Contents lists available at ScienceDirect

Journal of Visual Languages and Computing

journal homepage: www.elsevier.com/locate/jvlc

View Points

Enabling decision trend analysis with interactive scatter plot matrices visualization

Wen Bo Wang[a], Mao Lin Huang[a,b], Quang Vinh Nguyen[d], Tony Huang[e], Kang Zhang[b,c], Tze-Haw Huang[a]

[a] School of Software, University of Technology, Sydney, Australia
[b] School of Computer Software, Tianjin University, Tianjin, China
[c] Department of Computer Science, University of Texas at Dallas, Richardson, TX, USA
[d] MARCS Institute and School of Computing, Engineering and Mathematics, University of Western Sydney, Australia
[e] Collaboration and Visual Analytics Lab, University of Tasmania, Australia

ARTICLE INFO

Article history:
Received 19 November 2015
Accepted 24 November 2015

Keywords:
Rough set theory
Dimensionality reduction
Scatter plot matrices
Parralell coordinate
Visual analytics
Visual decision making

ABSTRACT

This paper presents a new interactive scatter plot visualization for multi-dimensional data analysis. We apply Rough Set Theory (RST) to reduce the visual complexity through dimensionality reduction. We use an innovative point-to-region mouse click concept to enable direct interactions with scatter points that are theoretically impossible. To show the decision trend we use a virtual *Z* dimension to display a set of linear flows showing approximation of the decision trend. We conducted case studies to demonstrate the effectiveness and usefulness of our new technique for analyzing the property of three popular data sets including wine quality, wages and cars. The paper also includes a pilot usability study to evaluate parallel coordinate visualization with scatter plot matrices visualization with RST results.

1. Introduction

Multi-dimensional data exploration presents a great challenge to information visualization because features of data are inherently sparse in high dimensional data and the over-plotting of visual display makes it even difficult to observe any useful patterns. Visualization methods for large dimensional data are not usually effective due to density of high dimensions and the limitation of screen display. Interactive zooming could be used to provide an aid for exploring and reducing the number of dimensions such as [39]. However, the interaction is still limited where some of the contextual information could be lost during the navigation.

The efficiency of knowledge discovery tends to decline while the processing cost of information interpretation tends to increase because some are noisy data and not necessary all the dimensions need to be analyzed. This phenomenon is also known as *curse of dimensionality* which was first apparently coined by Bellman to describe the problem that data samples will grow exponentially according to the changes of the number of dimensions because of the necessity of fitting a multivariate function for a given degree of accuracy.

Dimensionality reduction is important in many application domains to be facilitated with classification, visualization of dealing with the complexity of multi-dimensional data. It reduces the intrinsic dimensionality of the data in order to cut down the cost of time and space complexities required for subsequent computation and analytic task. The terms *variable*, *feature* and *attribute* are commonly quoted

E-mail addresses: wenbo.wang@student.uts.edu.au (W.B. Wang), mao.huang@uts.edu.au (M.L. Huang), q.nguyen@uws.edu.au (Q.V. Nguyen), Tony.Huang@utas.edu.au (T. Huang), kzhang@utdallas.edu (K. Zhang), tze-haw.huang@student.uts.edu.au (T.-H. Huang).

http://dx.doi.org/10.1016/j.jvlc.2015.11.002
1045-926X/© 2015 Elsevier Ltd. All rights reserved.

in various research fields hence we would use them interchangeably throughout the paper.

Dimensionality reduction can be divided into feature selection and feature extraction. Feature selection is mainly to select a subset of the original variables according to selection principals. In the supervised method, the general criteria requires user to guide the selection process through choosing weighted quality metrics, therefore the selection rule would prefer the attributes weighted above the threshold. However in this case, user's expertize about quantization would have a great influence on the effectiveness of variable selection as quantization is typically not a trivial task. More importantly, empirical studies are the fundamental basis of applying quantization; hence the method may work well on this data set but might completely fail on another. On the other hand, feature extraction is a typically unsupervised technique with minimal consideration about user factors. The absence of user guidance raises the challenge of information interpretation if the result is unintuitive or not expected by the user that is often criticized as information loss. Most techniques developed in the past are projection based, implying that phenomena of interest higher than second order could not be discovered. Strictly speaking, projection means orthogonal. The oversimplified pattern is not adequate to support interactive data exploration that requires iterative interaction through visualization for the adjustment of input vectors to increase the accuracy of analytical results for decision trend analysis. Multi-dimensional data exploration via dimensionality reduction is really a user centric task in information visualization. Most dimensional reduction methods do not provide multiple results and make no assumption with the consideration of user's concern. Ideally, an effective method should only require the user to guide the procedures of dimensionality reduction, in terms of specifying a most concerned attribute and adjusting the values of input vectors subjectively.

In our previous works, we integrated Rough Set Theory (RST) with parallel coordinates [2] and scatter plot [34] for interactive feature selection. RST is a mathematical approach to data vagueness and uncertainty, which can be considered as discovering facts from complicated data through dimension reduction with a given dimension known as decision specified by the user. In this paper, we further extend our prior work with additional contributions described as follows:

- A feature ranking method on the result to guide the user for multi-dimensional data analysis.
- Interactive data exploration support in scatter plot matrices for class data.
- Enhanced scatter plot matrices for decision trend analysis.
- Provide more case studies to illustrate the visualization on different data sets.
- Carry out a pilot usability study on the visualizations.

2. Related works

There are several techniques for visualizing multi-dimensional data, such as Parallel Coordinate, Start Plots, Scatter plot Matrix, Mosaic Plots, Heat Map, Glyphs and Icons. Among them, Parallel Coordinate and Scatter plot Matrix are considerably popular techniques for large scale data sets. Theoretically, they are capable to visualize the data with unlimited number of dimensions nevertheless their visual efficiencies tend to decline when number of dimensions grows.

Some developments addressed the problem by visual transformation. Guo et al. and Artero et al. used clustering to highlight the patterns of homogenous data in parallel coordinate. Peng et al. applied dimension reordering to rearrange the dimension axes based on visual neighboring similarity for clutter reduction. However, using visual transformation to enhance the visual structure still left data in high dimensional space with sparse features. Nguyen et al. [35] presented a multi-dimensional data visualization system based on scatter plot with flexible axis and attribute mapping. The tool also provided interaction, filtering, zooming and dynamically control to the visualization. Although these techniques are quite effective to visualize small numbers of dimensions, dealing with high numbers of dimensions remains a challenge.

The widely accepted dimensionality reduction methods are Principal Component Analysis (PCA), Multi-dimensional Scaling (MDS) and Self-Organizing Map (SOM). PCA is a linear transformation method that projects the original data onto a much smaller set without ordinarily result. The selection principles – are typically interested in dimensions with largest eigenvalues, known as principal components because they explain the majority of variability. The low dimensional view that represents the high dimensional dataset is formed by rotating the principal components along the linear directions of maximum variability. MDS aims to place the data points that the pairwise distances are preserved as well as possible. SOM is an unsupervised learning algorithm based on neural network model by reducing the dimensions to low-dimensional (typically 2D) layer of neurons. Locally Linear Embedding (LLE) is another popular unsupervised learning technique that computes nearest neighborhood of each dimension to obtain the low dimensional embedding of high dimensional data. One common drawback of these methods is that they project the dataset into extremely low dimensions that could oversimplify patterns. For projecting an information correlated dataset i.e. survey dataset, into 2D space is usually meaningless for human centric knowledge discovery.

Projection Pursuit (PP) is a type of statistical technique for the pursuit the choices of possible projections in multi-dimensional data that can reveal the most details about the structure defined by a projection index. The pursuit of the possible projections globally involves non trivial computational intensive task. XGobi is a visualization system that integrated PP for viewing high dimensional data. The choices of possible relevance are the commonality between our work and PP. The main problem of PP is the difficulty to quantize the value of projection index because it is possible to present spurious interesting structures with an inappropriate projection index.

Several Visual Dimensionality Reduction (VDR) methods have been proposed by taking advantages of information visualization at different stages. Yang proposed Visual Hierarchical Dimensionality Reduction (VHDR) method by visually grouping dimensions into a hierarchy and constructing a new representation through the clusters of the hierarchy… VHDR has been integrated into XmdvTool since version 6.0. Yang further extended VHDR to propose a hierarchical Dimension Ordering, Spacing and Filtering Approach (DOSFA). DOFSA is similar to VHDR with additional improvement on visual structure via dimension ordering and spacing. Guo contributed a method that computed the entropy matrix and hierarchical clustering for low dimensional feature selection. Johansson applied several user-defined combinations of quality metrics such as similarity, outlier and clustering to measure the importance of attributes. The attributes are selected for these weights above the threshold defined by the user. By strict definition, they are feature selection techniques using some quality metrics as a measure to determine the feature subset selection.

3. Dimensionality reduction

3.1. Rough set theory

RST was first introduced by Pawlak to distinguish objects into sets under the given conditions necessary to make decisions specified by *decision attributes*. In general, it seems to be fundamental importance to many fields that require classification task such as feature selection, decision analysis, knowledge discovery and pattern recognition, etc.. In RST terms, a dataset is called a *decision table* which contains a finite set of data, namely universe, denoted as U. In the *decision table*, rows of a decision table are known as *decision rules*, which give conditions to make decisions, and let $A = \{a_1, a_2, a_3, \ldots a_n\}$ represent a superset of attributes. A is further classified into two disjoint subsets $A = (C, A \cup \{D\}), C \cap D = \varnothing$ where C and D denote the condition and decision respectively. RST is unable to deal with single objects because of the impossibility of discerning some objects by the existed information, so the objects need to be grouped into a set of equivalent classes by finding their indiscernibility relation expressed as follows:

$$E(P) = \left\{(x, y) \in U \colon \forall_a \in P \colon a_i(x) = a_i(y)\right\} \tag{1}$$

where $P \in A$ and x, y are the objects in the universe, $a_i(x)$ is the value of attribute a, for object x. Equivalence classes are further classified into approximation space where RST defines three regions of approximations namely lower approximation, upper approximation and boundary region: The first one is where the union of all original sets are included in every set, the second is where the union of all original sets have nonempty intersection with every set, and the third is that represents the difference between the upper and lower approximation. In our work, we only care about the lower approximation as it determines the quality of classification. Lower approximation and upper approximation are also called positive and negative regions respectively in RST terms.

3.2. Variable precision rough set

Classic RST was designed to deal with consistent dataset by its assumption of being not possible under a certain level of error on classification. For example, if $ab \rightarrow D$ then $cd \rightarrow D$ is considered inconsistent. This assumption of failure-free-decision-making is unrealistic in most real world datasets. To deal with inconsistency, Ziarko argued that probabilistic classification rules should be incorporated and hence proposed Variable Precision Rough Set (VPRS) model as an extension to RST. Beynon, provides the detailed VPRS concept, notations and case study. VPRS model allows the probability classification by introducing a given probability value β to deal with the restricted classification in original RST. It introduces the concept of major inclusion to tolerate the inconsistent dataset and the definition of majority is defined to lie between 0.5 and 1, which implies less than 50% of classification error.

The β position region in VPRS model is approximated as

$$POS_p^\beta(Y) = \cup_{Pr(Y|x_i) \geq \beta}\left\{x_i \in E(P)\right\} \tag{2}$$

where $Y \in U, Pr(Y|x_i) = \left|Y \cap x_i\right|_{\overline{x_i}}$ is a conditional probability function and $E(P)$ denotes a set of equivalent classes partitioned using (1). Clearly, a portion of objects with specified value β in the equivalent classes need to be classified into Y for it to be included in the β positive region. Given (2), we could find the *quality of classification* that measures the percentage of objects in conditional classes C has approximated into the position region in decision attributes D. It is used to extract β reduct and we will explain the definition of *reduct* later. The quality of classification in VPRS model is defined as follows:

$$\gamma^\beta(C, D) = \frac{\left|\cup_{Pr(E(D)|x_i) \geq \beta}\ \left\{x_i \in E(P)\right\}\right|}{|U|} \tag{3}$$

A subset of attributes that meets the classification requirement is called a *reduct* which is sufficient to describe the original attributes without loss of classification. In VPRS model, *reduct* is called β *reduct* or *approximate reducts* denoted as $RED^\beta(C, D)$ and according to Ziarko that a subset $P \subseteq C$ is a reduct of C with respect to D if and only if the following two criteria are satisfied

1. $\gamma^\beta(C, D) = \gamma^\beta\left(RED^\beta(C, D), D\right)$ and
2. No attributes can be eliminated from $RED^\beta(C, D)$ without affecting the requirement (1).

In the first requirement, Ziarko has defined the strict satisfaction of β reduct that some attributes can only be removed if and only if its qualification of classification γ^β for subset $P \subseteq C$ must be not affected by the γ^β for the whole set of conditional attributes C.

Our task of dimensionality reduction is relatively computational expensive by exhaustive approaches. Basically, we generate all the possible candidates from the conditional attributes and test them for satisfaction of β reduct

Please cite this article as: W.B. Wang, et al., Enabling decision trend analysis with interactive scatter plot matrices visualization, Journal of Visual Languages and Computing (2015), http://dx.doi.org/10.1016/j.jvlc.2015.11.002

criteria. Given n conditional attributes, we start from $k=2$ until $k=n$ so there are $\binom{k}{n}=\frac{n!}{k!(n-1)!}$ combinations of search space. There is a more efficient algorithm called *QuickReduct*, but its discussion is outside the scope here.

Algorithm 1 describes the feature selection procedures where G is a function used to satisfy the second requirement defined by Ziarko.

Algorithm 1. Dimensionality reduction algorithm based on VPRS model.

```
Input: A dataset $U$ with conditional $C$, decision $D$ and precision $\beta$.
Output: A set of reducts with respect to $D$.
$R \leftarrow \varnothing$
for $k \in K$ do
  if $\gamma^{\beta}(C,D)=\gamma^{\beta}(RED^{\beta}(k,D),D)$ then
    if $G(RED^{\beta}(k,D))$ then
      $R \leftarrow k$
    end if
  end if
end for
return $R$
```

3.3. Feature ranking

Typically we expect to find many *reducts* from the procedures described earlier and they have no discrepancy from RST perspective because they are all sufficient to represent the whole set of attributes without loss of classification quality. Unfortunately, it might be a legitimate concern from the user perspective as to which attribute (??) is the most useful to start with if there exist more than one. Feature ranking is commonly used in this situation that measures the correlation between classes based on ranking criteria. The correlation here refers to the linear relationship between two variables.

We applied Spearman rank correlation coefficient which is a non-parametric measure of statistical dependence between variables and ranks the order of data items instead of calculating the mean value. Thus, it is less susceptible to outlier or boundary items over other algorithms. Given a *reduct*, we first compute the ranking coefficient for each conditional attribute against the decision attribute as follows:

$$r=1-6\sum_{i=0}^{N} d_i^2/n(n^2-1) \tag{4}$$

where d_i denotes the difference between ranks for data items and r measures the degree of linear dependency. The overall ranking weight of a *reduct* can be easily calculated by $\sum_i^N r_i$.

3.4. K-means for data discretization

Recall that RST get the results in the form of classification derived from a set of objects. If the underlying numerical attributes are continuous then there will be too many weak equivalent classes generated, remember that a continuous data range can be theoretically unlimited.

Discretization is a process that transfers the attributes with continuous data into their discrete counterparts. It has received significant attention as a data pre-processing technique in many data mining systems i.e. ROSETTA. Equal Interval Width is the simplest discretization method but it is vulnerable to deal with uneven distribution of data. In our work, the K-means clustering is extended to discretization on attributes with continuous data before doing classification. It computes object similarities through distance function which generates the minimum value of the average inner-cluster separation, and so the uneven distribution of values can be well separated. Although, its main disadvantage is that the input parameter of k clusters must be known in advance as opposed to hierarchical clustering. However, specifying k is considered easier than defining a stopping rule for optimal clusters in hierarchical clustering.

4. Visualization

4.1. Point-to-region interaction in multi-dimensional visualization

Identifying class pattern and their correlations, such as linear relationships is a fundamental task in multidimensional data exploration. We use scatter plot matrices to visualize the result of dimensionality reduction. A scatter plot matrix shows all the pairwise scatter plots of attributes on a single view with multiple scatter plots in a matrix format, as shown in Fig. 1. The motivations behind this choice are (1) it is generally more intuitive to perceive data correlation in low projection view and (2) it is less susceptible to visual clutter created by over-plotting as opposed to parallel coordinate visualization. Interaction is an important function in our visualization which turns the static info-graphics into a dynamic display to uncover insights by involving users to manipulate the data transformation directly through the visual interface. In the interaction design, we allow the user to use the "*focus*+*context*" concept in interacting with scatter points directly. This interaction method can achieve noise reduction in class selection process. For example, when visualization detects a point that has been clicked, the entire convex hull of a corresponding class will be highlighted and the background of the convex hull will be grayed out as illustrated in Fig. 2(c). In other words, the system provides the interaction for individual data at the class level granularity, focusing on the subset with the area covered by a convex hull.

4.2. Linear approximation of decision trend

Decision attribute is the most distinct conception in RST comparing with other methods. It explicitly asks users to decide a preferred attribute from a given dataset so the attributes are reduced according to it. It would be useful to the user if the data exploration task is designed with a more decision oriented approach. Since a scatter plot can only reveal data correlation between two variables, we augment a parameter to approximate its relationship with a corresponding point in a third virtual dimension Z_0, that

Please cite this article as: W.B. Wang, et al., Enabling decision trend analysis with interactive scatter plot matrices visualization, Journal of Visual Languages and Computing (2015), http://dx.doi.org/10.1016/j.jvlc.2015.11.002

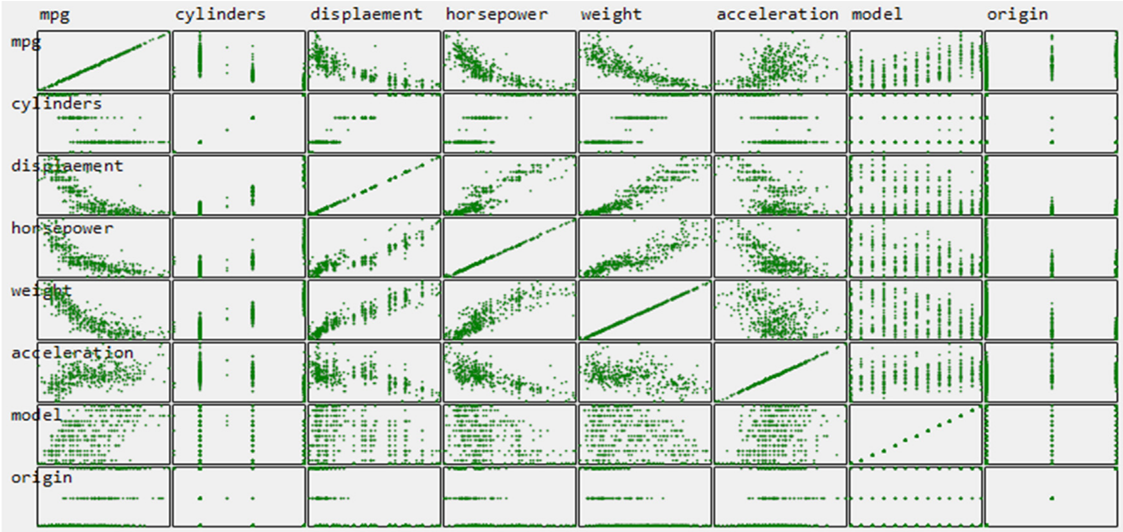

Fig. 1. An illustration of scatte plot metrics visualization.

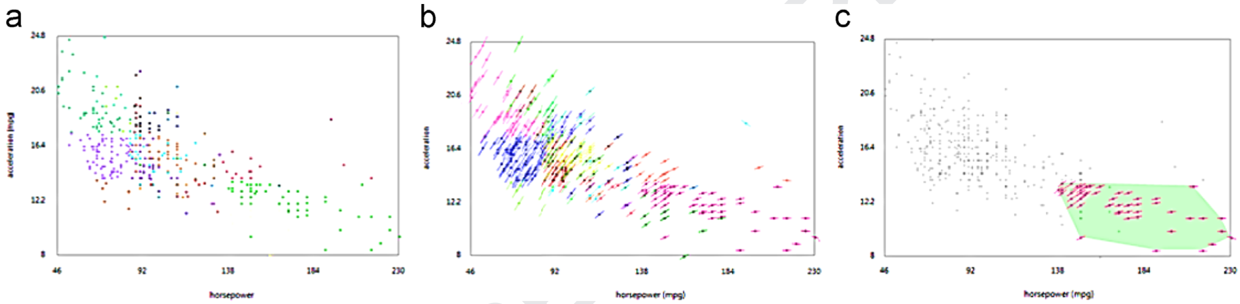

Fig. 2. (a) A classic scatter plots visualization. (b) Adding the decision flow where plots were augmented with respect to the decision variable. (c) Interaction (mouse click) by using point-to-region concept: that is, a point click causes an entire convex hull (a class) highlighted.

is, *decision attribute* in this case. We acknowledge that flow based scatter plot was previously discussed by Chan to study the sensitivity, but we further extend it to scatter plot matrices with interaction for class exploration by rough set model. A scatter point is positioned by its data value $(X_0, Y_{0,})$ with a line which represents the derivative of function *y* specifically, the slope indicates the positive or negative correlation with respect to (X_0, Z_0) or (Y_0, Z_0). In global linear approximation, all the points reveal the same trend when there is one slope. Chan computed local neighborhood of radius *w* to smooth the local trend around a given point. In our case, the equivalent class is already a set so we compute the local trend from the members in the class of a given point. Fig. 2(a) and (b) provides a visual comparison between the classic and flow based scatter plot representations. Clearly, it is simple yet powerful visual augmentation that helps the user to study the decision trend as opposed to classic metaphor which does not show the phenomena of interest as the decision trend.

To approximate the decision trend, we apply least squares in the linear regression model to best fit line of a given point (X_0, Y_0) with respect to the decision attribute. In linear regression model expressed by (5) and (6), there are two important coefficients namely b_1 and b_0 need to be solved first, where b_1 is the slope that measures the change in Y with respect to $X$and b_0 is the intercept. They are defined as follows:

$$b_1 = \frac{N\sum_i^N (X_i - X_0)(Y_i - Y_0) - \sum_i^N (X_i - X_0)\sum_i^N (Y_i - Y_0)}{N\sum_i^N (X_i - X_0)^2 - \left(\sum_i^N (X_i - X_0)\right)^2} \tag{5}$$

$$b_0 = \frac{\sum_i^N (Y_i - X_0) - b_1 \sum_i^N (X_i - X_0)}{N} \tag{6}$$

where $x_i \in E(P)$ and $X_0 \in E(P)$. Substituting b_0 and b_1 into the linear equation below to interpolate the best fitting line at point (X_0, Y_0)

$$Y_i(X_0 \pm k) = Y_0 + b_1(X_0 \pm k) + b_0 \tag{7}$$

where k is the desired length and please note that we have added the value of Y_0 because Y_i is a local linear approximation from a given point (X_0, Y_0).

In the interactive design for decision trend analysis, we enable the user to switch the view between (X_0, Z_0) and (Y_0, Z_0) by simply clicking on the coordinate label.

Please cite this article as: W.B. Wang, et al., Enabling decision trend analysis with interactive scatter plot matrices visualization, Journal of Visual Languages and Computing (2015), http://dx.doi.org/10.1016/j.jvlc.2015.11.002

4.3. Augmenting class converage

We mentioned earlier that the data covered by a convex hull belong to an equivalent class. It essentially represents a rule expressed as $E(P) \rightarrow D_i$ that has learned from approximating a set with respect to a decision class using (1)–(3). For example, the rule $E(P) = \{wight_{high}, accel._{low}\} \rightarrow 80\% cylinder_{high}$ means that there is eighty percent confidence. Cars having more *cylinders* should be with higher *weight* and lower *acceleration*. In fact, approximation regions are rule templates, a certain rule would classify the equivalent classes into positive regions, while uncertain or negative rules would classify the classes or negative regions. We are only interested in the rules that explain the phenomenon of interest. The two key elements associated with a rule are accuracy and coverage. Given a rule, its accuracy is defined as

$$accuracy\ (E(P) \rightarrow D_i) = \frac{|E(P) \cap D_i|}{|E(P)|} \tag{8}$$

where $E(P)$ and D_i denote the condition and decision class respectively. The accuracy measures the strength of a rule with respect to D_i. A weak rule has the accuracy less than β and is too weak to be meaningful. Similarly, the coverage of a rule can be measured by

$$coverage\ (E(P) \rightarrow D_i) = \frac{|E(P) \cap D_i|}{|D_i|} \tag{9}$$

The coverage measures the generality of a rule pointing to a certain class in D. In general, a rule with higher accuracy does not necessary to imply a lower coverage rule and vice versa.

In the visualization, we map the coverage to a hot–cold map with colors ranging from red to blue. For example, the background color for the area covered by the convex hull will be close to red for higher coverage.

5. Case Studies

We applied our technique on three popular datasets to demonstrate its effectiveness. The case studies are presented as follows.

5.1. Wine data

We used the wine dataset obtained from which consisted of 12 attributes with 4898 samples for modeling the wine quality based on physicochemical tests. The attributes cover the sufficient information to describe the characteristics of a wine such as *fixed acidity*; *volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulfates, alcohol* and *quality* (see Fig. 4 for the visualization of the entire data set using standard parallel coordinate and scatter plot matrix). Although the visualizations in Fig. 4 provide contextual information of entire data set, the inclusion of many dimensions makes the visualizations less readable due to

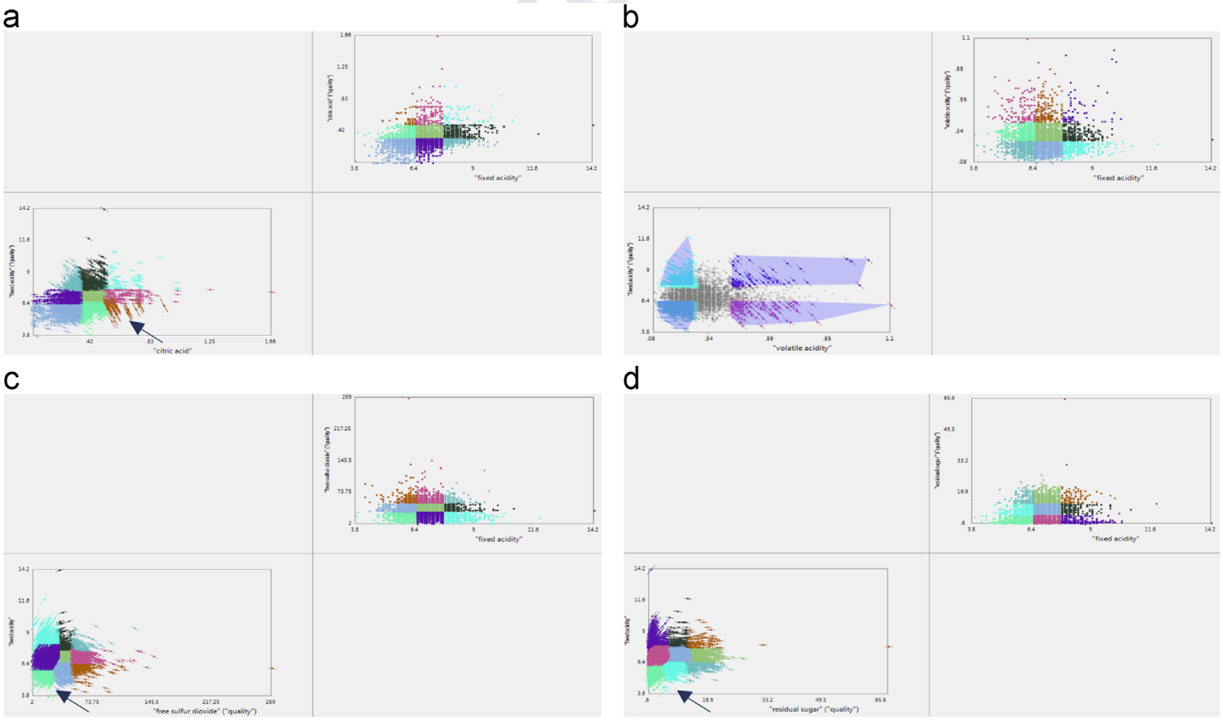

Fig. 3. Result obtained from the case study with wine data. The upper diagonal matrices displays the classic scatter plots and the lower diagonal matrices has been augmented with decision trend. (a) {citric acid, fixed acidity, quality} with (Y_0, Z_0). (b) {volatile acid, fixed acidity, quality} with (Y_0, Z_0). (c) {free sulfur dioxide, fixed acidity, quality} with (X_0, Z_0). (d) {residual sugur, fixed acidity, quality} with (X_0, Z_0).

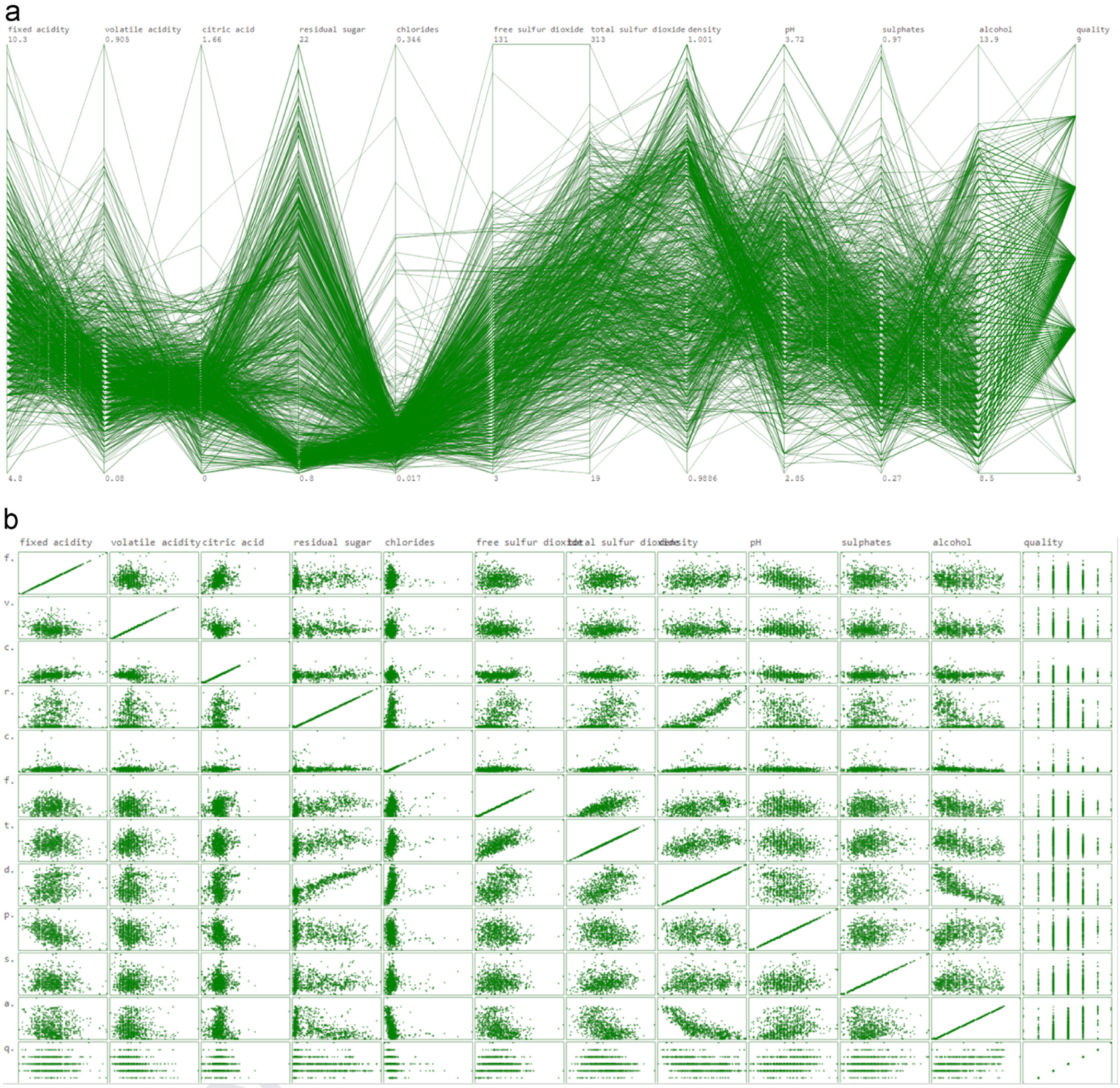

Fig. 4. Visualization of the entire wine dataset using (a) parallel coordinate and (b) scatter plot matrix.

the density at the parallel coordinate and the size reduction at the scatter plot matrix.

In our visualization of the dataset after using the rough set theory, the wine quality is the decision attribute (or dimension) and the rest become conditional attributes. The attributes are partitioned into three groups (or clusters) using K-means. There were five ranked feature sets obtained from VPRS procedures and each contains two condition attributes and one decision where we selected four of them as shown in Fig. 3. The points with the same color indicate that they belong to the same class. Some outlier classes are annotated with an arrow. In the visual data exploration of the scatter plot matrix, it displays that both fixed acidity and volatile acidity have negative impact on the wine quality revealed from the trends in Fig. 3(a) and (b). We further identified an outlier class which is with the worst impact to the wine quality in Fig. 3(a). It is also interesting to note that the lower free sulfur dioxide and residual sugar tend to have positive impact to the quality as displayed in Fig. 3(c) and (d).

Fig. 5 illustrates another example of the wine dataset where (1) the number of clusters was set to two, because without clustering, rough set will classify too many weak rules due to continuous variables, (2) Chose fixed acidity as the rough set decision attribute, and (3) the acceptable classification error rate of quality was set to 80%, which interpreted that it is allowable to have up to 20% incorrectness in the final clustering. After carrying out the RST process, there were 6 feature sets generated for classifying

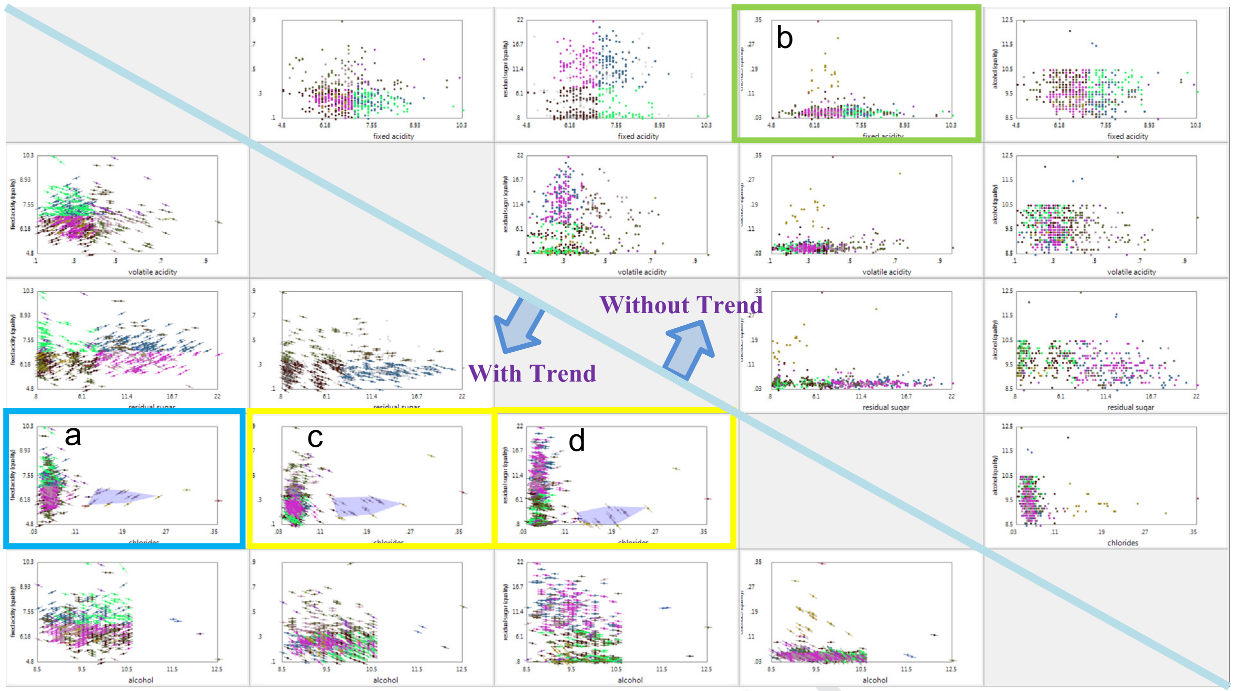

Fig. 5. Case study with wine dataset obtained from [31]. Boxes at “without-trend” area (area above the diagonal line) are scatter plots of each pair of attributes while boxes at “with-trend” area (area below the diagonal line) represent the same values and with changing trends.

data into classes, including: (a) {*fixed acidity, alcohol, quality*}, (b) {*fixed acidity, volatile acidity, residual sugar, chlorides, alcohol, quality*}, (c) {*fixed acidity, residual sugar, alcohol, quality*}, (d) {*residual sugar, PH, quality*}, (e) {*residual sugar, alcohol, quality*}, (f) {*fixed acidity, volatile acidity, alcohol, quality*}. The feature set (b) was used in our experiment after several trials in comparing the data quantity and visual quality of the classification results.

As seen at Fig. 5, *X*-axis and *Y*-axis give the same wine properties including {*fixed acidity, volatile acidity, residual sugar, chlorides, alcohol, quality*}. Boxes at “with-trend” area (area above the diagonal line) are scatter plots of each pair of attributes while boxes at “without-trend” area (area below the diagonal line) represent both the points' value and their changing trends. For example, it is easy to discover from box (a) and (b) that the wine data have been divided into 6 clusters. And in the box (a), the wine with higher *chlorides* and lower *fixed acidity* has positive influences on the *quality*, while the highest *chlorides* have invisible influence on wine. The visualization at boxes (a), (c) and (d) also indicated that, when the *chloride* is the same, *fixed acidity and volatile acidity and residual sugar* might have different impacts on the wine clusters. *Particularly*, both *volatile acidity* and *residual sugar* impacts the wine *quality* negatively, while *fixed acidity* impacts the wine *quality* positively.

5.2. Car data

In the next case study, we used a well-known car dataset obtained from http://lib.stat.cmu.edu/datasets/cars.data. The dataset contains 8 attributes with 392 samples after removed the missing attribute data. The dataset describes the car information about its *origin, model, acceleration, weight, horsepower, cylinder, mileage per gallon (mpg)* and *displacement*. The dimensionality reduction result is described in Fig. 6. This case study is used only for illustration purpose to show a feature set with more dimensions has been captured from the feature selection procedures.

Through the demonstration of the case study, we have shown the ease of use provided by the system for multi-dimensional data exploration, visual analysis and decision making.

5.3. Wage data

In the last case study, the wage dataset collected from http://www.nber.org/cps/ contains 534 observations on 11 variables sampled from the Current Population Survey of 1985. This data set includes attributes including *education, south, sex, experience, union, wage, age, race, occupation, sector and Marr* (*Marital Status*).

In this case study, *experience*, *wage* and *age* are the features in one rule of using *education* as the decision attribute. The visualization at Fig. 7 shows seven categories on the working *experience*.

More specifically, there are less people with higher wages, and they are either at a younger age (25–29) or an elder age (52–60). Although less young people earn a high wages, they have a positive influence on the classification with respect to *experience*, on the contrast; people who are elder with a higher wages tend to impact the classification negatively.

Please cite this article as: W.B. Wang, et al., Enabling decision trend analysis with interactive scatter plot matrices visualization, Journal of Visual Languages and Computing (2015), http://dx.doi.org/10.1016/j.jvlc.2015.11.002

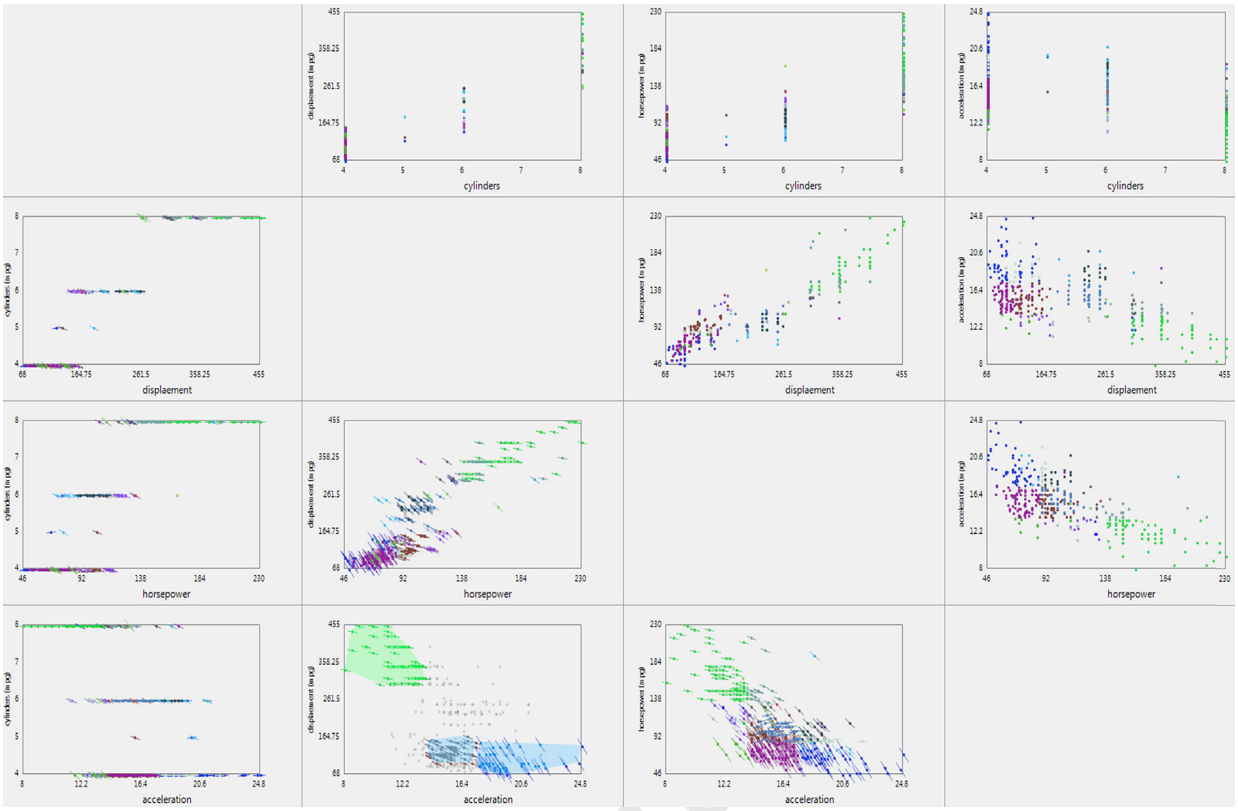

Fig. 6. Case study with car dataset obtained from http://lib.stat.cmu.edu/datasets/cars.data. We selected mileage per gallon (MPG) as the decision and the dataset has been reduced to 4 attributes namely acceleration, displacement, cyliners and horsepower.

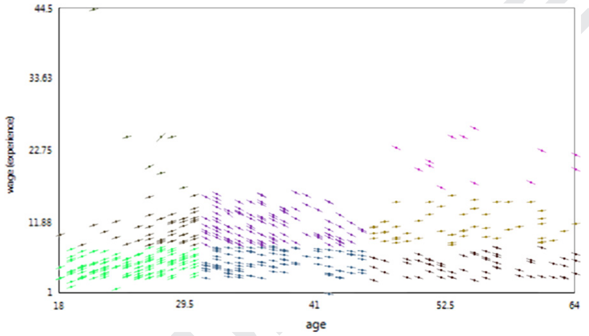

Fig. 7. Case study with Wages dataset obtained from http://www.nber.org/cps/. This figure show a box at "with-trend" area for correlation between *experience*, *wage* (*y*-axis) and *age* (*x*-axis).

6. Usability study

We conducted a piloted usability study with students in various backgrounds. The goal was to evaluate whether *Scatter Plot Matrix* is more effective than *Parallel Coordinate* when using RST results, in terms of accuracy and user preference.

6.1. Methods

1) *Participants:* recruited to the usability study were 16 participants (5 female, 11 male), ages ranged from 25 to 40 who were students with different backgrounds, including information technologies, sciences and business. Most of the participants indicated that they never used Parallel Coordinate and Scatter Plot Matrix before. None of them know about Rough Set Theory. All participants were fluent English speakers and accustomed with the methodology of understanding.
2) *Experimental design and tasks*: two similar datasets were used in the study. Each participant does two experiments on the two datasets using the two visualization techniques, Parallel Coordinate and Scatter Plot Matrix. The experiments were run on a 24 in. full HD screen. All the tasks in each trial took approximate 20 min to complete. For dataset 1, we set clusters to 2, and apply RST in the condition of using *mpg* as the decision attribute; the classification error rate will be set to 0.8 on *horsepower.* For dataset 2, we set clusters to 3, and apply RST in the condition of using *education as* the decision attribute; the classification error rate will be set to 0.8 on *experience.* We only collect the accuracy result in our study, each correct answer was marked as 1 and an incorrect answer was marked as 0.

Five questions were designed for the evaluation as follows, and listed on a multiple choice questionnaire.

Q1. After using Rough Set Theory, how many attributes we have on the visualization?

Q2. Which one is the decision attribute?

Please cite this article as: W.B. Wang, et al., Enabling decision trend analysis with interactive scatter plot matrices visualization, Journal of Visual Languages and Computing (2015), http://dx.doi.org/10.1016/j.jvlc.2015.11.002

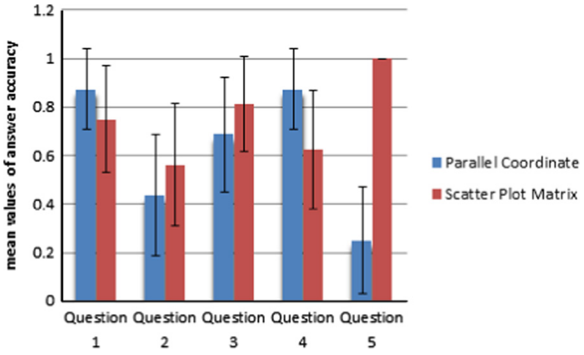

Fig. 8. Accuracy of parallel coordinate and scatter plot matrix visualizations corresponding to five questions (with 95% confidence intervals).

Q3. Whether an attribute is more influent than another attribute?

Q4. Whether selected attributes are more influent than the decision attribute?

Q5. Find the number of clusters in the visualization?

At the end, the participants were requested to rank each visualization technique on a 5-point Likert scale, from 1 (strongly disagree) to 5 (strongly agree).

6.2. *Results*

1) Accuracy-Fig. 8 shows the accuracy for the two visualizations corresponding to the five questions.

The early analysis of the results indicates that the average accuracy was significantly better in questions 2 and 3 for Scatter Plot Matrix than for Parallel Coordinate: Question 2 ($M=0.56$, $SD=0.26$ versus $M=0.44$, $SD=0.26$), Question 3 ($M=0.81$, $SD=0.16$ versus $M=0.69$, $SD=0.23$). The average accuracy was significantly lower in question 1 and 4 for Scatter Plot Matrix than for Parallel Coordinate: Question 1 ($M=0.75$, $SD=0.20$ versus $M=0.88$, $SD=0.12$), Question 4 ($M=0.63$, $SD=0.25$ versus $M=0.88$, $SD=0.12$). Noticeably, all the participants can identify the number clusters in Scatter Plot Matrix while they have difficulty in identifying them in the Parallel Coordinate (Question 5: $M=1.00$, $SD=0.00$ versus $M=0.25$, $SD=0.20$).

2) Subjective feedback about the Parallel Coordinate and Scatter Plot Visualizations: on the five-point Likert scale, the participants evaluated their overall preference using the two techniques on the RST datasets. The result indicated that that the participants prefer the Scatter Plot Matrix ($M=4.19$, $SD=0.16$) over the Parallel Coordinate ($M=2.13$, $SD=0.65$).

7. Conclusion and future work

Visual analysis is an important subject in multi-dimensional visualization, but it is often discussed in a standalone manner with many areas unexplored. Thus, it is often considered as a viewing step in multi-dimensional dataset. Dealing with high dimensional dataset is always challenging and we believe that the most effective way is through iterative visualization and interaction on the data subset. This is because that the iterative interaction process will involve human's eye-brain system into the data analytics and the eye-brain system would be considered to be the most efficient system for data analysis.

We contributed the novel scatter plot matrix visualization for multi-dimensional data and decision trend analysis. Our solution is a more comprehensive approach with a novel interaction model that is tightly integrated with the dimensionality reduction based on RST. We highlight the decision rule based concept offered by RST because it explicitly requires the user to establish a target interest in the visual analytic task. We illustrated the visualizations in three case studies with 3 popular datasets including wine quality, cars and wages. Our pilot usability study indicates the higher accuracy of Scatter Plot Matrix visualization over Parallel Coordinate visualization in determining the decision attribute, which attributes have more influence, and recognizing the clusters. The participants also preferred Scatter Plot Matrix than the Parallel Coordinate in the analysis task.

The current system implementation requires non-trivial computational time to search the solution space exhaustively. In the future work, we would like to integrate the evolutionary algorithm as suggested in with QuickReduct to improve the time complexity.

Uncited references Q4

[1,10–29,3,'30,'32,'33,'4,5–9].

References

[1] R.E. Bellman, Adaptive Control Processes: A Guided Tour, Princeton University Press, 1961.

[2] T.H. Huang, M.L. Huang, J.S. Jin, Parallel rough set: Dimensionality reduction and feature discovery of multi-dimensional data in visualization, Neural Inf. Process. (2011).

[3] A. Inselberg, The plane with parallel coordinate, Vis. Comput. 1 (2) (1985) 69–91.

[4] D.F. Andrews, Plots of high dimensional data, Biometrics 29 (1972) 125–136.

[5] P. Guo, H. Xiao, Z. Wang, X. Yuan, Interactive local clustering operations for high dimensional data in parallel coordinate, in: Proceedings of the IEEE Symposium on Pacific Visualisation, 2010.

[6] A.O. Artero, M.C.F. De Oliveira, H. Levkowits, Uncovering clusters in crowded parallel coordinates visualization, in: Proceedings of the IEEE Symposium on Information Visualisation, 2004.

[7] W. Peng, M.O. Ward, E.A. Rundersteiner, Clutter reduction in multi-dimensional data visualization using dimension reordering, in: Proceedings of the IEEE Symposium on Information Visualisation, 2004.

[8] K. Pearson, On llines and planes of closest fit to systems of points in space, Philos. Mag. 2 (6) (1901) 559–572.

[9] J.B. Jruskal, M. Wish, Multidimensional Scaling, Sage Publications, Beverly Hills, 1977.

[10] T. Kohonen, The self-organizing map, Neurocomputing 21 (1–3) (1998) 1–6.

[11] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by local linear embedding, Science 290 (5500) (2000) 2323–2326.

[12] J. Friedman, J. Tukey, A projection pursuit algorithm for exploratory data analysis, IEE Trans. Comput. 23 (9) (1974) 881–890.

[13] P.J. Huber, Projection pursuit, Annu. Stat. 13 (2) (1985) 435–475.

[14] D.F. Swayne, N. Hubbell, A. Buja, XGobi Meets S: Integrating software for data analysis, in: Computing Science and Statistics: Proceedings of the 23rd Symb Interface, 1991. Q5

[15] J. Yang, M. Ward, E. Rundensteiner, S. Huang, Visual hierarchical dimension reduction for exploration of high dimensional data, in:

Proceedings of Eurographics/IEEE TCVG Symposium on Visualisation, 2003.

[16] M. Ward, XmdvTool: integrating multiple methods for visualizing multivariate data, in: Proceedings of Visualisation, 1994.

[17] J. Yang, W. Peng, M.O. Ward, E.A. Rundensteiner, Interactive hierarchical dimension ordering, spacing and filtering for exploration of high dimensional dataset, in: Proceedings of IEEE Symposium on Information Visualisation, 2003.

[18] D. Guo, Coordinating computational and visual approaches for interactive feature selection and multivariate clusteriing, J. Inf. Vis. 2 (4) (2003) 232–246.

[19] S. Johansson, J. Johansson, Interactive dimensionality reduction through user-defined combinations of quality metrics, IEEE Trans. Vis. Comput. Graph. 15 (6) (2009) 993–1000.

[20] Z. Pawalk, Vagueness and uncertainity: a rough set perspective, Comput. Intell. 11 (2) (1995) 227–232.

[21] Z. Pawlak, Rough Set: Theoretical Aspects of Reasoning About Data, Kluwer Academic Publishing, 1991.

[22] W. Ziarko, Variable precision rough set model, J. Comput. Syst. Sci. 46 (1) (1993) 39–59.

[23] M. Beynon, Reducts within the variable precision rough set model: a further investigation, Eur. J. Oper. Res. 134 (3) (2001) 592–605.

[24] C. Spearman, The proof and measurement of association between two things, Am. J. Psychol. 15 (1904) 72–101.

[25] A. Ohrn, J. Komorowski, ROSETTA: A rough set toolkit for analysis of data, in: Proceedings of the JT Conference on International Science, Workshop Rough Sets and Soft Computing (RSSCS97), 1997.

[26] J.A. Hartigan, M.A. Wong, Algorithm AS 136: a K-means clustering algorithm, J. R. Stat. Soc. 28 (1) (1979) 100–108.

[27] Y.H. Chan, C.D. Correa, K.L. Ma, Flow-based scatterplots for sensitivity analysis, in: Proceedings of IEEE Symposium on VAST, 2010.

[28] S. Chatterjee, A.S. Hadi, Sensitivity Analysis in Linear Regression, Wiley, 1988.

[29] S. Tsumoto, Accuracy and coverage in rough set rule induction, in: Proceedings of RSCTS, LNAI, pp. 2475–2002.

[30] Y. Yao, Y. Zhao, Attribute reduction in decision-theoretic rough set momdels, Inf. Sci. 178 (1) (2008) 3356–3373.

[31] White Wine Dataset, [Online]. Available: 〈http://archive.ics.uci.edu/ml/machine-learning-databases/wine-quality/winequality-white.csv〉.

[32] A.T. Bjorvand, J. Komorowski, Practical application of genetic algorithm for efficient reduct computation, Wiss. Tech. Verl. 4 (1997) 601–606.

[33] Q.V. Nguyen, P. Alzamora, N. Ho, M.L. Huang, S. Simoff, D. Catchpoole, Unlocking the Complexity of genomic data of RMS patients through visual analytics, in: Proceedings of the International Conference on Computing Healthcare, IEEE, Hong Kong, 2012, pp. 134–139.

[34] T.H. Huang, M.L. Huang, K. Zhang, An interactive scatter plot metrics visualization for decision trend analysis, in: Proceeings of the 11th International Conference on Machine Learning and Applications, 2012, pp. 258–264.

[35] Q.V. Nguyen, Y. Qian, M.L. Huang, J. Zhang, TabuVis: a tool for visual analytics multidimensional datasets, Sci. China Inf. Sci. 052105 (12) .

Please cite this article as: W.B. Wang, et al., Enabling decision trend analysis with interactive scatter plot matrices visualization, Journal of Visual Languages and Computing (2015), http://dx.doi.org/10.1016/j.jvlc.2015.11.002