Multi-dimensional Data Visualization Techniques for Exploring Financial Performance Data.

Conference Paper · January 2007

Source: DBLP

Dorina Rajanen (Marghescu)
University of Oulu

MULTIDIMENSIONAL DATA VISUALIZATION TECHNIQUES FOR EXPLORING FINANCIAL PERFORMANCE DATA

Dorina Marghescu
Turku Centre for Computer Science, Finland
Åbo Akademi University, Dept. of IT, Finland
dorina.marghescu@abo.fi

Abstract

In this paper, we review nine visualization techniques that can be used for visual exploration of multidimensional financial data. We illustrate the use of these techniques by studying the financial performance of companies from the pulp and paper industry. We also illustrate the use of visualization techniques for detecting multivariate outliers, and other patterns in financial performance data in the form of clusters, relationships, and trends. We provide a subjective comparison between different visualization techniques as to their capabilities for providing insight into financial performance data. The strengths of each technique and the potential benefits of using multiple visualization techniques for gaining insight into financial performance data are highlighted.

Keywords: multidimensional data visualization techniques; financial performance; financial data visualization

Introduction

Many novel visualization techniques have been developed in the fields of information visualization (Card et al. 1999) and visual data mining (Keim 2002). However, the research literature concerning the use of visual data mining for gaining insight into *financial data* is relatively sparse, despite the fact that this technological approach is suitable for both financial data and business users. Financial data are very complex due to their high dimensionality, large volume and diversity of data types. Business users are demanding straightforward visualizations and task-relevant outputs, due to the time and performance constraints under which they work (Kohavi et al. 2002).

In this paper, we review nine visualization techniques that are suitable for representing multidimensional data. The aim is to examine the extent to which they are capable of providing insight into *financial performance data*. In particular, we focus on the problem of financial benchmarking, which is concerned with comparing the financial performance of companies.

The approach consists of the following steps. First, we formulate the financial benchmarking problem in terms of business questions and associated data mining tasks. Second, we investigate the capabilities of each visualization technique in solving the derived data mining tasks and uncovering interesting patterns in data. Third, we compare the visualization techniques from three different perspectives such as: 1) the capability of the techniques to uncover interesting patterns in the data (task fitness); 2) the capability to visualize data items or data models; and 3) the type of data processed (i.e., original data or normalized data).

The analysis highlights the strengths of each technique and the potential benefits of using multiple techniques for exploring financial data. In this paper, we do not address the interactive capabilities of the visualization techniques.

The paper is organised as follows. In the next section, we outline the problem of financial benchmarking, describe the dataset to which we applied the visualization techniques, and derive the business questions and data mining tasks. In Section three, we describe nine multidimensional data visualization techniques and highlight their capabilities for solving the derived data mining tasks. Section four provides a subjective comparison of the techniques and discusses the results. We conclude with final remarks and future work ideas.

The problem of financial benchmarking

One of the problems that business intelligence people are confronted with nowadays is performing comparisons of companies' financial performance. This problem of comparing financial performance of companies is known as *financial competitor benchmarking* (Eklund 2004). The problem is non-trivial since many variables (financial ratios) must be considered. One part of the problem is choosing the ratios to be used when describing the financial performance of a company. Eklund (2004) proposed a model for financial competitor benchmarking in the pulp and paper industry, with seven financial ratios as a basis for companies' performance comparison, and the Self-Organizing Map (SOM) as the method for data analysis. In this paper, we build on the mentioned research to explore the use of other visualization techniques for gaining insight into financial data.

Illustrative Dataset

The dataset analysed in this paper is a subset of a dataset whose collection process including variable and company selection are described by Eklund (2004). The data values are entirely based on the information obtained from companies' financial reports available on the Internet.

The data refer to 80 companies that function in the pulp and paper industry worldwide, observed during 1997 and 1998. A total of 160 observations are analysed. The dataset contains seven numerical variables, namely seven ratios that characterize the financial performance of companies in the pulp and paper industry. The ratios are grouped in four categories: *profitability* (Operating Margin, Return on Equity, and Return on Total Assets), *solvency* (Interest Coverage, Equity to Capital), *liquidity* (Quick Ratio), and *efficiency* (Receivables Turnover). In the following, we use acronyms when referring to any of the financial ratios (that is, OM, ROE, ROTA, IC, EC, QR, and RT respectively). The dataset contains three categorical variables: companies' name, region (Europe, Northern Europe, USA, Canada and Japan), and year (1997 or 1998). The choice of this particular dataset was due to the availability of the dataset, and to its suitability for data mining (e.g., cluster detection, cluster characterization, class characterization, outlier detection, and dependency analysis).

Business Questions and Data Mining Tasks

According to Soukup and Davidson (2002), in order to use information visualization for solving a business problem, the problem should be translated in terms of business questions and further into visualization or data mining tasks. For the problem of financial benchmarking we have derived the business questions and data mining tasks as follows:

a) Outlier detection: Do the data contain outliers or anomalies? Are there any companies that show unusual values of financial ratios?
b) Dependency analysis: Are there any relationships between variables?
c) Data clustering: Are there clusters (groups of companies with similar financial performance) in the data? How many clusters exist?
d) Cluster description: What are the characteristics of each cluster?
e) Class description: Are there any relationships (common features) among companies located in one region or another? What are these common features?
f) Comparison of data items: Compare two or more companies with respect to their financial performance.

For the task f), we have chosen three companies to be compared according to their financial performance in 1998: Reno de Medici, Buckeye Technologies, and Donohue. For Reno de Medici we look also at its evolution from 1997 to 1998. These companies are identified on the graphs using the letters A, B, C, and D, respectively. Table 1 presents the financial ratios of these companies.

Table 1. Financial ratios of the companies chosen for comparison

Company	*Id.*	*Year*	*Region*	*OM*	*ROE*	*ROTA*	*EC*	*QR*	*IC*	*RT*
Reno de Medici 1997	A	1997	Europe	4.02	-15.38	0.64	27.94	1.29	0.15	3.3
Reno de Medici 1998	B	1998	Europe	6.7	5.34	5.27	28.19	1.03	1.68	2.63
Buckeye technologies 1998	C	1998	USA	19.42	38.96	16.21	20.91	1.36	3.28	7.79
Donohue 1998	D	1998	Canada	21.24	17.96	15.92	46.35	0.91	5.15	7.96

Multidimensional data visualization techniques

Because our dataset is tabular data, that is, the rows represent records and the columns represent attributes or dimensions of data, and the data has more than two dimensions, we selected *multidimensional data visualizations* for analysis (Hoffman and Grinstein 2002). The multidimensional data visualization techniques that are reviewed in our paper are multiple line graphs, permutation matrix, survey plot, scatter plot matrix, parallel coordinates, treemap, Principal Components Analysis, Sammon's mapping, and the Self-Organizing Maps. In the following, we apply these visualization techniques on the financial performance data and highlight their capabilities for answering the business questions and data mining tasks formulated in the previous section. Due to page limitations, we are only discussing two to three ratios for each technique. A complete discussion can be found in Marghescu (2007).

Multiple line graphs

Line graphs are used for one dimensional data. On the horizontal axis (Ox) the values are not repeated (e.g., time or the ordering of the table). The vertical axis (Oy) shows the values of the variable of interest. Multiple line graphs can be used to show more than two variables or dimensions (x, y1, y2, y3, etc.).

Figure 1 shows line graphs for two ratios (OM and ROE), observed in 1997 and 1998. The companies are mapped to the horizontal axis, in the order of appearance in the data table. The graph presents companies from different regions (Europe, Northern Europe, USA, Canada and Japan) in different colours, facilitating the characterization of companies from one region or another. By positioning the two years of data one under the other, one can follow the evolution of some company's financial ratios, and make comparisons between companies' financial states.

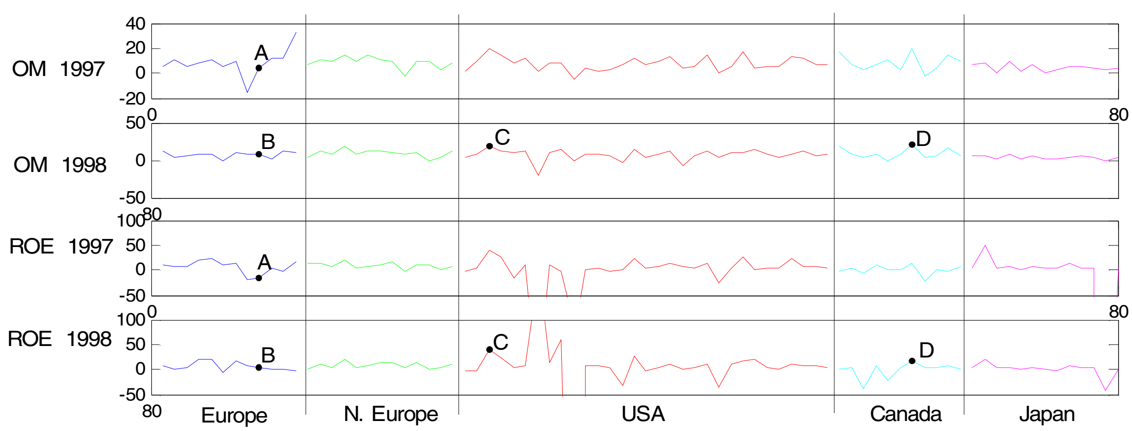

Figure 1 Multiple line graphs

This graph also facilitates the detection of outliers or anomalies in the data, for example, the very low and very high values of ROE for three of the companies, which were further removed from the dataset. By highlighting the companies to be compared, one can see the differences and similarities among them.

Permutation matrix

The permutation matrix is a special type of bar graph described by Bertin (1983). In a permutation matrix, each data dimension is represented by a bar graph in which the heights of the bars represent the data values. The horizontal axes of all bar graphs have the same information (e.g., the time or ordering of the data table). The below average data values are coloured black, and the above average data values are coloured white. A green dashed line plotted over the data represents the average value of each dimension. Implementations of permutation matrixes allow the interactive changing of the order of the records for observing interesting patterns.

Figure 2 displays a permutation matrix created with Visulab (Hinterberger and Schmid 1993). On the horizontal axes the companies are arranged in descending order of ROTA. The companies of interest are highlighted. This graph facilitates the detection of relationships between ratios and the comparison of companies. It also reveals anomalies in the data.

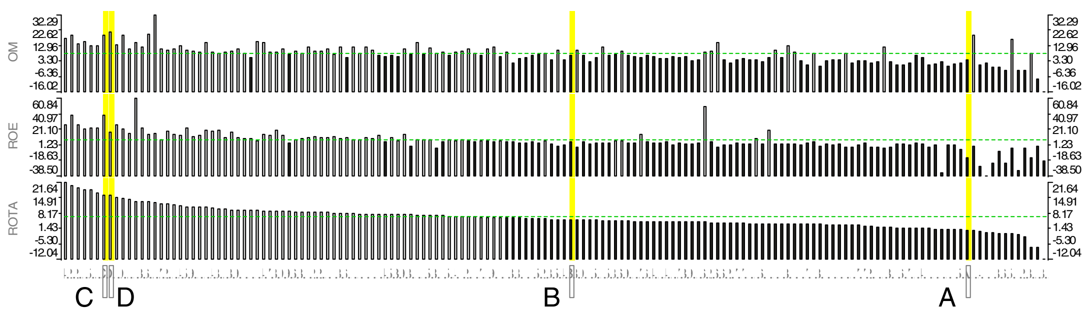

Figure 2 Permutation matrix created with Visulab

Survey plot

The survey plot is a variation of the permutation matrix. The values of each data dimension are represented as horizontal bars. The width of the bars is proportional to the data values. The bars are centred and there are no spaces separating the bars. One can use colours to distinguish between different classes in the data (if a class variable is present).

Figure 3 displays a survey plot, in which the data are sorted according to ROTA. This facilitates the detection of relationships between ROTA and other ratios, for example OM, ROE and IC.

Companies from different regions are displayed with different colours. The graph shows that the Japanese companies are not among the most profitable ones, while the American and European companies display the highest profitability. The technique facilitates the detection of outliers and comparison between two or more companies.

Scatter-plot matrix

A scatter plot is used to plot two dimensional data so that the horizontal axis shows the values of one variable and the vertical axis shows the values of another variable. The scatter-plot matrix is useful for looking at all possible pairs of variables in the dataset.

Figure 4 displays a scatter plot matrix for three financial ratios (OM, ROE and RT). The plots reveal relationships between ROE and OM. The visualization also reveals outliers, and facilitates the comparison of companies.

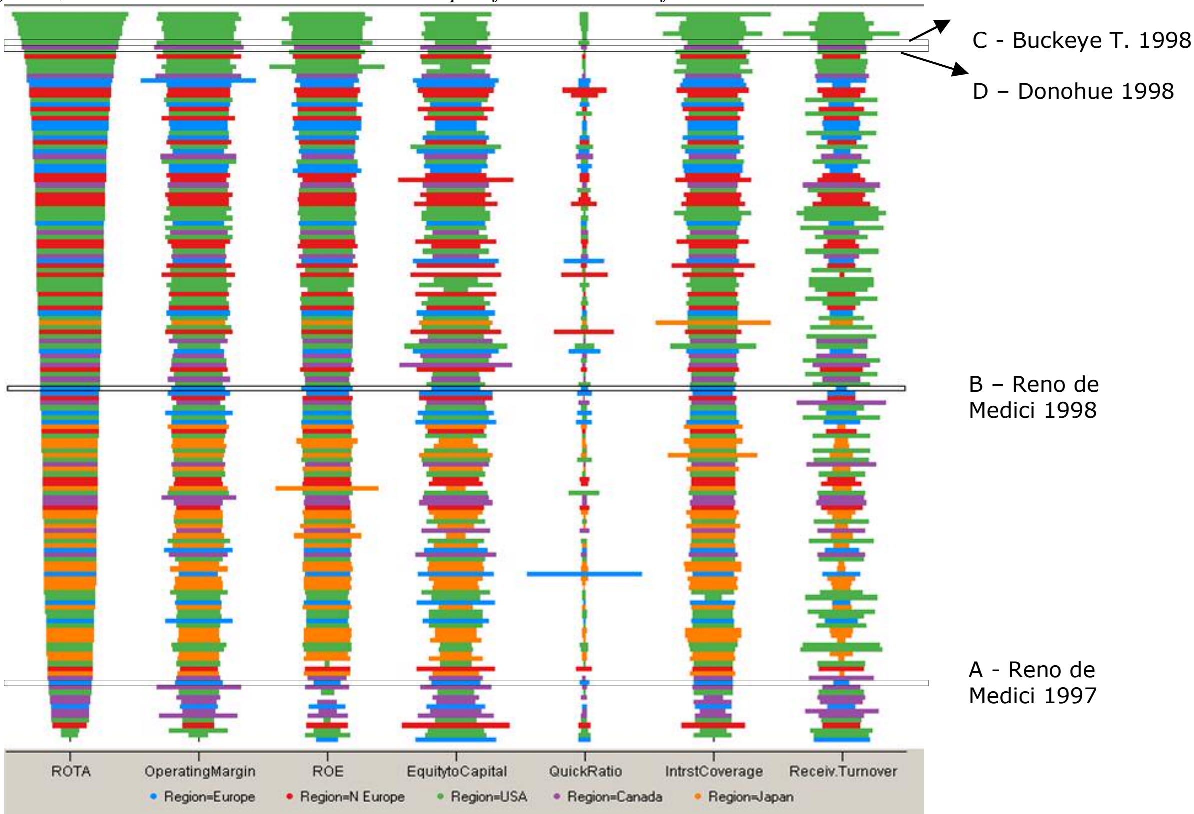

Figure 3 Survey plot created with Orange (Demsar 2004)

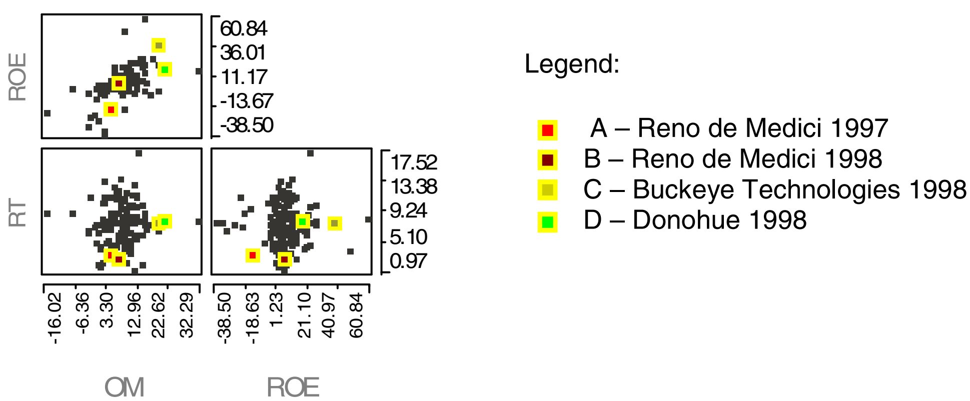

Figure 4 Scatter-plot matrix created with Visulab

Parallel coordinates

Introduced by Inselberg (1985), parallel coordinates represent multidimensional data using lines. The data dimensions are represented as parallel axes (coordinates). The maximum and minimum values of each dimension are scaled to the upper and lower points on a vertical axis. An n-dimensional data point is displayed as a *polyline* that crosses each axis at a position proportional to its value for that dimension.

Figure 5 represents the financial ratios as parallel axes and each company as a polyline that crosses each axis at a point proportional to the value of the ratio for the corresponding company. The companies of interest are highlighted using different colours.

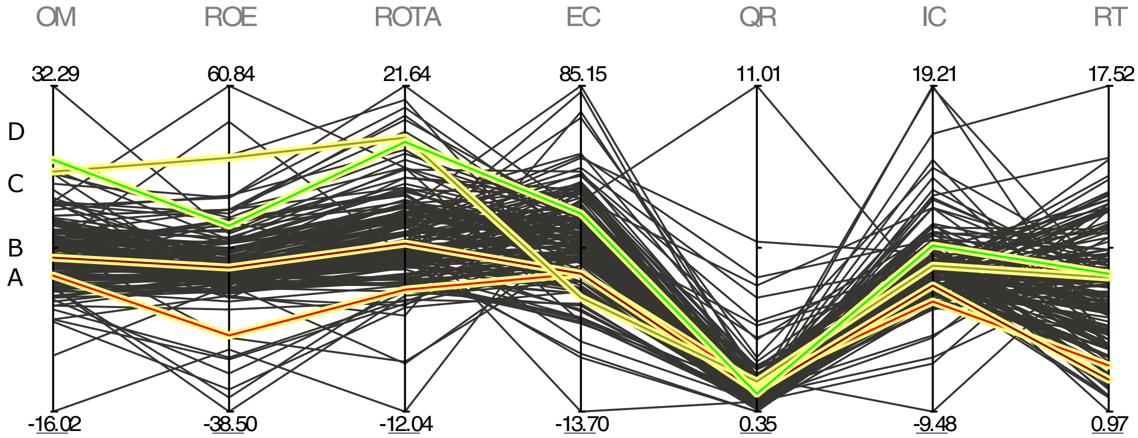

Figure 5 Parallel coordinates created with Visulab

The display facilitates detection and characterization of outliers. One can compare the financial performance of different companies. The relationships between two or more variables can be detected if the correlated variables are arranged consecutively (for example, ROE and ROTA).

Treemaps

The treemaps (Johnson and Shneiderman 1991) are hierarchical visualizations of multidimensional data. Data dimensions are mapped to the size, position, colour, and label of nested rectangles.

Figure 6 displays the dataset using the treemaps technique. The figure was created with Treemap 4.1 (2004). Each company is represented by a rectangle. The size of the rectangle indicates the value of RT. The colour of the rectangle indicates the value of the ROTA ratio as follows: light green indicates high values of ROTA; light red indicates small values of ROTA; dark red and dark green shows values of ROTA close to 14 (see the "colour binning" panel in the visualization below). In this visualization, the dataset is organised into categories such as year and region.

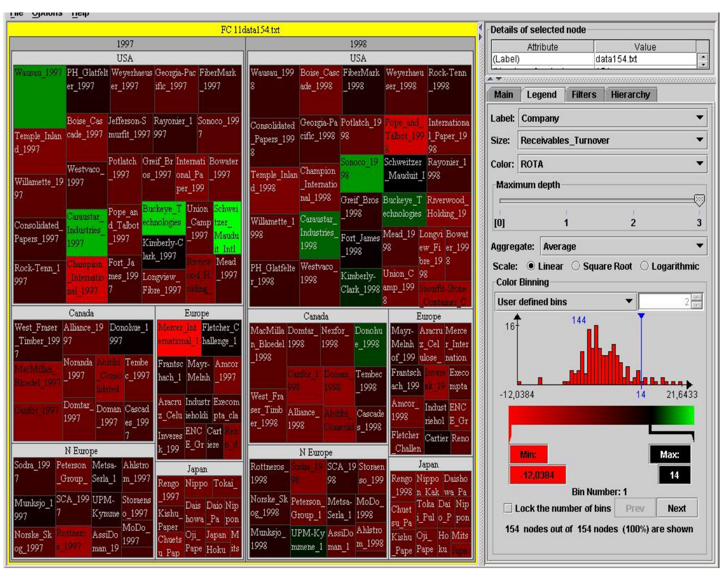

Figure 6 Treemap created with Treemap 4.1

This treemap representation shows where the most profitable companies in terms of ROTA are located, and how the companies of interest have evolved over time. In addition, one can identify common features or patterns in the industry, for example, that Japanese companies have the lowest values of the efficiency ratio. One can also compare the financial performance of different companies.

Principal component analysis (PCA)

PCA is a dimensionality-reducing technique employing *linear transformation of data* (Sharma 1995). The projection of high-dimensional data onto a lower-dimensional space tries to preserve the variance of the original data as well as possible. The PCA technique creates new variables (called principal components), which are linear composites of the original variables and are uncorrelated amongst themselves. The maximum number of new variables that can be formed is equal to the number of original variables. The PCA output is judged in terms of how well the new variables represent the information contained in data, or, geometrically, how well the new dimensions can capture the original configuration of the data.

Figure 7 shows PCA plot that was constructed from the standardized dataset. The red dot shows the observation closest to the centre of the dataset. The companies of interest are marked with a yellow star and labelled on the graph.

One can interpret the principal components by inspecting the loadings of each original variable to the PCs. The higher the loading of a variable, the more influence it has in forming the PC score and vice versa. In our case, the first PC (horizontal axis) is highly correlated with the profitability ratios and the IC ratio. Therefore, companies placed towards the right of the horizontal axis, have high values in profitability and IC. The second PC (vertical axis) is highly correlated with QR and EC. Companies located on the upper part of the graph have a high liquidity and high solvency. The amount of variation explained by the two PCs is 40.926% + 19.455%= 60.38% of the total variance. While this amount of variance accounts for the variation of six of the ratios, it does not consider the variation of efficiency (RT) among the companies.

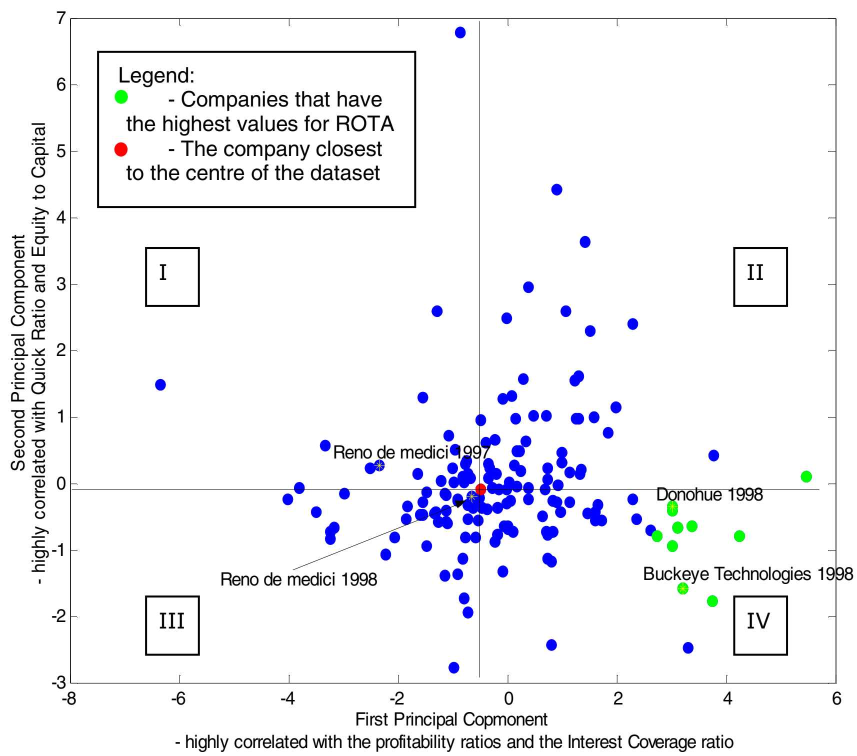

Figure 7 Data projected on the first two PCs created with the Statistics Toolbox (The MathWorks 2002). In area I: medium-high liquidity, low-medium profitability; II: medium-high liquidity, solvency and profitability; III: low-medium liquidity, solvency and profitability; IV: low-medium liquidity, medium-high profitability

In addition to its usefulness as a data reduction method, PCA is also useful in finding numerous patterns in data (Figure 7). The graph shows the high profitability of Buckeye Technologies 1998 and Donohue 1998, and the increase in profitability for Reno and Medici in 1998. The high correlation of the first PC with all profitability ratios and with the IC ratio indicates that there also exists a relationship between the profitability ratios and IC. Similarly, the high correlation of the second PC with EC and QR indicates that EC and QR are also correlated.

By splitting the visual representation in four areas by two orthogonal lines that intersect in the centre of the dataset, one can divide the dataset into four groups of similar observations as shown and described in Figure 7. Based on the meaning of the first two PCs, one can conclude that in area I there are companies with medium-high liquidity and low-medium profitability; in area II, companies with medium-high liquidity, solvency and profitability; in area III, companies with low-medium liquidity, solvency and profitability; and in area IV, companies with low-medium liquidity but medium-high profitability. Based on this evaluation, one can compare the financial performance of the companies of interest.

Sammon's mapping

Sammon's mapping is a *nonlinear projection* of the multidimensional data down to two dimensions so that the distances between data points are preserved (Kohonen 2001). It belongs to multidimensional scaling techniques.

Figure 8 illustrates our financial dataset using Sammon's mapping. The data values were normalized using the discrete histogram equalization method. The normalization method works in two steps: first, the data values of each variable are replaced by the order index, and then these values are normalized to be in the range [0, 1], by applying a linear transformation. Companies from different regions are displayed using different colours. The companies of interest are marked with yellow stars and labelled on the graph.

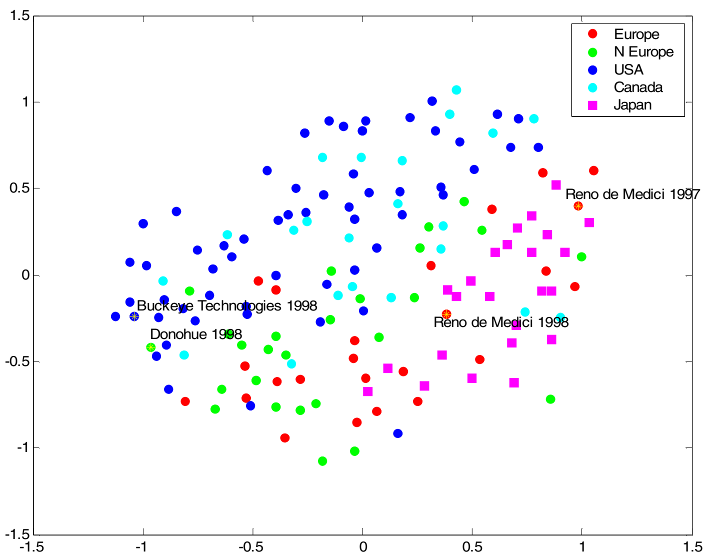

Figure 8 Sammon's mapping created with SOM Toolbox 2.0 (2005)

The technique is useful in visualizing class distributions, especially the degree of their overlap. One can see that companies from Canada and USA overlap and map to the same area of the graph, whereas Japan, Europe and Northern Europe form three separate groups. However, the degree of overlapping between all these classes is quite high; especially Europe and Northern Europe do not separate well from the other groups. The differences and similarities between the companies are easy to distinguish, but not easy to interpret.

Self-Organizing Maps (SOM)

The SOM technique, developed by Kohonen (2001) is a special type of neural network based on unsupervised learning. The SOM algorithm is similar to the K-Means clustering algorithm, but the output of a SOM is topological and neighbouring clusters are similar. As a projection technique of multidimensional data onto a two-dimensional grid, the SOM method is similar to multidimensional scaling techniques, such as Sammon's mapping. The grid consists of units that have assigned reference vectors with the same dimensionality as the original data. After learning is complete, the reference vectors are updated such that they resemble most of the data items, as much as possible. Each data item is then mapped to the unit where the highest similarity between the reference vector and the data item is calculated. Multiple data items mapped onto the same unit are similar and form a cluster.

We have used the SOM technique on normalized data obtained byapplying the d iscrete histogram equalization method. There are many ways to represent the SOM output. One way to represent the data is to use the *scatter plot* technique (usually with jittering), in which the horizontal and vertical axes are produced by the Kohonen network (i.e., the map size, in our case 6x5 units).

Figure 9 is a scatter plot of the dataset based on the SOM coordinates. Companies from different regions are highlighted using different colours. The technique of jittering was used in order to change with a small value the position of each company; otherwise the companies mapped to the same unit would have overlapped. Figure 9 shows many clusters in the data (if more companies are mapped to the same map unit, they may be interpreted as forming a cluster). One can also observe some isolated companies. However, the interpretability of this map is not easy. One can distinguish among the companies belonging to the same region, or identify the placement of these companies on the map but cannot interpret these classes or the clusters formed.

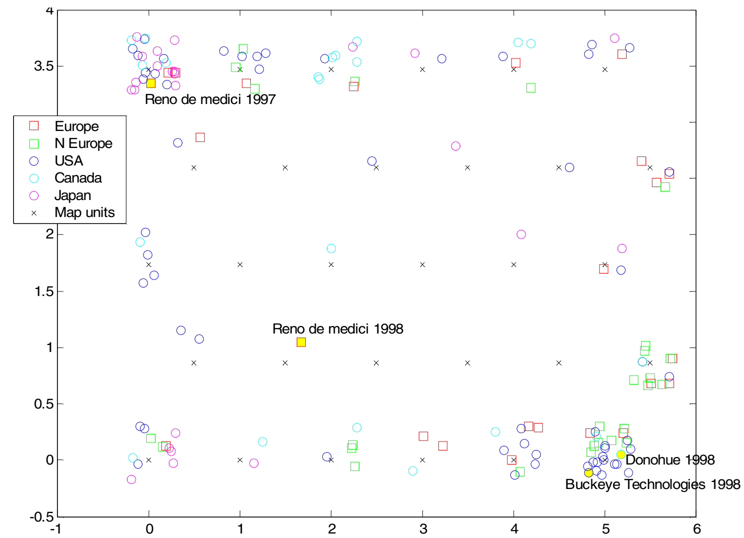

Figure 9 Self-Organizing Map – scatter plot view created with SOM Toolbox 2.0

Ultsch and Siemon (1989) developed the *U-matrix* graphic display to illustrate the clustering of the reference vectors, by representing graphically the distances between map units. In this visual representation, each map unit is typically represented by a hexagon. The line or border between two neighbouring map-units (hexagons) has a distinguishable colour that signifies the distance between the two corresponding reference vectors. Dark green signifies for large distances, and light green signifies similarities between the vectors, as indicated by the colour bar (Figure 10-Left).

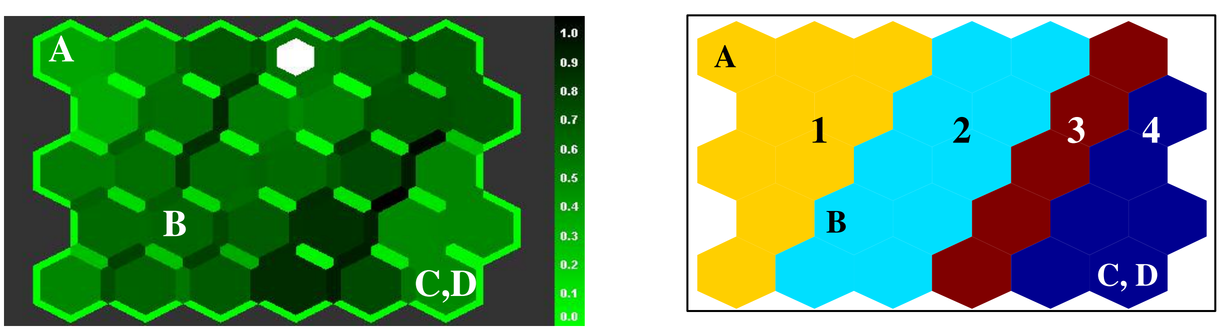

Figure 10 Left: Self-Organizing Map - U-matrix view created with Nenet 1.1 (1999); Right: Clustering of SOM view created with SOM Toolbox 2.0

By looking at the borders' colours in Figure 10-Left, one can distinguish the main clusters that exist in the data. A clustering algorithm (e.g., K-means) can be used to automatically partition the map into similar clusters (Figure 10-Right), creating the *clustering of SOM* view. The dataset appears to contain four main clusters. Based on Figure 9 and Figure 10, one can compare the companies of interest with respect to their membership in the identified clusters. Moreover, one can see the composition of each cluster with respect to the variable Region (e.g., Cluster 4 contains mostly American, Northern European and European companies).

It is also possible to visualize each data dimension using *feature planes*. These represent graphically the levels of the variables in each map unit. The colour red signifies high values of the variables, and blue and black correspond to low values of the variables (as indicated by the colour bars, Figure 11).

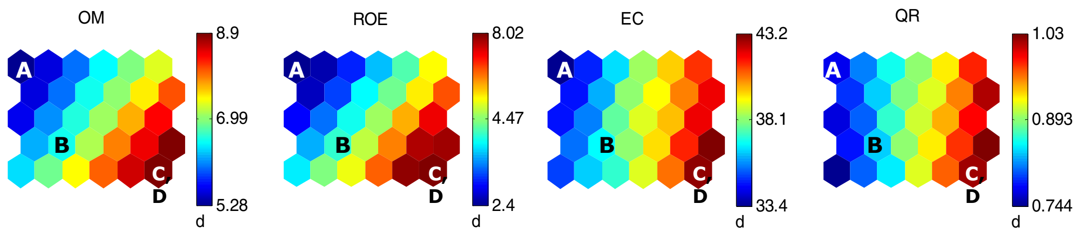

Figure 11 Feature planes created with SOM Toolbox 2.0

The feature planes facilitate the comparison of the companies of interest. The feature planes also help to identify the relationships between variables (e.g., OM is correlated with ROE; EC is correlated with QR).

By examining the features planes in parallel with the clustering of the SOM, one obtains the description of the four clusters identified previously as follows. Cluster 1 shows very low profitability, liquidity, solvency and efficiency. It contains the companies with the poorest financial performance. Reno de Medici 1997 is situated in this cluster (A). Cluster 2 shows medium profitability, solvency, and liquidity, but low efficiency. Reno de Medici 1998 belongs to this cluster (B). Cluster 3 shows good profitability, liquidity and solvency. Efficiency is medium to low. Cluster 4 shows very high profitability, solvency, liquidity and efficiency. It contains the companies with the best financial performance, among which Buckeye Technologies 1998 and Donohue 1998 (C and D) are situated.

Comparison of visualization techniques

In the previous section, we illustrated the use of multidimensional data visualization techniques for exploring financial performance data. All visualization techniques used are capable of providing an overview of the dataset under analysis, and different techniques uncover different patterns in the data. We highlighted the capabilities of each technique for answering the business questions and data mining tasks related to the financial benchmarking problem. In this section, we compare the techniques with respect to three criteria: 1) their capabilities to answer the questions and data mining tasks formulated for the financial benchmarking problem; 2) their capabilities to show data items or data models; and 3) the type of data used as input for the visualization technique (i.e., original data or normalized data).

First, we provide in Table 2 a *subjective comparison* of the techniques with respect to their capabilities for solving the data mining tasks related to financial benchmarking. The assessment concerns only this business problem and the associated dataset. We do not intend to generalize the results to other datasets, because for a different dataset (with different types of data, number of variables, number of observations, underlying structure) the results of the evaluation could be different. Table 2 can be used as a means to map the data mining tasks to different visualization techniques for this dataset. This table can, therefore, be used in the process of selection of visualization techniques suitable for representing and exploring the data in the financial benchmarking problem.

Table 2 shows that there are data mining tasks for which more than one visualization technique can be used. On the other hand, one data mining task may be addressed using different visualization techniques but with a different outcome (e.g., clustering solutions produced by the SOM and PCA). Almost all visualization techniques can facilitate the comparison among companies. Moreover, all techniques, except Sammon's mapping, are effective for finding outliers or anomalies in this dataset. The scatter plots, survey plot, permutation matrix, PCA and the SOM are good in showing relationships between ratios. The SOM and PCA are capable of showing and describing clusters. The treemap, line graphs, and survey plot are capable of describing class characteristics. Treemap is typically effective in displaying hierarchical data, and in our example proved to be very effective in making comparisons between companies and highlighting the characteristics of companies from one region or another with respect to the values of financial ratios. Sammon's mapping is effective in displaying class distribution, but does not provide means to describe the characteristics of the classes.

Table 2 also shows that the most effective techniques in uncovering patterns in this specific dataset are the SOM (when all views are analysed together) and PCA. If we assess separately each SOM-based visualization technique, the results show that all SOM views show the clustering of the data, but the other patterns are uncovered to a different extent by each SOM view.

Table 2. The capabilities of the visualization techniques on the dataset under analysis

Visualization technique	*Outliers detection*	*Dependency analysis*	*Clustering*	*Cluster description*	*Class description*	*Comparison*
Line graphs	✓	✓			✓	✓
Permutation matrix	✓	✓				✓
Survey plot	✓	✓			✓	✓
Scatter plot matrix	✓	✓				✓
Parallel coordinates	✓	✓				✓
Treemaps	✓				✓	✓
PCA	✓	✓	✓	✓		✓
Sammon's mapping					✓*	
SOM – scatter plot	✓		✓		✓**	
SOM – U-matrix			✓			
SOM – clustering			✓			
SOM – feature planes		✓	✓	✓		✓
SOM – all views combined	✓	✓	✓	✓	✓	✓

* Sammon's mapping is capable of organizing the dataset so that different classes are distinguishable but does not provide a means to interpret and describe the classes.
** SOM – scatter plot view is capable of showing where the companies from different classes (regions) are mapped but does not provide a means to interpret and describe the classes.

Second, the visualization techniques are compared with respect to their capability for showing data items or data models. All techniques display the data items. The SOM displays also a data mining model (e.g., the clustering of the data). In the former case, the user has to use his/her perceptual abilities to distinguish the patterns of interest. In the later case, the data model is automatically generated and displayed by the visualization.

Third, the visualization techniques are compared with respect to the type of data processed. The following visualization techniques represent the original data: multiple line graphs, permutation matrix, survey plot, scatter plot, parallel coordinates, and treemap. The other techniques represent standardized or normalized data: PCA, Sammon's mapping, and SOM. The visualizations obtained using standardized or normalized data are more difficult to interpret.

In summary, the techniques reviewed in this paper complement each other in uncovering all patterns in the financial benchmarking dataset. Using a single technique for data exploration may result in a limited understanding of the data. Therefore, the use of multiple techniques could be beneficial for the user. For example, combining different visualizations that are based on the SOM, we obtain a very good understanding of the data, while if we consider only one view, we understand very little about the data. One benefit of using multiple visualizations is that one can see different facets of the data and problem under investigation by using visualizations that uncover distinct patterns. Another benefit is that the analyst has the possibility to confirm that the patterns or outliers highlighted by one visualization technique are indeed real, and not an artefact, thereby increasing confidence in the findings. A third benefit is given by the descriptive power of some techniques over the others.

Conclusion

In this paper, we reviewed nine multidimensional data visualization techniques for representing financial performance data. We investigated the capabilities of different visualization techniques for uncovering interesting patterns in financial performance data, patterns described in terms of outliers, clusters, classes, relationships and trends.

By deriving business questions and data mining tasks from the financial benchmarking problem as recommended by Soukup and Davidson (2002), and mapping these tasks to appropriate visualization techniques, we provided a means to subjectively compare and assess the capabilities of different visualization techniques to solve the financial benchmarking problem. This approach can serve visual data mining systems' developers in assessing the strength of various techniques in the early stage of system development and accordingly select the most appropriate techniques. The approach can be extended by involving users in further evaluation studies of the selected visualization techniques.

We highlighted the potential benefits of using multiple visualization techniques for solving a business problem such as the financial benchmarking problem and uncovering all interesting patterns in data. Empirical studies of users facing multiple visualizations are needed to quantify these benefits. The study can also be extended by analysing other financial datasets and/or other multidimensional data visualization techniques.

Acknowledgements

I gratefully acknowledge the contributions of Professor Barbro Back and Dr. Tomas Eklund, and the financial support of Tekes (grant number 40435/05) and the Academy of Finland (grant number 104639).

References

Bertin, J., Semiology of Graphics: Diagrams, Networks, Maps, (W.J. Berg, Trans.), Madison, WI: University of Wisconsin Press, 1967/1983.

Card, S.K., Mackinlay, J.D., and Shneiderman, B., Readings in Information Visualization - Using Vision to Think, San Francisco, CA: Morgan Kaufmann Publishers, 1999.

Demsar J., Zupan B., and Leban G., "Orange: From Experimental Machine Learning to Interactive Data Mining", White Paper (www.ailab.si/orange), Faculty of Computer and Information Science, University of Ljubljana, 2004.

Eklund, T., The Self-Organizing Map in Financial Benchmarking, TUCS Doctoral Dissertation, Åbo Akademi University, Turku, 2004.

Hinterberger, H. H., and Schmid, C., "Reducing the Influence of Biased Graphical Perception with Automatic Permutation Matrices", SoftStat '93, Proceedings of the Seventh Conference of the Scientific Use of Statistic-Software, March 1993.

Hoffman, P. E., and Grinstein, G. G., "A Survey of Visualizations for High-Dimensional Data Mining", in Fayyad, U., Grinstein, G. G., and Wierse, A. (eds.), Information Visualization for Data Mining and Knowledge Discovery, San Francisco: Morgan Kaufmann, 2002.

Inselberg, A., "The Plane with Parallel Coordinates", Special Issue on Computational Geometry, The Visual Computer 1, 1985, pp. 69–91.

Johnson, B., and Shneiderman, B., "Treemaps: A Space-Filling Approach to the Visualization of Hierarchical Information Structures", Proc. of IEEE Visualization '91 Conf., San Diego, 1991, pp. 284 291.

Keim, D.A., "Information Visualization and Visual Data Mining", IEEE Transactions on Visualization and Computer Graphics, Vol. 8, No. 1, January-March 2002.

Kohavi, R., Rothleder, N. J., and Simoudis, E., "Emerging Trends in Business Analytics", Communications of the ACM, 45(8), 2002, pp. 45-48.

Kohonen, T., Self-Organizing Maps, Third edition, Berlin: Springer-Verlag, 2001.

Marghescu, D., Multidimensional Data Visualization Techniques for Financial Performance Data: A Review, TUCS Technical Report no. 810, Turku Centre for Computer Science, Turku, Finland, February 2007.

Nenet 1.1, http://koti.mbnet.fi/~phodju/nenet/Nenet/General.html, 1999.

Sharma, S., Applied Multivariate Techniques, New York, NY: John Wiley & Sons, 1995.

SOM Toolbox 2.0 for MATLAB, http://www.cis.hut.fi/projects/somtoolbox/, 2005.

Soukup, T., and Davidson, J., Visual Data Mining: Techniques and Tools for Data Visualization and Mining, New York, NY: John Wiley & Sons, 2002.

The MathWorks, Statistics Toolbox for Use with MATLAB - User's Guide, Version 4, 2002.

Treemap 4.1, http://www.cs.umd.edu/hcil/treemap/, 2004.

Ultsch, A., and Siemon, H., Technical Report 329, Univ. of Dortmund, Dortmund, Germany, 1989.