the crossing of the row and the column represents the confidence of the rule. As an example, in Figure 3, the user has placed the mouse cursor on one of the rules, and the attributes of the rule are shown at the upper left corner of the display. The antecedent (row) of the rule is "f6_t_tgt=890 - 900," the consequent (column) of the rule is "COIL_T-_CLASS3=80 - 100," and confidence is 0.52.

Measuring the performance of a visualization or a user interface quantitatively is related to usability, and requires an extensive group of testees. Thus, comparing the visualization of this study quantitatively to, for example, some competent 2D method is considered out of the scope of this article. However, there are some strengths, as well as weaknesses, in the interactive interestingness-augmented 3D correlation matrix that can be pointed out. The most important benefit of the visual matrix representation over list-based structures is the capacity of providing a quick overview of the rules, in the form of a concise visual summary. The matrix provides a practical way of comparing rules with their neighbors. Showing pointers of statistical interestingness reduces the glut of rules to be examined. In addition, all the items the user has chosen during the mining session are shown in the matrix, even if there are no rules. This makes it possible to also detect such cases where a rule is absent. For example, if there is a rule concerning furnace 1, it is important from the domain point of view to know whether there are similar rules for other furnaces as well. If not all the items were displayed, the user would not necessarily even know that the other furnaces were also included in the source data. Considering the visualization and comparison of neighboring rules, it must be noted that this advantage only exists if the items in the matrix are related to each other in such a way that the item values can be ordered. In the case of independent items, the neighbor comparison is not useful.

The maximum practical size of the matrix, to be viewed without rotating or moving, is about $30 \times 30$. Larger matrices are slower to handle (with standard PC hardware and monitor) and focusing on a single rule becomes more difficult. The number of simultaneously viewable rules is limited, even if 3D presentation is used, due to the limited screen resolution or further, eye resolution. To overcome these restrictions, interactive mechanisms will have to be used to bring up more information. This interaction should, of course, be as natural and easy as possible, so as to enable efficient access to the information. In this work, brushing is used as an interaction method for swiftly gaining access to the information related to the currently visualized information. The user is able to view the history or details of the rule by simple mouse-over interactions, although only one set at a time.

### 3.2. Visualization methods

The correlation matrix is presented using a virtual model, described in VRML2. Each model is generated automatically by the tool according to the user settings made during each mining session. There are several benefits in using virtual modeling in the visualization of data mining results compared to using a hard-coded interactive visual model integrated in a data mining tool. The virtual model is hardware independent, transferable, and it can be viewed, rotated, panned, zoomed, and so forth with a web browser shareware plug-in. The user can carry out several rule mining sessions, and after finding an interesting model, for instance, send it to colleagues by e-mail, who can examine it with their ordinary web browsers. VRML2 is a standard language containing numerous visual structures and features, with a relatively simple syntax that supports interactivity. More complex functionality can be added using Java or JavaScript. To implement the interactive functions in this work, VRML2 features TouchSensor node and isOver event were applied by routing the isOver event to modify the transparency field of the Material node of the shape group to be visualized.

In the visualization model used, individual rules are displayed using a glyph representation. A glyph is a multivariate visualization method. A glyph is an object, which has attributes such as size, shape, and color. These attributes can be assigned to describing different variates. In this work, each glyph describes one rule, with the following attributes: The height of the BlueBox object (the blue bar-shaped object) corresponds to rule confidence, the y-axis position of the GreenCylinder is the rule consequent (right hand item) occurrence, the y-axis position of the RedCylinder is the rule antecedent (left hand item) occurrence (Fig. 2). The occurrences are, in addition to the absolute values, normalized between 0 and 1, and thus represent the probability of

https://doi.org/10.1017/S0890060401155034 Published online by Cambridge University Press

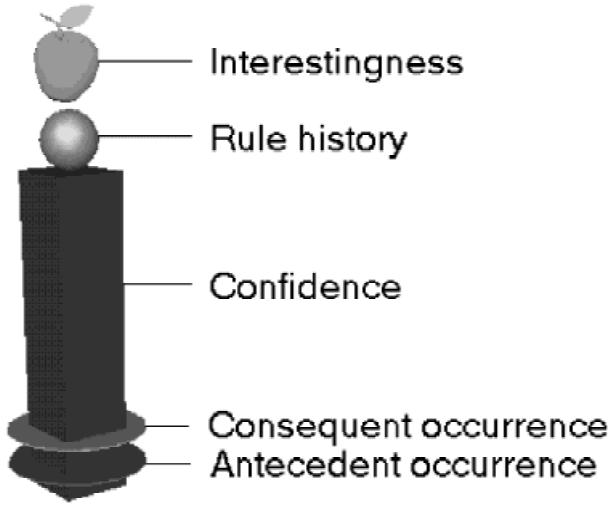



**Fig. 2.** The structure of the rule glyph.

the item appearing in the data set. The interestingness calculated from the matrix and rule attributes is shown by the size of the RedApple above the rule. Temporal rules are coded into GoldBall objects. Glyphs are positioned on the matrix, the row and column names of which are items chosen by the user during the mining session. Each crosspoint of the matrix represents a possible rule (see Fig. 3).

Brushing is an interactive method allowing the user to select on-screen specific data points or subsets of data and to identify their characteristics. In effect, the use of brushing makes it possible to show an increased number of variates. In the visualization model of this work, brushing is used in the following two ways:

1. When the user moves the mouse cursor on the BlueBox object, the rule attributes are displayed on the screen (Fig. 3). Although the rule can be seen also straight from the matrix, in a large matrix it is, however, difficult to distinguish the names or exact values of the rule, which are provided here by means of brushing. Also the absolute numbers of both items in the rule are shown. The main purpose of the matrix is to give the user an overview, while the purpose of the brushing operations is to provide more accurate information and additional figures, which would be difficult to present otherwise.

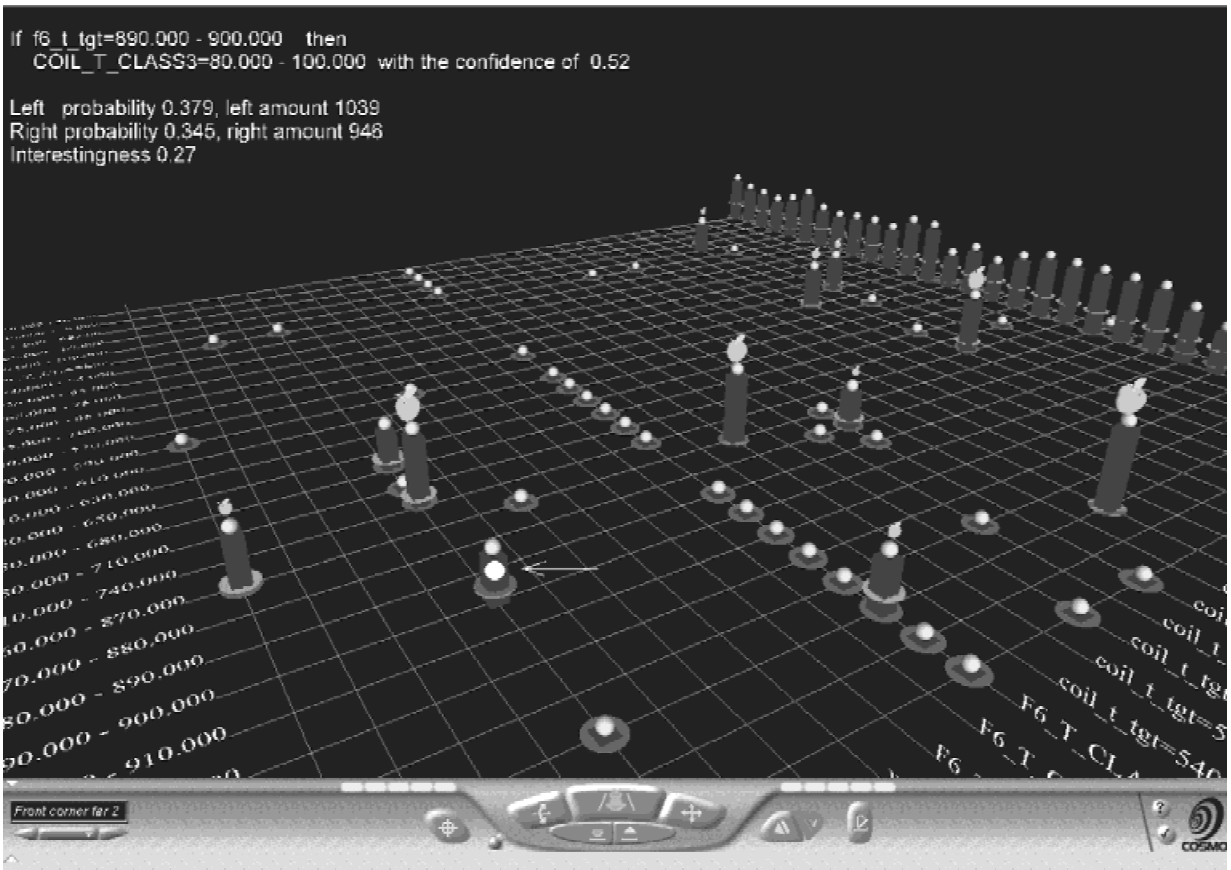



**Fig. 3.** When the user moves the mouse cursor (represented by the arrow) onto the BlueBox object of the rule glyph, the rule and its attributes are displayed on the upper left corner of the screen.

https://doi.org/10.1017/S0890060401155034 Published online by Cambridge University Press

2. When the mouse cursor is moved onto the GoldBall object, the temporal association rules are displayed (see Fig. 4). Above the rule glyphs, the dates are shown for each time period the rules have been searched from. A flat rule denotes that there has been no rule with a confidence level exceeding the threshold. In the matrix, the flat rules are used for indicating that there has been a rule, but in the most recent period, shown by the visible matrix, the rule is no longer there.

The rules with multiple items in both the antecedent and consequent are not visualized in 3D in this work, but are presented in a normal text form. Three-dimensional visualization of multiple-item association rules is also feasible, and it is discussed by Wong et al. (1999). The reason for visualizing only the single-item rules here is that most domain experts in this case have been primarily interested in simple rules, stating the main correlations between items. After finding an interesting simple rule, it can then be expanded to contain more items by choosing a single item for which all the rest of the rules are searched. This procedure is supported by the tool.

### 3.3. Interestingness measures

To further enhance the efficiency of the visualization, we adopt a statistical interestingness measure. Indicating rule interestingness is an ambiguous task. From the domain point of view, only domain experts have the knowledge to tell which rules are interesting and which are irrelevant. However, the matrix and the rule attributes provide the possibility of assessing the statistical interestingness of the rule. The purpose of the interestingness measure is to give the user priorities on where to look first and to provide information on which rules are valuable and which are not worthy of further consideration.

The size of the interestingness apple I is

$$I = \frac{cw_1 + Jw_2 + |c - \mu|w_3}{w_1 + w_2 + w_3} \tag{1}$$

where $c$ is the confidence of the rule, $J$ is the $J$-measure (Smyth & Goodman, 1992) shown in Eq. (2), $\mu$ is the column mean, with column being the column of the rule in the matrix, and $w_1$, $w_2$, $w_3$ are weighting factors.

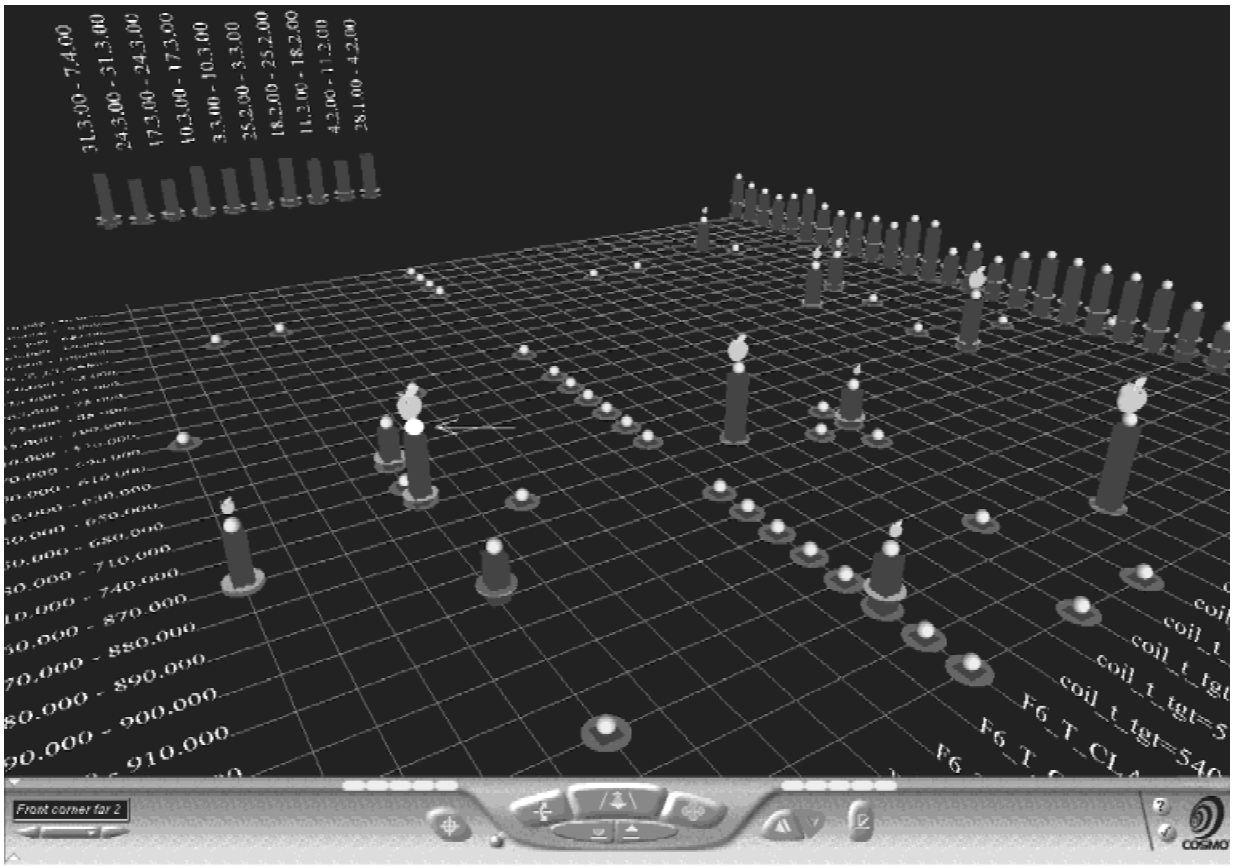



**Fig. 4.** When the user moves the mouse cursor (represented by the arrow) onto the GoldBall object of the rule glyph, the history of the rule is displayed on the left side of the matrix.

https://doi.org/10.1017/S0890060401155034 Published online by Cambridge University Press

The tests indicated a balanced result with, for example, the weighting factors $w_1 = 0.5, w_2 = 2, w_3 = 3$. The trial and error approach is used to set the weights. The purpose of these weights is to balance the ratio between the three interestingness components, which are described below. The weights can also be used to emphasize or to reduce the magnitude of the corresponding interestingness component. The components are partly overlapping, but emphasize different features. The number of the weight in the equation corresponds to the number of the component. The equation consists of the following components:

1. *Rule confidence*. This component corresponds to the term $cw_1$ in the equation.
2. *J-measure*. This component corresponds to the term $Jw_2$ in the equation. A short introduction: For the association rule $Y \rightarrow X$, let $p(Y)$ and $p(X)$ denote the probability of antecedent and consequent occurrence. Rule support is $p(XY)$, where $XY$ refers to the conjunction of $X$ and $Y$, and rule confidence $p(X|Y) = p(XY)/p(Y)$. J-measure is defined as

$$J(X,Y) = p(Y)\left[p(X|Y)\log_2\frac{p(X|Y)}{p(X)} + (1-p(X|Y))\log_2\frac{1-p(X|Y)}{1-p(X)}\right]. \quad (2)$$

   The first term $p(Y)$ measures the generality of the rule. The term inside the square brackets indicates how dissimilar the priori $p(X)$ and the posteriori $p(X|Y)$ are from $X$.
3. *Rule difference from its neighbors*. This component corresponds to the term $|c - m|w_3$ in Eq. (1). If a rule in a column differs from the other rules in the column, thus also from its neighbors, the rule is usually more interesting. If all the rules in the matrix column have nearly the same confidence (no difference between mean confidence and confidence), the consequent is usually quite a high probability. This means that all antecedents are likely to exist with a consequent, thus they are not interesting rules. The interestingness is increased by the difference between mean confidence of the column and rule confidences. Items with no rule in the column are accounted as zero confidence rules in the calculation of the column mean. Therefore, if there are, for example, only two rules in the column with the same confidence, this component is still valid.

EXAMPLE 1. There are $n = 10$ items in the column and 2 rules with the confidence of $c = 1$; the interestingness component $I_3$ for a rule is

$$I_3 = |c - (c + c + 8 * 0)/n| = 0.8.$$

■

In other words, the final interestingness $I$ is a weighted sum of the interestingness components described in these three components. Due to the use of weights, this method is not exact, but as a trade-off, it offers flexibility. If the user wishes to emphasize or exclude some of the interestingness components, all that is needed is to set the corresponding weight accordingly.

A thorough examination of the usefulness of the presented visualization method and interestingness measure would require an extensive usability study with a sufficiently large group of users. The methods to be compared could include list presentation, list presentation with interestingness measures, matrix presentation, and matrix presentation with interestingness measures. Such a study should discuss, for example, the following features:

- the average time for finding certain rules which are interesting from the domain point of view;
- the average time used for examining noninteresting rules from the domain point of view;
- the average time used for examining rules not interesting statistically nor from the domain point of view;
- the average percentage of rules marked interesting that are interesting from the domain point of view;
- the average percentage of rules marked interesting that are not interesting from the domain point of view;
- the average percentage of rules not marked interesting that are interesting from the domain point of view;
- user opinion: score of each method on the scale from 1 to 10.

Such a study requires much further work, and is considered to be out of the scope of this article. However, the matrix presented in this article represents a good example of a successful rule interestingness assessment. It is to be noted that the rules of the column at the end of the matrix have no real interestingness although they show high confidence. The difference between antecedent and consequent is high and consequent probability is high; thus there tends to be a rule for all the antecedents. Without the interestingness measures used in this example, the user, in vain, would have had to look at a number of rules about two to three times higher than with the interestingness indicators present. The most interesting solitary rules quite correctly have the biggest apples here. However, as mentioned before, the total interestingness for the user depends mainly on domain-specific knowledge; thus the apples simply offer additional information, serving as quick pointers and marking the rules with the highest potential.

## 4. HOT STRIP MILL EXAMPLES

There is a considerable amount of information in the matrix visualization that a domain expert can efficiently access. The examples described in this section are only a fraction of it, and their purpose is merely to briefly illustrate the

https://doi.org/10.1017/S0890060401155034 Published online by Cambridge University Press

possibilities of the presented visualization and exploration interface for transferring the domain-specific information from the visual model into knowledge for a human. The purpose is thus not to formally present any particular domain problem and its solution. All the facts described in these examples cannot be seen from a static picture, but by using an interactive model, the train of thought presented below becomes visible.

By analyzing the case of this matrix from the domain point of view, one may infer, for example, that when the target temperature of the strip in the last finishing mill phase (f6) of the process is between 890 and 900 degrees, the temperature is most likely to hit class 3, the optimal temperature class for the process, in the coiling phase (the rule in Fig. 3 with the white dot; if f6_t_tgt=890 - 900 then COIL_T_CLASS3=80 - 100 with the confidence of 0.52). Although being of low confidence, this rule is somewhat interesting since its neighbors are not producing a similar rule at all. Concerning f6 temperatures, the leftmost rule in the matrix indicates problems with temperatures between 860 and 870 (if f6_t_tgt=860 - 870 then COIL_T_CLASS3=0 - 46 with the confidence of 0.60). When mouse cursor is put over the gold ball object above that rule, it instantly becomes evident that the situation has for some reason worsened after the week 3.3.2000–10.3.2000 (these facts are not shown in the example figures).

Concerning coiling phase temperatures, we can see that the intervals between 650 and 740 are handled well, especially the temperatures 710–740 (if coil_t_tgt=710 - 740 then COIL_T_CLASS3=80 - 100 with the confidence of 0.90; the rule in the Fig. 4 with the white dot). An interesting exception here is that temperatures between 680 and 710 are usually "easy" for the process, and in the current matrix, produce a flat rule. From the history of the corresponding rule it can be seen that the situation here has become worse after 25.2.2000–3.3.2000. Below temperature 650, the process is less likely to be able to produce class 3, since there are no corresponding rules for the neighboring items.

Concerning the rule in Fig. 4 with the white dot, we can see that in the temperature interval 710–740, coiling phase temperatures have been most satisfactory during the observation period. However, the rule confidence shows slight, seemingly periodical, deviation during the 10-week time period of the example, although the changes are not very dramatic. To reveal any possible periodical behavior, the observation time should be lengthened. Weeks 4.2.2000–11.2.2000 and 17.3.2000–24.3.2000 have slightly lower confidence than the others. To track down the possible reasons for this, the user could, for example, examine those particular weeks on a daily basis, to find out if there have been any major differences between the days causing the lower average. Another possible way to examine an abnormal week would be to focus on that week using different constraints, or to perform a full search on that week, extracting all the rules containing the chosen temperature item.

To conclude, the general results are evaluated briefly. The presented tool with the visualization model satisfies the definition of knowledge discovery in databases (KDD; Fayyad et al., 1996), and can accordingly be used for identifying valid, novel, potentially useful, and ultimately understandable patterns in the hot strip mill data. It can be used to verify hypotheses, to discover interesting information about the process, and most importantly, by observing rule histories interactively, to reveal time-dependent connections of the associations with the actions made to modify the process.

However, for a domain expert, the tool is at first difficult to use even after a considerable effort put on the design of the user interface and usability aspects. Hence, our experience of practical data mining suggests that the fully efficient use of any data mining tool requires the user to master the data mining methods involved. It is very difficult to set the algorithm parameters correctly or to be able to translate the results fruitfully if the mechanisms of the data mining method itself are not familiar to the user. This can be identified as one of the tenacious problems of data mining technology transfer, hard to overcome as the aim is to swiftly communicate the benefits of new technology into industry. Effort and time has to be put on training the domain experts to know the new methods of data mining.

## 5. CONCLUSIONS

A method and a tool for solving the problem of efficiently visualizing time-dependent changes in association rules were introduced. The rules are presented in a 3D matrix with temporal rule exploration available through simple mouse-over interactions. The matrix, modeled in the VRML2, gives an overview of the rules and of the absence of the rules to the user. The virtual model used as a representation format of the results offers practical benefits such as transferability and hardware independence; furthermore, interactive visualization can be performed using, for example, a standard web browser with a shareware plug-in.

Strengths of the visualization model include the following: the user is given an overview of the simple associations of the specified context, the rule neighbors can be compared effortlessly, and the user can efficiently access more precise information and rule history in an interactive way. Rule history can be used to connect time-dependent associations of the process to different external actions made to modify the process, such as maintenance, repairs, and device updates. A restricting factor in visualization is the matrix size—with large matrices, the advantages of visualization diminish due to performance and display resolution issues of normal PC hardware. To give the user priorities in viewing the rules, a method for calculating rule interestingness was presented. The evaluation of interestingness is mainly knowledge driven; thus the role of the interestingness measures is to give complementary aid to the user, to show which rules are statistically more interesting than

https://doi.org/10.1017/S0890060401155034 Published online by Cambridge University Press

others. The importance of user-centered data mining was emphasized, supporting a constraint-based approach, as opposed to a full search of all rules. The constraint-based approach was considered better due to the substantial increase in performance, smaller memory consumption, and its user-centric nature.

The presented methods have been verified, in practice, with the developed tool which has been applied for mining data on a hot strip mill in a steel plant. Concluded from the experiences of creating and using the tool and the models, the use of virtual modeling for visualizing data mining results in general can be encouraged, thereby heading towards the vision of a 3D information space with instant response interaction in a guiding self-explaining knowledge discovery environment.

## REFERENCES


Agrawal, R., Imielinski, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. *Proc. ACM SIGMOD Conf. on Management of Data*, 207–216.

Agrawal, R., Mannila, H., Skrikant, R., Toivonen, H., & Verkamo, A.I. (1996). Fast discovery of association rules. In *Advances in Knowledge Discovery and Data Mining*, (Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P., & Uthurusamy, R., Eds.), pp. 307–328. Menlo Park, CA: AAAI Press.

Ames, A., Nadeau, D., & Moreland, J. (1997). *VRML 2.0 Sourcebook*, 2nd ed. New York: Wiley Computer Publishing.

Becker, R., & Cleveland, W. (1987). Brushing scatterplots. *Technometrics 29*, 127–142.

Beddow, J. (1990). Shape coding of multidimensional data on a microcomputer display. *Proc. IEEE Visualization*, pp. 238–246. Los Alamitos, CA: IEEE Computer Society Press.

Berndt, D.J., & Clifford, J. (1996). Finding patterns in time series: A dynamic programming approach. In *Advances in Knowledge Discovery and Data Mining*, (Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P., Uthurusamy, R., Eds.), pp. 229–248. Menlo Park, CA: AAAI Press.

Brunk, C., Kelly, J., & Kohavi, R. (1997). Mineset: An integrated system for data mining. *Proc. Third Int. Conf. Knowledge Discovery and Data Mining*, pp. 135–138. Menlo Park, CA: AAAI Press.

Card, S., Macinlay, J.D., & Scneiderman, B. (1999). *Information visualization: Using vision to think*. San Francisco, CA: Morgan Kaufmann.

Fayyad, U.M., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery: An overview. In *Advances in Knowledge Discovery and Data Mining*, (Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P., & Uthurusamy, R., Eds.), pp. 1–34. Menlo Park, CA: AAAI Press.

Han, J., Lakshmanan, L., & Ng, R. (1999). Constraint-based, multidimensional data mining. *Computer 32(8)*, 46–50.

Klemettinen, M., Mannila, H., Ronkainen, P., Toivonen, H., & Verkamo, A.I. (1994). Finding interesting rules from large sets of discovered association rules. *Proc. Third Int. Conf. Information and Knowledge Management*, 401–407.

Kurki, M., Taipale, T., & Korpipää, P. (1998). A large scale fault diagnosis system for a hot strip mill. *Proc. Ninth IFAC Symposium on Automation in Mining, Mineral and Metal Processing*.

Long, J.M, Irani, E.A., & Slagle, J.R. (1991). Automating the discovery of causal relationships in a medical records database. In *Knowledge Discovery in Databases* (Piatetsky-Shapiro, G. & Frawley, W.J., Eds.), pp. 465–476. Menlo Park, CA: AAAI Press.

Mannila, H., Toivonen, H., & Verkamo, A.I. (1995). Discovering frequent episodes in sequences. *Proc. First Int. Conf. on Knowledge Discovery and Data Mining*, pp. 210–215. Menlo Park, CA: AAAI Press.

Nielson, G., Hagen, H., & Muller, H. (1997). Thirty years of multidimensional, multivariate visualization. *Scientific visualization: Overviews, methodologies, techniques*, pp. 3–33. Los Alamitos, CA: IEEE Computer Society Press.

Risch, J., Rex, D., Dowson, S., Walters, T., May, R., & Moon, B. (1999). The STARLIGHT information visualization system. In *Information Visualization: Using Vision to Think*, (Card, S., Mackinlay, J., & Schneiderman, B., Eds.), pp. 551–560. San Francisco, CA: Morgan Kaufmann.

Roddick, J. & Spiliopoulou, M. (1999). A bibliography of temporal, spatial and spatio-temporal data mining research. *SIGKDD Explorations 1(1)*, 34–38.

Skrikant, R., Vu, Q., Agrawal, R. (1997). Mining association rules with item constraints. *Proc. Third Int. Conf. Knowledge Discovery and Data Mining*, pp. 67–73. Menlo Park, CA: AAAI Press.

Smyth, P. & Goodman, R. (1992). An information theoretic approach to rule induction from databases. *IEEE Transactions on Knowledge and Data Engineering 4(4)*, 301–316.

Wang, K., Hock, S., Tay, W., Liu, B. (1998). Interestingness-based merger for numeric association rules. *Proc. Fourth Int. Conf. Knowledge Discovery and Data Mining*, pp. 121–127. Menlo Park, CA: AAAI Press.

Westphal, C. & Blaxton, T. (1998). *Data mining solutions*. New York: Wiley Computer Publishing.

Wong, P.G., Whitney, P., & Thomas, J. (1999). Visualizing association rules for text mining. *Proc. IEEE Symposium information visualization*, pp. 120–123. Los Alamitos, CA: IEEE Computer Society Press.


---

**Panu Korpipää** received his M.Sc. (Eng) degree in 1996 from the University of Oulu, Finland. He is currently a Ph.D. student at the University of Oulu and works as a research scientist at the Finnish state research center, VTT Electronics, in the Knowledge Engineering group. His main research interests include information visualization, context awareness, and data mining.

https://doi.org/10.1017/S0890060401155034 Published online by Cambridge University Press