



DART: a visual analytics system for understanding dynamic association rule mining

Huijun Zhang^{1,2} · Junjie Chen¹ · Yan Qiang¹ · Juanjuan Zhao¹ · Jiangyang Xu¹ · Xiaobo Fan¹ · Yemin Yang¹ · Xiaolong Zhang³

Published online: 5 February 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

Dynamic rule mining can discover time-dependent association rules and provide more accurate descriptions about the relationship among items at different time periods and temporal granularities. However, users still face some challenges in analyzing and choosing reliable rules from the rules identified by algorithms, because of the large number of rules, the dynamic nature of rules across different time periods and granularities and the opacity of the relationship between rules and raw data. In this paper, we present our work on the development of DART, a visual analytics system for dynamic association rule mining, to help analysts gain a better understanding of rules and algorithms. DART allows users to explore rules at different time granularities (e.g., per hour, per day, per month, etc.) and with different time periods (e.g., daily, weekly, yearly, etc.), and to examine rules at multiple levels of detail, including investigating temporal patterns of a set of rules, comparing multiple rules, and evaluating a rule with raw data. Two case studies are used to show the functions and features of DART in analyzing business data and public safety data.

Keywords Visual analytics · Dynamic association rule · Sensemaking · Data mining

1 Introduction

Visual analytics helps users gain insight into data through interactive visualization [1–3]. To respond to the rapid growth of data volumes and the increasing complexity of data types in analysis, researchers have advocated leveraging both human analytical skills and machine computational powers data in user-centered visual analytical systems [4–8].

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00371-020-01803-x>) contains supplementary material, which is available to authorized users.

✉ Xiaolong Zhang
lzhang@ist.psu.edu
Huijun Zhang
zhanghuijunsxm@gmail.com

- ¹ College of Information and Computer, Taiyuan University of Technology, Taiyuan 030600, China
- ² Department of Digital Media Technology, Communication University of Shanxi, Jinzhong 030600, China
- ³ College of Information Sciences and Technology, Pennsylvania State University, University Park, PA 16801, USA

Effective use of powerful data-processing algorithms requires users to have necessary knowledge and skills to interpret and evaluate algorithms and their results [9]. However, users may lack basic knowledge about an algorithm, feel uncertain about its results, or fail to see the implications of the results for real-world applications. Even for users with necessary knowledge about the algorithm, they may want to be more involved in data processing by controlling algorithm parameters so that they can know more about the data, the algorithm, the results, and the semantic implications of the results.

More efforts are needed to help users better understand algorithms and their results. Often algorithms are used in systems as a black box, and users interact with them by just providing inputs and receiving outputs. A more explicit integration of algorithms into visual analytics systems would provide users with visual representations for algorithms, algorithm parameters, and their results.

The focus of our research in this paper is on dynamic association rule mining. Association rule mining [10,11] can help to find interesting relationships among data items based on the frequency of their co-occurrences and has been used in decision-making in various areas. Extended from associ-

ation rule mining, dynamic association rule mining [12–14] provides more information about such relationships by considering temporal features of data.

Current support for the analysis of dynamic association rule mining results is insufficient. Users often face various challenges in understanding and interpreting dynamic association rules identified by algorithms. The first challenge stems from the temporal nature of dynamic rules. Dynamic rules are time-dependent. While some rules may be very strong across all the time (e.g., everyday in the whole week), some rules may be valid only within a specific time range (e.g., weekends only). To know better and more accurately the rules, users may need to analyze them in various ways, such as choosing different temporal granularities (e.g., per hour, per day, per month, etc.) in analysis, or to compare the validity of rules in different time periods (e.g., rules for Monday vs. rules for a week). Another challenge is associated with the massive rules returned by an algorithm [15]. Facing hundreds, or sometimes thousands, of derived rules, users often find it hard to interpret individual rules effectively and select appropriate ones accurately [16]. The complex data structures involved with dynamic rules make the situation even worse. In addition to understanding what data items are included in a rule and whether the rule makes sense, users also need to evaluate and compare rules with several measures, such as support, confidence and lift. Furthermore, similar to other data-mining algorithms, dynamic association rule mining algorithms are opaque to users, making it difficult to understand, explain and use the results obtained from algorithms [9]. Existing data-mining tools, such as Weka [17] and RapidMiner [18], only support static association rules and lack the support for the analysis of dynamic association rules.

In this paper, we report our research on the design of a system to support the visual analysis of dynamic association rule mining. After reviewing relevant work in Sect. 2, the paper presents design requirements in Sect. 3. In Sect. 4, we provide the design details of individual visualization tools. After two case studies in Sect. 5, we discuss the implications of our work and future research directions in Sects. 6 and 7. Our contributions can be summarized as the following:

- An interactive visual analytics system, DART, that helps users to analyze and compare dynamic association rules across various time periods and at different temporal granularities;
- An approach to support multi-level analysis of temporal data, in particular in situations where periodic data are the focus; and
- Two case studies that demonstrate the usefulness of our system in the scenarios of business analysis and public safety analysis.

2 Related work

Our research concerns visual analysis of dynamic association rule mining. Thus, in this section we review literature on association rule mining and discuss research on visualization designs for association rule mining.

2.1 Association rule mining

Association rule mining [10] is a widely used data mining method to identify those data items in a data set that appear together frequently. The input of this method is a set of itemsets, each of which contains several items. The output is a set of rules, which is a set of items. The co-appearances of these items in the input itemsets must meet certain percentage-based measures, such as support and confidence. Association rule mining algorithms search for rules based on raw itemset data and based on user-specified measure values. Various algorithms [19–22] have been developed for association rule mining. In addition, Djenouri et al. [23] presented a bio-inspired approach to improving performance in frequent itemsets mining. Recently, a clustering-based pattern mining technique [24] was developed to support the discovery of relevant rules in data.

An association rule contains a set of items and has several measures. The number of items in a rule varies. The minimum number is 2. A rule itself can be measured by support, confidence, lift and other criteria. Each item in a rule can also be compared by its support. A rule L and its relevant measures can be written as:

$$L = \{I_1, I_2, \dots, I_i, \dots, I_k\} \quad (1)$$

$$S = \{s_1, s_2, \dots, s_i, \dots, s_k\} \quad (2)$$

$$M = \{s, c, l\} \quad (3)$$

where I_i is an item; k is the number of items in the rule L ; S is the support set, of which the member s_i corresponds to the support of the i th item; and M is the measure set for the rule, with three most commonly used measures: s as support, c as confidence and l as lift.

Extended from association rule mining, temporal association rule mining discovers time-related rules from temporal database. Various types of temporal association rule mining methods had been proposed, such as sequence rules mining [25], cyclic rules mining [26], incremental association rule mining [27] and dynamic rule mining [12–14]. More specifically, sequential association rule mining [25] extracts relationships between data items while considering the time ordering from the sequence database; cyclic rule mining algorithms [26] find rules having regular cyclic variation over time from the whole dataset; and incremental association rule mining [27] discovers rules from databases that update

over time, instead of mining the entire dataset from scratch. Although these works take the time factor into account, they still assume that the data characteristics and the underlying associations hidden in the dataset are stable over time, and thus these rules from the whole dataset are also static.

Different from the above methods, dynamic rule mining [12–14] identifies association rules that provide more accurate descriptions about the relationships among items in different time periods and at different temporal granularities. For example, when applied in analyzing people's purchasing behaviors, dynamic association rule mining can help to discover purchasing patterns at different times, such as what gifts people buy together during the holiday season, or what stuffs people usually buy with beers in the evenings of weekends and whether the patterns may differ from weekday evenings. Such temporal patterns may be unavailable under generic association rule mining, because the frequencies of relevant time-dependent records may never be above the minimum frequency due to a low number of instances of relevant records out of all records. To discover these dynamic rules, data sets must first be appropriately clustered according to the purposes of analysis, such as grouping purchasing records based on the month or hour of transaction time in the above examples, and then apply an association rule mining algorithm to individual data clusters.

Let $I = \{i_1, i_2, \dots, i_m\}$ be a set of m different items. Let $D = \{d_1, d_2, \dots, d_n\}$ be a set of n different transactions collected within a time period τ . Let $T = \{t_1, t_2, \dots, t_k\}$ be a set of k time segments, which are disjoint, where $\sum t_i = \tau$, and let A and B be two sets of items, where $A, B \subset I$ and $A \cap B = \emptyset$. The dynamic association rule can be defined as follows:

$$R : A \rightarrow B, \quad (s_1, s_2, \dots, s_k), \quad (c_1, c_2, \dots, c_k) \quad (4)$$

$$s_j(AB) = f_j(AB)/|D_j| \quad (5)$$

$$c_j(AB) = f_j(AB)/f_j(A) \quad (6)$$

where $j \in \{1, 2, \dots, k\}$, s_j and c_j are the support and confidence values of the rule during the time period t_j . $|D_j|$ is the number of transactions collected within the time period t_j and $f_j(x)$ measures the frequency of the set x in D_j . In this paper, the Fp-Growth algorithm [19] was used to extract association rules for consecutive time intervals with different time granularities. Then these derived rules were combined to dynamic rules. Using this approach, the users can observe the changes and fluctuation in the association rules over the time period when these rules are valid.

Although dynamic association rule mining was proposed years ago [12–14], tools to support its use are rarely seen. Different from traditional association rule mining, which often only requires measure threshold of support and confidence, dynamic association rule mining also requires controls over

temporal parameters. Usually, users need to specify certain parameters for data mining [28,29] to obtain such dynamic rules as hourly patterns in each day, daily patterns in each week, or monthly patterns in each year. Thus, users often need to examine temporal patterns to find interesting rules and modify time granularity back and forth based on previously found rules and involved raw data. This is where interactive visualization can help.

2.2 Visualization designs for association rules

Research has shown that visualization can facilitate association rule mining from three aspects: visualizing the rules, assisting rule evaluation, and controlling rule generation [30].

2.2.1 Visualization of rules

Most of the data-mining tools list the derived results in text, but visualization of data-mining results can provide immediate insights into important features of algorithms. Liu and Salvendy [31] argued that visualization of association rules should present all the rules generated by an algorithm, show interesting items involved in a rule, and provide effective interestingness measures.

Some visualization designs have been proposed for visualizing association rules. For example, rules can be visualized as a grid [32,33], a node-link network [34,35], parallel coordinates [36–38], or information landscape [39,40]. Users can also examine the details of the rule subset or a specific rule with tools like SARV [35], which presents rules with three synchronous views: a matrix view for rule preview, a node-link view to show the relationship among selected rules, and a view to display texts of the selected rules and items.

2.2.2 Assisting rule evaluation

Rule evaluation is a fairly complicated process, which often requires users to examine and analyze a significant amount of rules. Various methods have been proposed to evaluate the interestingness of rules [16] with objective measures (data-based methods) and/or subjective measures (user-oriented methods). Bruzzese et al. [41], targeting for objective measures, defined utility index for items of rules to measure the impact on confidence exerted by the inclusion or non-inclusion of a certain item in a rule, and then used parallel coordinates to visualize the association rules and the utility index of each item. Berzal et al. [42] introduced an assessment framework based on Shortliffe and Buchanan's certainty factors [43] to discard misleading rules. In Liu et al. [44], a subjective interestingness method was used to measure the unexpectedness and actionability of rules based on user prior knowledge. And further, Delgado et al. [45] discussed a good assessment measure for association rule

evaluation should fulfill and provided a new formulation for both strong and very strong rules based on a logical model. With visualization techniques, users can explore the rules of interest more effectively.

2.2.3 Controlling rule generation

With data-mining algorithm being considered as a highly automated model, visualization techniques have been combined with algorithm modeling processes recently in many studies [46,47]. In a model building process, visualization plays various roles, including displaying the operation results obtained in model building, fulfilling interactive functions to enable user participation in model building, and giving feedback about user operations.

When visualization is combined with a modeling process of association rules, it is necessary to increase the participation of users [8]. For association rule mining algorithms, Liu et al. [44] first proposed an interactive visual exploration tool to control the derivation of rules. Similarly, Chen et al. [32] used visual analysis techniques for modulating the constraints during iterative mining processes. Recently, Zhao et al. [48] focused on progressive techniques that execute data-mining processes step-by-step and show results to facilitate analysts to detect interesting patterns and factors effectively and efficiently.

In sum, research has been done to support the visualization and evaluation of association rule mining algorithms and their results. However, to our best knowledge, tools to support the understanding of dynamic association rule mining and derived results are rare. Our research is an effort to fill the gap.

3 Design requirements

The focus of this research is on the design of a user-driven visual analytics system for dynamic association rule mining. In this section, we present the analysis of design requirements.

Our goal is to design a set of visualization tools to help users better understand and analyze the results of dynamic association rule mining, as well as control rule mining processes. Our design is grounded in interviews with two data analysis experts, whose work involved the use of dynamic association rule mining. One expert is a business analyst who often needs to analyze business data for marketing. The other works on public safety data analysis. Both experts were asked to show us how they performed data analysis in their work. They demonstrated their tools, such as Weka [17] and Tableau Desktop [49], for data analysis. To this end, based on the observation of their analysis process, we conducted several rounds of discussions and system prototyping with

them, and the following requirements were distilled based on such work to guide the system development.

R1: Support temporal pattern driven analysis Here, the temporal pattern is defined as the distribution of dynamic rules in the time dimension. To obtain dynamic rules and their temporal patterns, the experts first divided the dataset into multiple disjoint sub-datasets according to a certain time granularity (e.g., hour, day, week, or month) and a desirable temporal period (e.g., daily, weekly, monthly, or yearly). Then, these data subsets were separately imported into the Weka tool to perform association rule mining. Weka presented the static rules obtained in each run as a list with various numerical measures (e.g., support, confidence, or lift). Next, these rules derived from different subsets of data were integrated into a file to examine temporal patterns, which was a daunting task even with tools like Tableau. Furthermore, both experts noted that they often need to examine and compare temporal patterns under multiple time granularities to get a comprehensive understanding of the data, but none of available software packages allows them to do so. Thus, they hoped to get a tool that can help them get an overview of the dynamic rules across all time periods and at different time granularities.

R2: Support item-driven analysis Through the investigation, we learned that analysts usually have some preliminary ideas in their minds before starting data analysis. Often they want to know about some specific information from a particular perspective. For example, the business analyst mentioned that he always wanted to know information about particular products (such as newly promoted goods, high-margin items, slow-moving stock), rather than the rules for all products. This goal requires tools for item-driven rule analysis. Combining temporal pattern-driven and item-driven analysis could help the selection of specific rules based on temporal criteria and item interest.

R3: Support detailed analysis of dynamic rules Both experts said they often analyzed rules in different levels of detail. After selecting several potentially interesting rules based on temporal patterns and/or data items involved, they usually viewed and manipulated the details of these rules (e.g., support, confidence, lift) at various time periods to further narrow down the number of rules to be analyzed. They hoped to have tools for the analysis of rules in different ways, such as sorting rules based on various criteria, comparing rules that contain similar items, etc. They preferred intuitive and user-friendly visualization-based tools.

R4: Support the analysis and interpretation of individual rules The experts indicated that for ordinary users, the black box nature of the algorithm makes the results difficult to understand. In particular, they felt that without knowing the relationship among raw data and derived rules, it was hard to judge where these rules came from, whether they were correct and valid, and which rules should be chosen. They felt

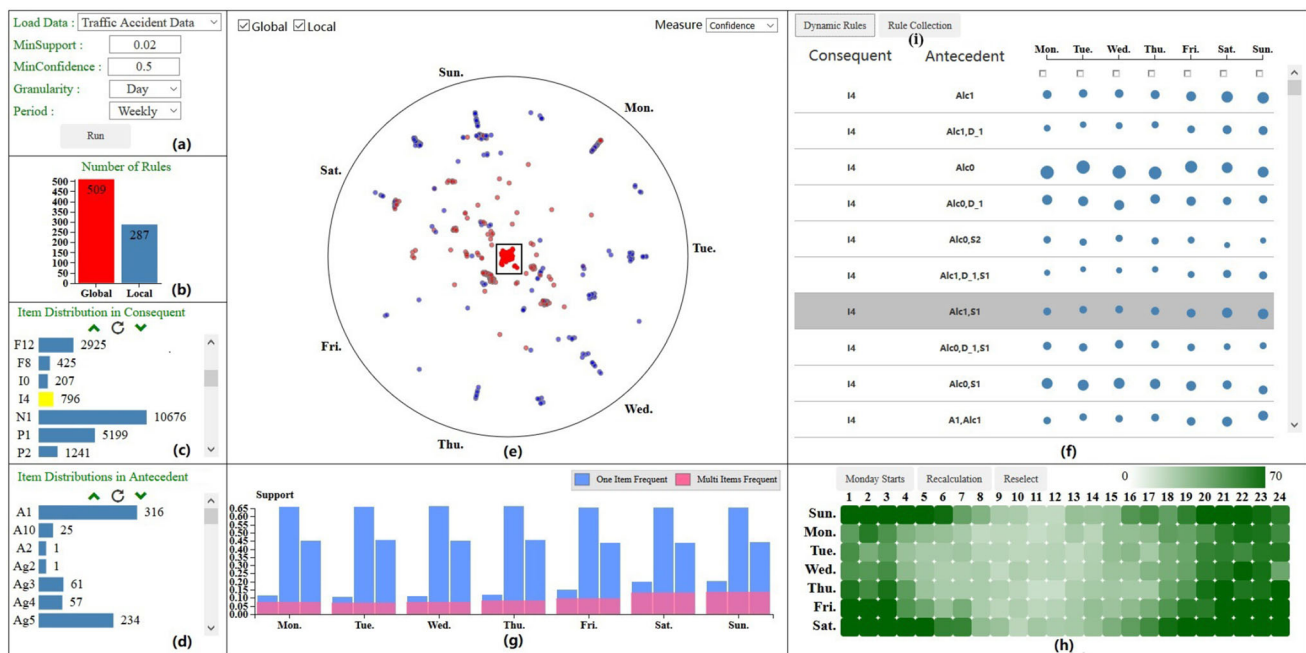


Fig. 1 User interface of DART: **a** parameter panel to set parameters and start analysis; **b** summary view to show the number of rules; **c**, **d** views to show the items appearing in rules; **e** overview to show temporal patterns of multiple rules; **f** rule comparison view for the evaluation of

multiple rules; **g** itemset view for the analysis of a rule and its frequent items; **h** rule data distribution for the examination of raw data related to a rule; and **i** tabbed view panel for rule collection (not shown here)

that to better interpret the results and choose correct rules, it was important for them to have a deeper understanding of the relationships among the data, the algorithm parameters, and the results. Specifically, they hoped that they could examine the frequent items and items corresponding to individual rules to evaluate the rules and understand the semantics of the rules.

R5: Support the collection and management of rules

Both experts indicated that they needed tools to help them collect and organize rules, in particular when the number of the rules generated was significant. If they needed to collect a rule of interest, their current practice was very similar: copying and pasting the rule to a separate document (e.g., a text file or a spreadsheet). This approach is less efficient because they had to move back and forth between the analysis tools (e.g., Tableau) and the rule collection documents to search and compare rules. They hoped to have a tool that allows them to collect and compare rules of interest directly and interactively.

4 Visualization design

Our system, DART, was designed based on these requirements. Our design includes several panels that support user interaction and visualize dynamic rules in different perspectives (Fig. 1). Panels for parameter control and rule summary

(Fig. 1a–d) on the left are presented to set up parameters and summarize the rule results. The overview (Fig. 1e) in the upper middle shows temporal pattern of multiple rules. The rule comparison view (Fig. 1f) on the top right allows users to view and manipulate the details of dynamic rules. The itemset view (Fig. 1g) and the view of rule distribution (Fig. 1h) are updated by selecting individual dynamic rule from the rule comparison view. In this section, we first present the data attributes and structures of dynamic rules used in visualization design and then describe individual panels and views. Tools we used to implement the system are also briefly introduced. The brief introduction and analysis process about the system can also be found at <https://www.dropbox.com/s/aulwshhu6ln256u/DART.mp4?dl=0>.

4.1 Data attributes and structures of dynamic rules

A dynamic association rule actually is a set of simple association rules that contain the same set of items, but have different measures at different time points. For example, a dynamic rule that describes the hourly purchasing patterns of beer and diaper often have different supports and/or confidences in each hour during a day. Thus, the analysis of a dynamic rule actually involves the analysis of multiple simple rules with the same frequent items but different measures in the time domain. The itemset of these rules can still be described by Eq. (1), because of the same set of items they

have, but their measures become matrices as shown below, rather than what Eqs. (2) and (3) describe:

$$\mathbb{S} = \begin{pmatrix} s_{1,1} & s_{2,1} & \cdots & s_{i,1} & \cdots & s_{k,1} \\ \vdots & \vdots & & \ddots & & \vdots \\ s_{1,j} & s_{2,j} & \cdots & s_{i,j} & \cdots & s_{k,j} \\ \vdots & \vdots & & \ddots & & \vdots \\ s_{1,t} & s_{2,t} & \cdots & s_{i,t} & \cdots & s_{k,t} \end{pmatrix} \quad \mathbb{M} = \begin{pmatrix} s_1 & c_1 & l_1 \\ \vdots & \vdots & \vdots \\ s_p & c_p & l_p \\ \vdots & \vdots & \vdots \\ s_t & c_t & l_t \end{pmatrix} \tag{7}$$

where t is the total number of time points in analysis, \mathbb{S} is the item-support matrix, in which $s_{i,j}$ is the support of the i th item among all k items at time j ; and \mathbb{M} is the rule-measure matrix in which s_p , c_p , and l_p are the support, confidence, and lift, respectively, of the p th rule out of the total t rules.

The visualization design in our research focused on the visualization of data structures seen in Eqs. (1) and (7), as well as other information relevant to these structures.

4.2 Panels for parameter control and rule summary

The left side of the user interface (Fig. 1) is a set of panels to set up algorithm parameters and summarize the rule results. Users start an analytical process by selecting a data set and specify algorithm parameters (e.g., minimum support, minimum confidence, temporal granularity and period) in the upper-left panel (Fig. 1a). The histogram below it shows the total counts of dynamic rules returned by the algorithm (Fig. 1b). The histogram color-codes the rules that are valid in the whole time period, or *global rules*, and those only valid in certain periods, or *local rules*. Their heights indicate the total numbers of global and local rules.

Below the summary histogram is a bar chart to show the item distribution in consequent of dynamic rules (Fig. 1c). All these bars can be toggled on and off by clicking. Toggling on an item will only keep those rules that contain the item in consequent for further analysis.

Under this bar chart is another bar chart, in which users can check what other items are contained in the antecedent of rules that are kept and the corresponding rule distribution (Fig. 1d). Items in this bar chart can also be toggled on and off for further rule filtering.

These two bar charts allow users to examine what items are involved in rules and initiate item-driven rule analysis (R2). Rules returned from the algorithm or filtered by users will be projected to the overview panel where their temporal patterns can be analyzed.

4.3 Overview of dynamic rules

The overview (Fig. 1e) allows users to gain a big picture of a set of dynamic rules and lets users choose some rules for

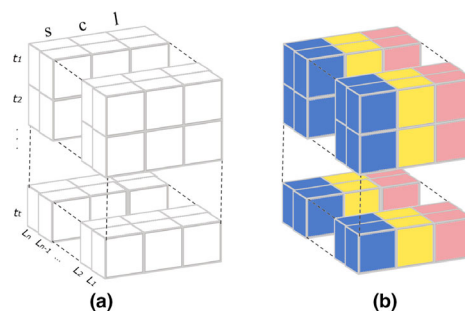


Fig. 2 Data cube involved in analyzing multiple rules: **a** 3 dimensions of the data cube: measuring dimension with 3 measures (support, confidence, and lift), time dimension with t time points, and rule dimension with n rules; **b** data slices, distinguished by color

further analysis based on their temporal distribution patterns. When comparing multiple rules, users often need to evaluate them based on certain measures that are available in their rule-measure matrix, or \mathbb{M} in Eq. (7). However, visualizing multiple matrices could be a challenge, because putting these matrices together actually makes a data cube (Fig. 2a) with three dimensions: measure dimension, rule dimension and time dimension. The rule dimension is fixed with three measures, but the number of rules in the rule dimension and time points in the time dimension vary from analysis to analysis.

To deal with the challenge in visualizing a data cube with two undetermined dimensions, we adopted an approach to convert the visualization of the whole data cube into the visualization of user-controlled three data slices (Fig. 2b)—support slice, confidence slice, and lift slice, because in rule analysis users usually compare rules with the same criterion. Each slice is a matrix, in which one dimension has individual rules, and the other has time points. The visualization in the overview is based on data slice specified by users through a drop-down menu, as seen in Fig. 1e.

The interest here is the temporal distribution patterns of the measures of these rules. On the surface, the measures of a dynamic rule may look like multidimensional data. For example, for a rule that is valid for 7 days each week, it has 7 supports, one for each day; or a rule that is only valid in weekends has 2 supports for Saturday and Sunday.

However, in nature the measures of dynamic rules are periodic data. Unlike regular multidimensional data, periodic data usually implies a specific order of individual dimensions and data analysis of periodic data must consider dimension order and adjacency. Thus, traditional tools for multidimensional data, such as parallel coordinates [50] and scatterplot matrix [51], are inappropriate here.

Our visualization design in this overview is based on Radviz [52,53]. Radviz considers both the multidimensional and periodic characteristics of data and maps data into a 2D plane. In our design, each time dimension is designed as an anchor and all anchors are evenly distributed on a circle. Differ-

ent choices of temporal granularity and period will lead to different numbers of anchors. For example, when analyzing daily rules in a week, there will be 7 anchors corresponding 7 days. If the interest is monthly rules in a year, 12 anchors are needed.

Based on its measure, each rule is visualized as a dot inside a circle, and the dot location is determined by the measures of the rule at all time points, as if the dot is tied to each anchor by a spring, the stiffness of which is determined by the rule measure in the dimension. For example, a rule that is valid on weekends will be visualized as a dot sitting between the Saturday and Sunday anchoring points and its location is determined by the measures on these two days; a rule that is only valid on Friday will be a dot on the radius connecting the center of the circle and the Friday anchoring point, and the bigger the rule measure is, the closer the dot is to the anchoring point. It should be noted that all measures used to calculate the locations of the dots are normalized to a value between 0 and 1. A value of 1 will put a dot on the circle, while with a value of 0 a dot will be at the center.

The overview also color-codes global and local rules as red and blue, respectively. Two check-boxes are provided in this view to let users decide the visibility of global and local rules.

The designs of the overview support temporal pattern driven analysis (**R1**). Users can quickly see the distribution patterns of rules, such as dot clusters where similar rules gather, or lone blue dots close to the circle, which may be unique local rules. Users can select such dots in the overview as the rules of interest to further compare their distribution patterns or to examine their details.

4.4 View for rule comparison

The view for rule comparison (Fig. 1f) supports the exploration and comparison of those rules of interest (**R3**). This view is designed as a table. Each row is a rule, and the columns include rule items and time points. The two left-most columns are reserved for items: one for items selected by users in the control panels and the other for the rest items in a rule. The other columns are for time points, and the number depends on the temporal granularity and period in analysis: 7 columns for an analysis of daily patterns in a week (7 days), or 24 columns for hourly pattern analysis in a day.

In each rule, its measures are visualized as circular dots. The size of a dot is determined by the corresponding support of the rule, and its vertical location is by its confidence. Our visual encoding considers only these two measures largely because of their popularity.

Users can interact with the table in various ways. They can order rules based on their measures at each time point through a drop-down menu. Users can also learn more about a rule by hovering the cursor over it. Hovering over the text

of a rule on the two left columns brings up a tooltip to tell what this rule is about. Hovering over a dot will show the quantitative measures at a particular time.

With these designs, users can compare rules directly based on the visualized measures at different time points and then identify those rules that need further investigation. Such rules can be those with the larger dots (higher support) compared with other rules with similar temporal patterns, or those with dots only appearing at some times. To analyze such rules of interest, users can look at the detail of a rule, such as its itemset composition and the strength of item support in the itemset view.

4.5 Itemset view

The itemset view (Fig. 1g) is designed to let users see the item composition of a rule to assist the understanding and evaluation of the rule (**R4**). When users select a rule from the rule comparison view, its items are displayed as a set of histograms. The bars in a histogram represent the supports of the rule and the supports of its items. The horizontal axis represent time. The data involved in this view includes the supports of a dynamic rule, i.e., the first column of the rule-measure matrix \mathbb{M} in Equation (7), and the item-support matrix \mathbb{S} .

The view usually contains multiple groups of bars. Each group corresponds to a rule at a time point. All groups have similar patterns: the same number of bars and comparable bar heights. Blue bars represent the items of the rule, and their heights are mapped to their supports. The pink bar at the bottom of a bar group represents the rule, and its height is proportional to the support of the rule itself. Because the support of a rule is usually smaller than the supports of all its items, the pink bar is generally lower than the blue bars, making it easy for users to see all items.

A rule that is valid at different times may have different supports across time and the supports of its items may also vary from time to time. Thus, the bar heights in all groups are usually inconsistent, allowing users to easily compare the temporal patterns of the supports of the rules and its items and evaluate the strength of the rule.

In addition to seeing directly the items involved in a rule and their supports, users can also obtain more information about the rule and its items interactively. Hovering the cursor over a bar group, users can read these items and their supports in a tooltip.

With this view, users can know more about how a rule is made, and compare the rule across different time points. Doing so will allow users to better understand the principles of dynamic association rule mining and enhance their skills in the evaluation of a rule (**R4**).

4.6 View for rule data distribution

The view of rule data distribution (Fig. 1h) is designed to help users further investigate the temporal characteristics of dynamic rules (**R4**). The goal of this view is to let users verify a rule based on the temporal granularity and period given in the beginning of analysis. These two parameters are often estimated and chosen based on the intuition, knowledge, or experience of a users, but temporal patterns of dynamic rules obtained from real-world data may not follow the same temporal patterns as what the user may think. For example, in analyzing the purchasing behaviors in a store, the user may take day as the granularity and week as the period, but interesting shopping patterns may occur from the evening of a day to the early morning the next day, a period that does not fit into what the user has defined. Consequently, such purchasing patterns, even through significant enough to become a rule, may never be discovered, if the relevant records are spread into multiple periods and become infrequent in any of them.

The data distribution view helps users to see how data related to a rule is distributed across time and guides users to choose new and probably more appropriate parameters for rule mining. The view appears as a calendar-like heatmap to show the frequencies of itemsets that contain all frequent items of the rule. The temporal scale of this view is determined by the time granularity specified by the user. Each cell in the view represents a time period at the granularity that is one level finer than the specified granularity. For example, for an analysis at a granularity of day, each cell in the view represent an hour. The color of a cell is based on the frequency of all relevant itemsets during the time the cell represents.

This view serves two purposes. First, it allows users to examine distribution patterns of data and verify the rule. For example, the view in Fig. 1h, which is about the hourly patterns of data in an analysis of the daily pattern in a week, shows that data distributions vary significantly between day and night: heavy activities at night but light in day. This implies that using day as the time granularity for rule mining may not be the best choice and results may be inaccurate. The second purpose of this view is to let users directly define new time granularity for rule mining. Based on what they see, users can manually specify the boundaries of time periods by clicking relevant cells. New time periods do not have to follow natural time units, such as day or week, and they can even be unequal. New rules can be generated with the changed parameters.

This data distribution view supports the analysis based on temporal patterns of data (**R1**) by drawing on more detailed information from another level (**R4**). Furthermore, with the help of this view, users can start with roughly-defined parameters and then use intermediate results to fine-tune them for more accurate results.

4.7 View for rule collection

Our system provides a view for collecting and managing rules of interest (**R5**). The view is in a tabbed panel (Fig. 1i) together with the panel that contains the rule comparison view. A rule can be added into the collection by double-clicking it in the rule comparison view. Its layout is very similar to that in the rule comparison view. Users can delete a rule from the view if it is no longer interesting.

4.8 Tools used for design and implementation

DART is a web-based visual analytic system. The front end focuses on visualization and interaction functions and was built with HTML5, D3.js, and jQuery.js. The server end provides computation and data management services and was implemented with Java and MySQL.

5 Usage scenarios

In this section, we describe two case studies to demonstrate the functions and features of DART. The first case study involved a sales analyst as the subject to analyze online retail data, while in the second case a transportation expert was recruited to use DART to analyze data related to fatal car accidents. Neither subject had prior knowledge on dynamic association rule mining before the study. We introduced the basic concepts of dynamic association rule mining and other relevant concepts, such as rule measures and itemsets, through system demonstration. We also asked them to talk aloud about their actions during the study session. In this section, we describe the first case study in a brief manner to show how the system was used, and provided more detail on the second case study to show how DART can be used for in-depth analysis of dynamic rules.

5.1 Online retail data analysis

Our data in this study were from an online database [54], and the data set we used contain 25,899 valid transactions in an online store occurring from December 1, 2010 to December 9, 2011. The subject was interested in the relationship among gift-related products sold during the holiday season, so he used month as the time granularity and year as the period. Some relevant information displayed in the system interface during the analysis is shown in Fig. 3. Specifically, the overview (Fig. 3a) presents a big picture of all rules in every month. Items that appear in the antecedent and consequent of rules are shown in Fig. 3b, c, respectively. A rule of interest to the user displays in Fig. 3d, and the frequent itemset and the rule data distribution corresponding to this rule are shown in Fig. 3e, f. After the user adjusted the parameters

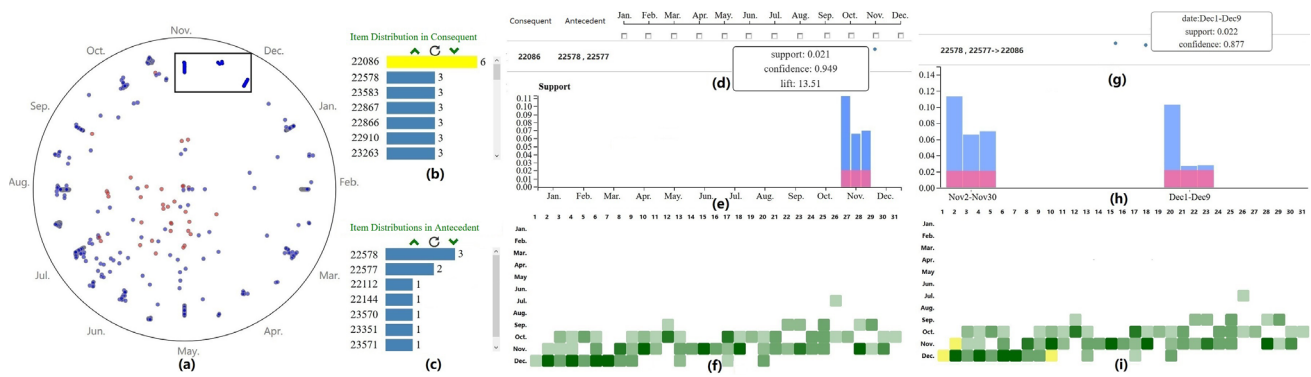


Fig. 3 Online retail data analysis: **a** overview; **b**, **c** items appearing in rules; **d** temporal pattern of a rule; **e** frequent items of the rule; **f** data distribution of the rule; **g** new rules after adjusting parameters; **h** frequent items of a new rule; and **i** adjusting parameters

(Fig. 3i), a new rule is generated (Fig. 3g), and its frequent itemset is shown in Fig. 3h.

Examining temporal patterns of dynamic rules The overview gave a big picture of all rules identified by the system (Fig. 3a). The subject said, “I want to view the relationship between the products sold during the Christmas holiday season, so I intend to examine the rules that only appear in November and December.” Then the subject selected the rule cluster shown in Fig. 3a for analysis.

Exploring and comparing dynamic rules Among 75 selected rules, the subject wanted “to look at the specific semantics of these rules, that is, which products they are about”. The subject checked the data items contained in these rules and sorted them in a descending order based on the number of rules that contained them, as shown in Fig. 3b. Through exploration, the subject take 22086 as the interest, which is a Christmas gift—*paper chain kit 50’s Christmas*, and explored all rules including it. After comparing 6 relevant rules, the subject finally took the rule with the highest confidence. The rule included three items: 22086 (*paper chain kit 50’s Christmas*), 22577 (*wooden heart Christmas Scandinavian*), and 22578 (*wooden star Christmas Scandinavian*), and was valid only in November (Fig. 3d).

Exploring items, rule data and new rules To further understand how this rule was derived from data, the subject clicked on the rule in the rule comparison view to explore its frequent items and the data distribution, as shown in Fig. 3e–f. After observing the data distribution of the rule, the subject said that “this rule appeared only in November. However, in fact, I noticed that there are many data records in early December. Maybe I should change the time periods to see whether the rule is good in early December too.” As shown in Fig. 3i, the subject set three new time points: November 2nd, December 1st, and December 10th to cut the time into four periods: before November 2, between November 2 to November 30, between December 1 to December 10, and after December 10. Running the algorithm with the new parameters, the subject got two new rules: one for Novem-

ber and the other for December 1 to December 10. After comparing these two rules with the previous one, the subject believed both rules were as good as the previous one, with similar support and confidence. The new rule on December indicated that if promotion for these items were needed, the promotion period should include the whole November as well as early December, rather than just November.

5.2 Traffic accident data analysis

The data used in this case study were from FARS (Fatality Analysis Reporting System) [55]. It contains 72,591 records of car crash accidents in 2011 and each accident contains such attributes as driver age, driver gender, driver alcohol test result, driver drug test result, road condition, crash date, injury severity, etc.

5.2.1 Obtaining a big picture of rules

The subject first looked at daily patterns in a week and obtained 20,468 global rules and 10,693 local rules. The subject was interested in fatal accidents. All these accidents have a value of I4 in the attribute of injury severity. Thus, the subject clicked the bar representing I4 in the item distribution chart to narrow down the number of rules. This choice reduced the rules to 509 global rules and 287 local ones (Fig. 1b). Next, the subject examined their temporal patterns in the overview (Fig. 1e), and felt that the overview was very “beneficial” to having a big picture of these rules. Then he checked the data items contained by these rules, and saw many human-related factors. Interested in such driver-related factors, he decided to focus on alcohol test result and driver age in his analysis.

5.2.2 Investigating alcohol factor

First of all, the user investigated the impact of alcohol factor on fatal traffic accidents at weekly granularity. As shown in

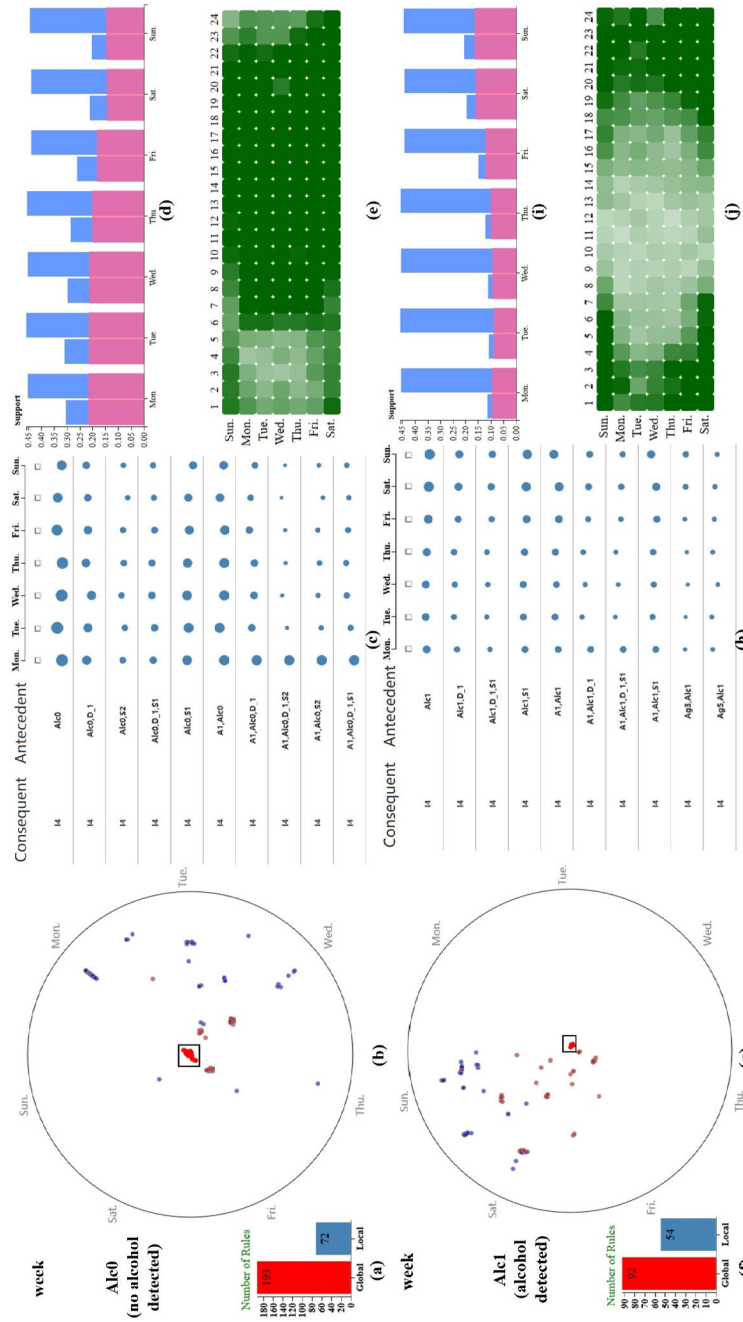


Fig. 4 Analysis of daily rules on alcohol factor and fatal injury: **a** numbers of rules having Alc0 and I4; **b** overview of the rules having Alc0 and I4; **c** comparison of rules having Alc0 and I4; **d** frequent items of a rule (Alc0, I4); **e** data distribution of the rule; **f** numbers of rules having Alc1 and I4; **g** overview of the rules having Alc1 and I4; **h** details of frequent items of a rule (Alc1, I4); and **j** data distribution of the rule

Fig. 4, the system provided some important information in the user's analysis process. For details, please refer to the user's analysis process below. The subject first examined the rules that contain negative alcohol test result (coded as Alc0). The overview (Fig. 4b) showed the rules in the middle of the circle as well as spreading over weekdays. Seeing some rules at the center of the circle, the subject wanted to "explore these rules that seem good at every day and want to see if they may differ between weekdays and weekends".

Selecting a rule cluster at the center (Fig. 4b), the subject began to examine the supports and confidences of these rules in the rule comparison view (Fig. 4c). The rule at the top of the table, (Alc0, I4) interested the subject, so he clicked it to check its frequent items (Fig. 4d) and the distribution of relevant accident records (Fig. 4e). He indicated that these two views were very "informative" to understanding the relationships among dynamic rules, itemsets, and raw accident records. Helped by these views, he was confident about the validity of the rule, as evidenced by the high supports and the fact that the accidents happened largely during day time and early evening from 6am to 10pm. Thus, he collected the rule.

The subject used the same procedure to analyze rules that contain positive alcohol test result (coded as Alc1). He noticed the number of the rules was fewer (Fig. 4f), compared with that with negative test result. He also saw a different distribution pattern in the overview (Fig. 4g): although there were still rules concentrating in the center, other rules actually appeared mostly during weekends. Selecting a rule cluster at the center again, the subject browsed the measures of the rules in the cluster (Fig. 4h). After checking the frequent items of the rule (Alc1, I4) (Fig. 4i) and its data distribution (Fig. 4j), he found another difference: most alcohol-related fatal accidents happened at night. Seeing this rule valid too, he collected it.

In the rule collection panel, the subject compared these two rules side by side (Fig. 5). He said, "the support values of the rule with Alc0 and I4 during weekdays are considerably higher than those at weekends. In contrast, the support values of the rule with Alc1 and I4 during weekends are significantly higher than those during weekdays, but their confidences do not vary too much." He also noticed that the rule (Alc1,I4) had lower supports than the rule (Alc0, I4) almost every day (measured by dot size), but its confidence levels (the vertical position of the dot) seemed higher consistently. He summarized what these rules implied as "well, among fatal traffic accidents, more may be caused by people not drinking alcohol, but drinking alcohol is more likely to cause such accidents". Drawing on his expertise, he liked these rules.

To understand more the relationship between alcohol test result and fatal traffic accident, the subject then analyzed the rules related to alcohol factor at two other temporal levels:

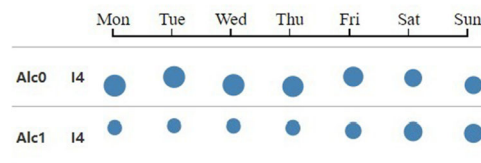


Fig. 5 Comparison of rules related to alcohol test result

monthly patterns and hourly patterns with the same procedure as the above.

The overview of the monthly patterns shows high rates of alcohol-related accidents from April to July and in December (Fig. 6g), while accidents with negative alcohol test result were scattered more broadly across the year (Fig. 6b). After the detail analysis of two rules—(Alc0, I4) and (Alc1, I4), the subject found that their supports and confidences did not change too much in the whole year (Fig. 6c, h). Their frequent itemset views (Fig. 6d, i) and data distribution views (Fig. 6e and 6j) indicate these two rules were fairly stable.

For the hourly patterns, the rules containing Alc0 appeared more during day time and from 5am to 2pm (Fig. 7b), while the rules containing Alc1 were more at night, from 9pm to 5am (Fig. 7g). Analyzing the details of two rules—(Alc0, I4) and (Alc1, I4) indicated that the supports of the rule (Alc1, I4) were larger at night (larger dot size), while the confidences were higher during the day time than at night (higher dot position) (Fig. 7h). In contrast, the supports of the rule (Alc0,I4) were higher during the day time, and the confidences were much more stable (Fig. 7c). The data distribution shown in Fig. 7e, j further confirmed the validity of these two rules.

5.2.3 Investigating age factor

The subject also analyzed the relationship between driver age and fatal accident. In our system, age values were coded as Ag1, Ag2, Ag3, Ag4 and Ag5 to represent to 5 age groups, respectively: child, teen, young adult, adult, and senior.

The subject examined the overall patterns of the rules related to each age group. Results showed very few rules related to Ag1 and Ag2. He found the rule patterns for Ag3 and Ag4 were similar. Figure 8 compares the rule patterns for these two age groups at different levels side by side. As seen, the weekly rules for these two groups were mainly distributed in weekends (Fig. 8a, d), their monthly rules concentrated between May and August (Fig. 8b, e), and their hourly rules were largely distributed between 1am to 12pm (Fig. 8c, f). Further analyzing some weekly rules on Ag3 that were only valid during weekends (those being selected in Fig. 8a), the subject saw the presence of Item S1 in many rules (Fig. 9a), indicating that all drivers were male.

The subject then saw something interesting on the rule (Ag3, I4, S1): "this rule only appears on Sunday, so I am

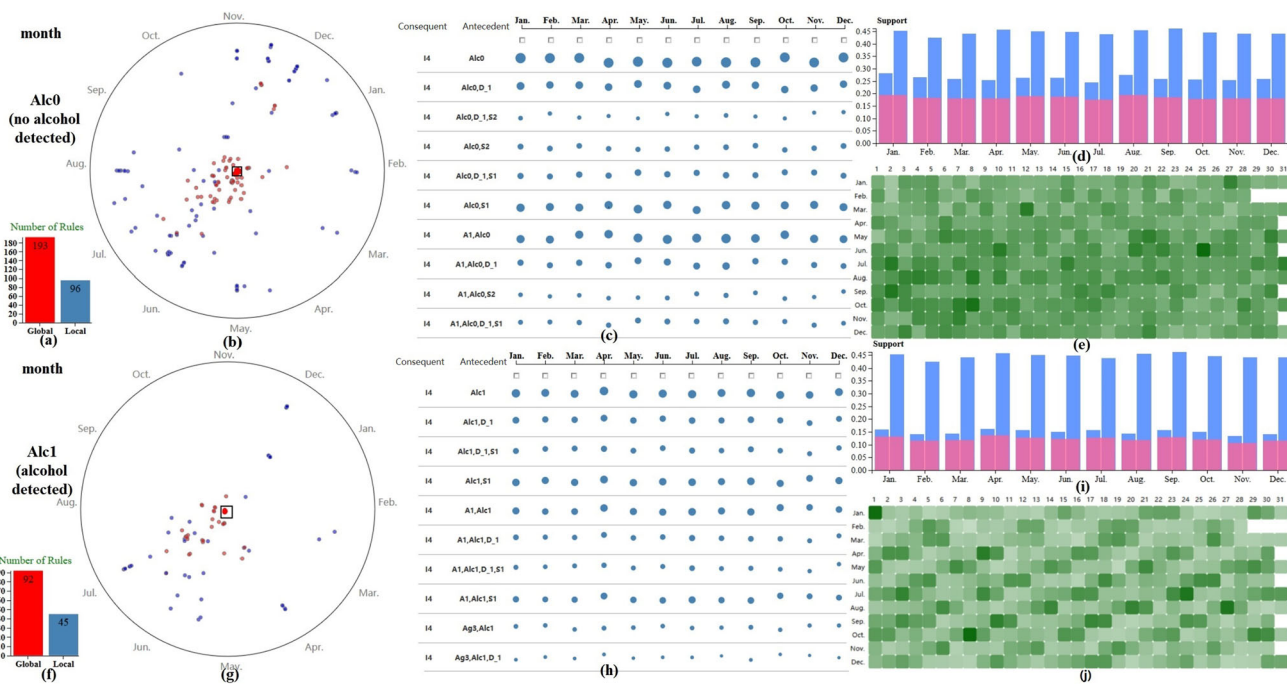


Fig. 6 Analysis of monthly rules on alcohol factor and fatal injury: **a** numbers of rules having Alc0 and I4; **b** overview of the rules having Alc0 and I4; **c** comparison of rules having Alc0 and I4; **d** frequent items of a rule (Alc0, I4); **e** data distribution of the rule; **f** numbers of rules

having Alc1 and I4; **g** overview of rules; **h** details of rules having Alc1 and I4; **i** frequent items of a rule (Alc1, I4); and **j** data distribution of the rule

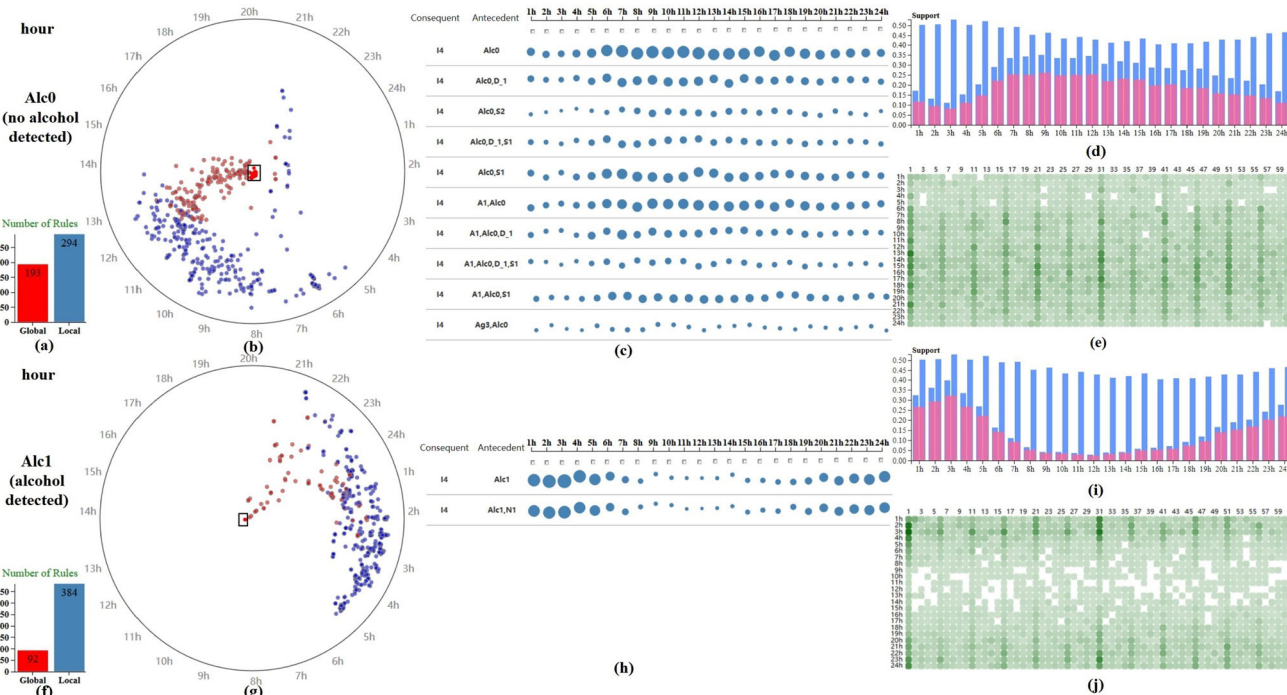


Fig. 7 Analysis of hourly rules on alcohol factor and fatal injury: **a** numbers of rules having Alc0 and I4; **b** overview of the rules having Alc0 and I4; **c** comparison of rules having Alc0 and I4; **d** frequent items of a rule (Alc0, I4); **e** data distribution of the rule; **f** numbers of rules

having Alc1 and I4; **g** overview of rules; **h** details of rules having Alc1 and I4; **i** frequent items of a rule (Alc1, I4); and **j** data distribution of the rule

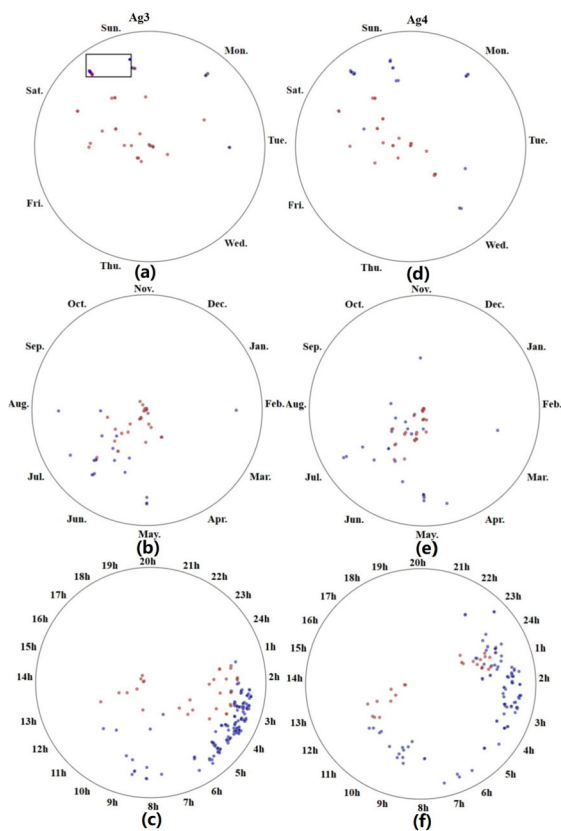


Fig. 8 Rule patterns for Ag3 and Ag4. **a–c** weekly, monthly, and hourly rule patterns for Ag3; and **d–f** patterns for Ag4

going to explore it further.” He analyzed its frequent itemsets (Fig. 9b) and data distribution (Fig. 9c) and found that the relevant data concentrated from 10pm on Friday to 8am on Saturday. The subject redefined time periods so that two night periods—10pm on Friday to 8 am Saturday and 10pm Saturday to 8am Sunday—would be singled out (Fig. 9f). The new parameters led to two new rules that corresponded to the two night periods, with comparable supports and better confidence (Fig. 9e). Comparing the rules obtained before and after the change of time parameters, the subject was confident that the new rules were “more accurate and more reliable” than the old one and attributed the improvement to “more accurate control of time periods”.

Finally, the subject analyzed the weekly rules related to senior drivers (Ag5). Unlike the rules on Ag3 or Ag4 (Fig. 8a, d), the rules on Ag5 appeared almost everyday (Fig. 10a), and many rules contained Alc0 (no influence of alcohol). The subject chose a rule (Ag5, Alc0, I4) to check its frequent itemsets (Fig. 10c) and data distribution (Fig. 10d). His conclusion was that “accidents involving this age group largely occurred during day time, but afternoon hours were more dangerous. Need further investigation to explain why.”.

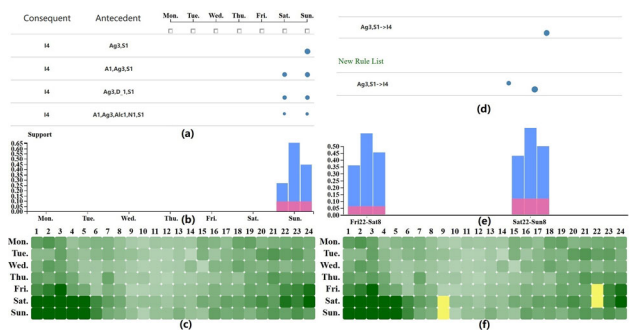


Fig. 9 Analysis of Ag3 and fatal accident: **a** rules having Ag3 and I4; **b** itemset of Ag3, S1 and I4; **c** data distribution; **d** rule collection with the old rule and new rules; **e** itemsets of S1, Ag3 and I4 under a new rule; and **f** changing time parameters for mining

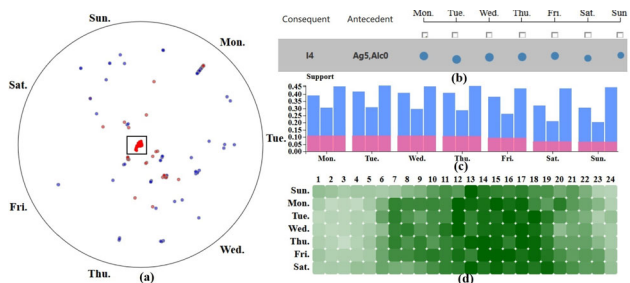


Fig. 10 Analysis of rules involving senior drivers Ag5: **a** the overview of rules having Ag5 and I4; **b** temporal pattern of a rule (Ag5, Alc0, I4); **c** frequent itemset of the rule; and **d** accident record distribution of the rule

5.2.4 User feedback

After the subjects completed their analysis, we conducted an interview with each to collect their feedback on DART. Overall, the subjects were impressed by DART. One subject said that “the system helped me steer the analysis effectively. In particular, I appreciate the tools to support the exploration of temporal patterns of rules at different levels”. The other subject believed that the improved knowledge on the concepts related to dynamic association rules (e.g., frequent itemsets and various measures) by using the system would give him “more confidence in evaluating rules from algorithm and choosing appropriate rules for future event predictions”.

Regarding visualization tools and user interface, they liked the way that different views worked together to support in-depth analysis of rules. In particular, they thought that the tool for adjusting time periods in the data distribution view is valuable and innovative in this type of scenarios. They also mentioned that the system was intuitive to use.

The second subject made some suggestions for system improvement. One suggestion was to add tools to deal with spatial attributes, such as accident locations. He believed that by supporting both spatial and temporal factors, the system would be stronger.

6 Discussion

By focusing on temporal pattern analysis of dynamic rules at different time granularities, DART enhances the support for the visual analysis of association rule. In recent years, researchers have paid attentions to the visualization of association rules. Some data-mining tools such as Weka [17] and RapidMiner [18] integrate visualization into system, but most of them mainly focus on the visual display of static rules, and only provide limited interaction. In addition, some visual analysis tools have emerged to interpret the algorithms and results of association rules, such as AssocExplorer [15] and PatternDiscover [48]. Furthermore, all these tools aimed at static rule analysis. Our research on the visual analysis of dynamic association rules will fill in the gap and offer people a powerful approach to understanding association rules and making better decisions.

DART was designed to support the analysis of dynamic association rules, but the methods we proposed can be generalized to the analysis of temporal data, in particular in situations where analysis needs to be conducted with different temporal granularities and periods. The combination of the RadViz in the overview (Fig. 1e) and the table layout in the rule comparison view (Fig. 1f) offers a way to enhance the understanding of global patterns of periodic data, as well as to support data evaluation at local levels. Also, our idea of offering relevant raw data in the itemset view (Fig. 1g) and the data distribution view (Fig. 1h) can be informative to design efforts to connect the results of an opaque algorithm with the raw data used by the algorithm.

Our designs to support data analysis at multiple scale levels (e.g., hourly, daily, and monthly rule patterns) can be extended to the analysis of other types of data that have hierarchical structures, such as spatial data. Our design advocates multiscale approaches to support easy shift of the level of analysis from one level to another, as evidenced by the analysis of the relationship between alcohol factor and fatal accident in the second case study, and to use cross-scale data to improve data analysis, as seen in the use of lower-level data to support the direct manipulation of data-mining parameters at the current level (Fig. 3i, f).

The scalability of our approach is reasonable. Because our tool is a web-based system with sufficient system memory on the server side, space complexity is less a concern of us. Thus, here we discuss the time complexity only. The core component of the time complexity of our approach is the time complexity of the process to generate various association rules. This process is based on the Fp-Growth algorithm, which has a time complexity of $O(n^2)$ [56]. In the analysis of dynamic association rules, users choose a specific temporal granularity, and this choice divides the whole dataset, which has a total n data records, into k subsets. Here, k represents the number of data segments in analysis at a given granular-

ity. For example, when the granularity is at the level of hour, k is 24, while when the granularity is at the level of week, it is 7. Assume all data records are evenly distributed into all data segments. Then, we have k data segments, each of which contains n/k records. For each data segment, the time complexity is $O((n/k)^2)$, so the total complexity is around $k \cdot (O(n/k)^2)$ for rule mining, or $O(n^2/k)$. This time complexity can support real-time online data analysis for common datasets. Of course, for every large datasets, other methods, such as parallel algorithms could be considered to further reduce the time complexity of our approach.

Some limitations exist in our research. One of them is the lack of support for linking the dynamic rules at different levels. Although DART allows users to switch the analysis between different levels, currently users can only examine these patterns at different levels separately and cannot see how they may be related to each other. Easy connections among rules at different levels could be useful for better decision-making and sensemaking (e.g., explaining whether high accident rate in a day is related to accidents in rush hours or during lunch time.)

7 Conclusion

In this paper, we have presented DART, a visual analytics system of dynamic association rule mining. Our system offers a set of visualization and interaction designs to assist the control of data-mining processes, the examination of rules at different temporal levels, and the interpretation of the results from algorithm. Our two case studies involving domain experts analyzing relevant data with our system show that DART supports the analyses of dynamic rules, the acquisition of knowledge on algorithm, and the interpretation of data-mining results.

Our research can be extended in several ways. On the one hand, we will enhance the system by adding such tools as those to recommend appropriate temporal parameters for dynamic rule mining, to support the connection among rules at different levels, and to facilitate temporal-spatial rule analysis. On the other hand, we will have more comprehensive evaluation studies by making DART available to the public and collecting system usage data in the wild to deepen our understandings of how people use this kind of systems.

Funding This work was supported in part by the National Natural Science Foundation of China (No. 61572344).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Keim, D., Andrienko, G., Fekete, J.-D.: Visual analytics: definition, process, and challenges. In: Kerren, A., Stasko, J.T., Fekete, J.-D., North, C. (eds.) *Information Visualization*, pp. 154–175. Springer, Berlin (2008)
- Thomas, J.J., Cook, K.A.: A visual analytics agenda. *IEEE Comput. Graph. Appl.* **26**, 10–13 (2006)
- Liu, S., Cui, W., Wu, Y., Liu, M.: A survey on information visualization: recent advances and challenges. *Vis. Comput.* **30**, 1373–1393 (2014)
- Keim, D.A.: Information visualization and visual data mining. *IEEE Trans. Vis. Comput. Graph.* **8**, 1–8 (2002)
- Ferreira de Oliveira, M.C., Levkowitz, H.: From visual data exploration to visual data mining: a survey. *IEEE Trans. Vis. Comput. Graph.* **9**, 378–394 (2003)
- Bertini, E., Lalanne, D.: Surveying the complementary role of automatic data analysis and visualization in knowledge discovery. In: *Proceedings of the ACM SIGKDD Workshop on Visual Analytics and Knowledge Discovery: Integrating Automated Analysis with Interactive Exploration, VAKD '09*, (New York, NY, USA), pp. 12–20. ACM (2009)
- Holzinger, A., Jurisica, I.: *Interactive Knowledge Discovery and Data Mining in Biomedical Informatics: State-of-the-Art and Future Challenges*, vol. 8401. Springer, New York (2014)
- Endert, A., Ribarsky, W., Turkay, C., Wong, B.W., Nabney, I., Blanco, I.D., Rossi, F.: The state of the art in integrating machine learning into visual analytics. *Comput. Graph. Forum* **36**, 458–486 (2017)
- Mühlbacher, T., Piringner, H., Gratzl, S., Sedlmair, M., Streit, M.: Opening the black box: strategies for increased user involvement in existing algorithm implementations. *IEEE Trans. Vis. Comput. Graph.* **20**, 1643–1652 (2014)
- Agrawal, R., Imieliński, T., Swami, A.: Mining association rules between sets of items in large databases. *SIGMOD Rec.* **22**, 207–216 (1993)
- Han, J., Kamber, M., Pei, J.: 6-mining frequent patterns, associations, and correlations: basic concepts and methods. In: Han, J., Kamber, M., Pei, J. (eds.) *Data Mining. The Morgan Kaufmann Series in Data Management Systems*, 3rd edn, pp. 243–278. Morgan Kaufmann, Boston (2012)
- Rong, G., Liu, J.-F., Gu, H.-J.: Mining dynamic association rules in databases. *Control Theory Appl.* **24**(1), 127–131 (2007)
- Zhonglin, Z., Zongcheng, L., Chongyu, Q.: Tendency mining in dynamic association rules based on SVM classifier. *Open Mech. Eng. J.* **8**, 303–307 (2014)
- Uz Tansel, A., Imberman, S.P.: Discovery of association rules in temporal databases. In: *Fourth International Conference on Information Technology (ITNG'07)*, pp. 371–376 (2007)
- Liu, G., Suchitra, A., Zhang, H., Feng, M., Ng, S.-K., Wong, L.: AssocExplorer: an association rule visualization system for exploratory data analysis. In: *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '12*, (New York, NY, USA), pp. 1536–1539. ACM (2012)
- Geng, L., Hamilton, H.J.: Interestingness measures for data mining: a survey. *ACM Comput. Surv.* **38**, 9 (2006)
- Weka 3: data mining software in java. <https://www.cs.waikato.ac.nz/ml/weka/>
- Rapidminer. <https://rapidminer.com/>
- Han, J., Pei, J., Yin, Y., Mao, R.: Mining frequent patterns without candidate generation: a frequent-pattern tree approach. *Data Min. Knowl. Discov.* **8**, 53–87 (2004)
- Ogihara, Z., Zaki, M., Parthasarathy, S., Ogihara, M., Li, W.: New algorithms for fast discovery of association rules. In: *The 3rd International Conference on Knowledge Discovery and Data Mining* (1997)
- Pasquier, N., Bastide, Y., Taouil, R., Lakhal, L.: Discovering frequent closed itemsets for association rules. In: *Proceedings of the 7th International Conference on Database Theory, ICDT '99*, (London, UK, UK), pp. 398–416. Springer (1999)
- Brin, S., Motwani, R., Ullman, J.D., Tsur, S.: Dynamic itemset counting and implication rules for market basket data. *SIGMOD Rec.* **26**, 255–264 (1997)
- Djenouri, Y., Comuzzi, M.: Combining apriori heuristic and bio-inspired algorithms for solving the frequent itemsets mining problem. *Inf. Sci.* **420**, 1–15 (2017)
- Djenouri, Y., Chun-Wei Lin, J., Nørsvåg, K., Ramampiaro, H.: Highly efficient pattern mining based on transaction decomposition. In: *2019 IEEE 35th International Conference on Data Engineering (ICDE)*, pp. 1646–1649 (2019)
- Agrawal, R., Srikant, R.: Mining sequential patterns. In: *Proceedings of the Eleventh International Conference on Data Engineering*, pp. 3–14 (1995)
- Özden, B., Ramaswamy, S., Silberschatz, A.: Cyclic association rules. In: *Proceedings of the Fourteenth International Conference on Data Engineering, ICDE '98*, (Washington, DC, USA), pp. 412–421. IEEE Computer Society (1998)
- Nath, B., Bhattacharyya, D.K., Ghosh, A.: Incremental association rule mining: a survey. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **3**, 157–169 (2013)
- Bettini, C., Wang, X.S., Jajodia, S., Lin, J.: Discovering frequent event patterns with multiple granularities in time sequences. *IEEE Trans. Knowl. Data Eng.* **10**, 222–237 (1998)
- Bettini, C., Wang, X.S., Jajodia, S.: Testing complex temporal relationships involving multiple granularities and its application to data mining (extended abstract). In: *Proceedings of the Fifteenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, PODS '96*, (New York, NY, USA), pp. 68–78. ACM (1996)
- Liu, Y., Salvendy, G.: Visualization to facilitate association rules modelling: a review. *Ergon. IJE&HF* **27**(1), 11–23 (2005)
- Liu, Y., Salvendy, G.: Design and evaluation of visualization support to facilitate association rules modeling. *Int. J. Hum. Comput. Interact.* **21**(1), 15–38 (2006)
- Chen, W., Xie, C., Shang, P., Peng, Q.: Visual analysis of user-driven association rule mining. *J. Vis. Lang. Comput.* **42**, 76–85 (2017)
- Ong, K.-H., Ong, K.-L., Ng, W.-K., Lim, E.-P.: CrystalClear: active visualization of association rules. In: *International Workshop on Active Mining (AM-2002)* (2002)
- Appice, A., Buono, P.: Analyzing multi-level spatial association rules through a graph-based visualization. In: *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, pp. 448–458. Springer (2005)
- Sekhvat, Y.A., Hoerber, O.: Visualizing association rules using linked matrix, graph, and detail views. *Int. J. Intell. Sci.* **3**(01), 34 (2013)
- Yang, L.: Visualizing frequent itemsets, association rules, and sequential patterns in parallel coordinates. In: *Proceedings of the 2003 International Conference on Computational Science and Its Applications: Part I, ICCSA'03*, (Berlin, Heidelberg), pp. 21–30. Springer (2003)
- Yang, L.: Pruning and visualizing generalized association rules in parallel coordinates. *IEEE Trans. Knowl. Data Eng.* **17**, 60–70 (2005)
- Ankerst, M.: Visual data mining with pixel-oriented visualization techniques. In: *Proceedings of the ACM SIGKDD Workshop on Visual Data Mining* (2001)
- Chakravarthy, S., Zhang, H.: Visualization of association rules over relational DBMSs. In: *Proceedings of the 2003 ACM Symposium*

- on Applied Computing, SAC '03, (New York, NY, USA), pp. 922–926. ACM (2003)
40. Wong, P.C., Whitney, P., Thomas, J.: Visualizing association rules for text mining. In: Proceedings 1999 IEEE Symposium on Information Visualization (InfoVis'99), pp. 120–123 (1999)
 41. Bruzzese, D., Davino, C.: Visual post-analysis of association rules. *J. Vis. Lang. Comput.* **14**(6), 621–635 (2003)
 42. Galiano, F.B., Blanco, I.J., Sánchez, D., Vila, M.A.: Measuring the accuracy and interest of association rules: a new framework. *Intell. Data Anal.* **6**, 221–235 (2002)
 43. Shortliffe, E.H., Buchanan, B.G.: A model of inexact reasoning in medicine. *Math. Biosci.* **23**(3), 351–379 (1975)
 44. Liu, B., Hsu, W., Wang, K., Chen, S.: Visually aided exploration of interesting association rules. In: Proceedings of the Third Pacific-Asia Conference on Methodologies for Knowledge Discovery and Data Mining, PAKDD '99, (London, UK, UK), pp. 380–389. Springer (1999)
 45. Delgado, M., Ruiz, M.D., Sánchez, D.: Studying interest measures for association rules through a logical model. *Int. J. Uncertain. Fuzziness Knowl. Based Syst.* **18**, 87–106 (2010)
 46. Zhao, X., Wu, Y., Lee, D.L., Cui, W.: iforest: Interpreting random forests via visual analytics. *IEEE Trans. Vis. Comput. Graph.* **25**, 407–416 (2019)
 47. El-Assady, M., Sperrle, F., Deussen, O., Keim, D., Collins, C.: Visual analytics for topic model optimization based on user-steerable speculative execution. *IEEE Trans. Vis. Comput. Graph.* **25**, 374–384 (2019)
 48. Zhao, H., Zhang, H., Liu, Y., Zhang, Y., Zhang, X.L.: Pattern discovery: a progressive visual analytic design to support categorical data analysis. *J. Vis. Lang. Comput.* **43**, 42–49 (2017)
 49. Tableau desktop. <https://www.tableau.com/products/>
 50. Guo, P., Xiao, H., Wang, Z., Yuan, X.: Interactive local clustering operations for high dimensional data in parallel coordinates. In: 2010 IEEE Pacific Visualization Symposium (PacificVis), pp. 97–104 (2010)
 51. Sedlmair, M., Munzner, T., Tory, M.: Empirical guidance on scatterplot and dimension reduction technique choices. *IEEE Trans. Vis. Comput. Graph.* **19**, 2634–2643 (2013)
 52. Hoffman, P., Grinstein, G., Marx, K., Grosse, I., Stanley, E.: DNA visual and analytic data mining. In: Proceedings. Visualization '97 (Cat. No. 97CB36155), pp. 437–441 (1997)
 53. Wang, Y.C., Zhang, Q., Lin, F., Goh, C.K., Seah, H.S.: Polarviz: a discriminating visualization and visual analytics tool for high-dimensional data. *Vis. Comput.* **35**, 1567–1582 (2019)
 54. Online retail dataset. <http://archive.ics.uci.edu/ml/datasets/>
 55. Nhtsa fatality analysis reporting system. <http://www.nhtsa.gov/FARS>
 56. Kusters, W.A., Pijls, W., Popova, V.: Complexity analysis of depth first and fp-growth implementations of apriori. In: Proceedings of the 3rd International Conference on Machine Learning and Data Mining in Pattern Recognition, MLDM'03, (Berlin, Heidelberg), pp. 284–292. Springer (2003)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Huijun Zhang received the master's degree in computer science and technology from Taiyuan University of Technology, China, in 2010, where she is currently pursuing the Ph.D. degree with the College of Information and Computer. And she is also a Lecturer with the Department of Media Engineering, Communication University of Shanxi. Her research interests include visual analytics and data mining.



Junjie Chen received the Ph.D. degree from Beijing Institute of Technology in 2003. He is currently a Professor with college of Information and Computer, Taiyuan University of Technology. His main research interests include data mining, image processing, brain science, and intelligent computing.



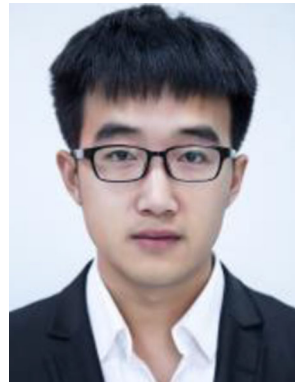
Yan Qiang received the Ph.D. degree from Taiyuan University of Technology in 2010. He is currently a Professor with college of Information and Computer. His main research interests include artificial intelligence, image processing, cloud computing technology, and database performance optimization.



Juanjuan Zhao received the Ph.D. degree from Taiyuan University of Technology in 2010. She is currently a Professor with college of Information and Computer. Her main research interests include artificial intelligence, image processing, cloud computing technology, and database performance optimization.



Jiangyang Xu received the master's degree in computer science and technology from Taiyuan University of Technology, China, in 2019. His research interests include big data analytics and visualization.



Yemin Yang received the bachelor's degree in computer science and technology from Tianjin Chengjian University, China, in 2017, where he is currently pursuing the master degree with the College of Information and Computer. His research interests include big data analytics and visualization.



Xiaobo Fan received the master's degree in computer science and technology from Taiyuan University of Technology, China, in 2019. His research interests include big data analytics and visualization.



Xiaolong Zhang is an Associate Professor at the College of Information Sciences and Technology at the Pennsylvania State University, University Park, PA. His research interests are in the area of human-computer Interaction (HCI) and focus on information visualization, visual analytics, multiscale interaction tools and systems, virtual environments, and social network analysis. His graduate degrees are from the University of Michigan and Tsinghua University, China, and his undergraduate degree is from Tsinghua University, China.