A Novel Explorative Visualization Tool for Financial Time Series Data Analysis

Matthias Schaefer[1]
schaefer@dbvis.inf.uni-konstanz.de

Franz Wanner[1]
wanner@dbvis.inf.uni-konstanz.de

Roman Kahl[1]
roman.kahl@uni-konstanz.de

Leishi Zhang[1]
zhang@dbvis.inf.uni-konstanz.de

Tobias Schreck[1]
schreck@dbvis.inf.uni-konstanz.de

Daniel A. Keim[1]
keim@uni-konstanz.de

ABSTRACT

Analyzing high dimensional time series data can be challenging, especially when the data consists of a large number of long time series. The task involves supporting interactive analysis of the data in both global and local fashion. In this paper we introduce a novel visual data analysis tool, which facilitates such type of tasks. The tool advances existing visualization techniques by integrating a pixel based visualization technique, which can display a large volume of data without overlaps, and line graph visualization, which provides an intuitive understanding of patterns and trends in the time series data.

Keywords

Time Series, Financial Data, Visualization

1. INTRODUCTION

A time series is a sequence of observed values ordered in time. Such data exist in many application domains such as finance, communication and science. Extracting trends and patterns from time series data provides essential knowledge for many data analysis tasks such as performance analysis, fraud detection and decision support. Various visualization techniques have been proposed and applied in many application scenarios [5][4][10][12][3][11][6][8][1][13] to help extracting patterns from time series data. However, challenges still exist: for example preserving global views of long sequences of observations while displaying details-on-demand information, showing multiple time series data while avoiding a cluttered screen, or supporting a comparison of different time series.

In this paper we introduce a novel visualization system for analyzing share performance from historical stock price time series and sector indices data. Typically, the stock price data contains time series which record the share price measured over a long period of time. Each dataset contains the time series of a large number of shares. Analysts often look at the risk and return of the shares by comparing the performance of individual share prices against the sector index which records the average performance of all shares in the sector. Performance of different shares is often compared to support decision making in financial investments. During the explorative analysis process, the focus of interest may switch between the global view of all share performance and performance of individual shares over a specific time period. To facilitate such analysis tasks, we design our system in such a way that the users can easily filter the data, switch the focus of the exploration, compare share performance, and see the volatility (the deviation of the share performance) of shares. Our tool starts with a pixel based view which shows a global picture of the whole market performance. Each pixel unit represents the performance of one share at one time point and each row represents one share. Analysts can click on a pixel unit to see the detailed view of the selected share prices as line graph(s) lying on top of the colored background. Various user interactions are implemented to allow analysts to see the correlations, filter the data, change time intervals and to zoom into a particular time period to obtain detailed information.

The main contributions are: 1) integration and extension of pixel based visualization and line graph visualization for the analysis of share performance; 2) an interactive data analysis system for analyzing stock market time series data. The tool is developed for financial time series data analysis, but the proposed visualization techniques can be applied to other time series visualization applications. The paper is organized as follows: In Section 2 we introduce background and discuss related work, the design and implementation of the system is shown in Section 3, and Section 4 shows the benefits by use cases on real world applications. In Section 5 we draw conclusions and discuss future work.

2. RELATED WORK

Time series analysis plays an important role in stock market trading, where trading prices change every second. These changes are recorded in a database as time series. The data is often used to extract trends and patterns to support decision making in trading and investment. Typically, given historical stock market data, analysts want to look at the risk and return of individual shares as well as evaluate the performance of a share against the average performance of shares of a similar type. In the past decades various visual-

VAW2011 6-8 September 2011, University College London, UK

[1]University of Konstanz, Germany

ization techniques have been proposed to support time series data analysis. For example, spiral techniques [11], cluster and calendar based visualization techniques [12], VizTree [6], TreeMap [8] and displays based on Self-Organizing Map-based clusterings [7]. In this section, we focus on three visualization techniques, TimeSearcher[1, 2], Recursive Pattern [4] and Growth Matrix [13] which are used to visualize financial data and closest to the techniques we applied in our system design. We discuss the strengths and weaknesses of each technique which motivate our system design.

TimeSearcher is designed for interactive querying and exploration of time series data. The software applies traditional line graphs to display time series as a river-plot view. The interactive query facilitates the explorative analysis on the time series, and the pattern based search makes it possible to forecast the development of a time series. However with large multiple time series the overlap between lines can obscure individual time series and makes it difficult to find patterns.

Recursive Pattern is a pixel based approach which was designed to avoid the over plotting problem when visualizing large time series. The technique combines the ratio between the stock price and the sector index into a single input variable and maps this ratio to colors. Each value is visualized as a colored pixel in the display. The advantage of the pixel based technique is its ability to visualize the global view of a large volume of share performance data without cluttering the display. Some patterns can easily be detected in such a visualization, for example we can observe maxima or minima at certain time stamps. The disadvantages of this technique are that it lacks comparison on some detailed information such as the separate share price and sector index, and the static view makes it difficult to compare individual shares.

Growth Matrix visualization is proposed to avoid large data volumes causing an over plotted display. This technique provides an effective way for analyzing financial data, since it allows the identification of strong and weak periods of assets as compared to global market characteristics, and thus allows a more encompassing visual classification into "good" and "poor" performers than existing chart techniques. Like most other pixel based visualization techniques, the drawback of Growth Matrix is the visualization stays as static views and the implementation does not support dynamic queries. The quantities of differences between values are hard to estimate.

3. DESIGN AND IMPLEMENTATION

None of the existing visualization techniques on its own solves all the problems in visualizing large-and-long time series. Hence, in the design of the tool we take advantage of several visualization techniques and integrate them such that the large-and-long time series can be analyzed in both global and local fashion without losing information or over plotting the display.

Performance measure: Given a particular time frame, we define the performance measure to be the ratio of the share price against the corresponding sector index over the same period of time. This new variable is mapped into the visualization by using color. We use a red and green color map based on the convention in the financial domain that red indicates negative performance and green indicates positive performance. Thus no extra cognitive load is needed while

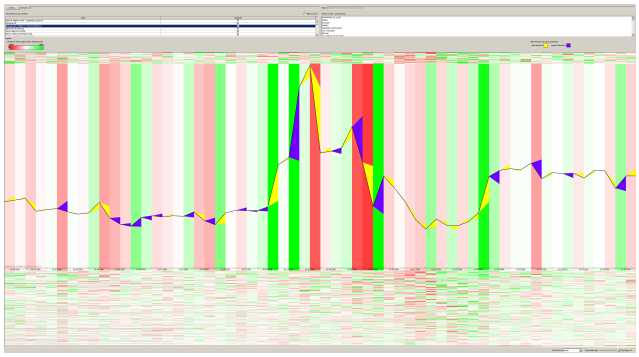

Figure 1: Overview of the system: GUI interface with a global view of stock price time series visualization and details-on-demand option for a selected share.

examining the graphical representations. The performance measures are mapped to the background of the line graph in vertical strips corresponding to the respective time stamps (see Figure 1).

Visualization of an individual share price series: We use a dynamic line graph to visualize a stock price series for good and easy understanding of trends and changes. The dynamic line graph visualization allows the analyst to zoom in or out of a particular time period or to change time intervals. The performance measures are mapped to the background of the line graph (see Figure 1). Red colour means that the share performs worse than the sector average; green means that the share performs better than the sector average.

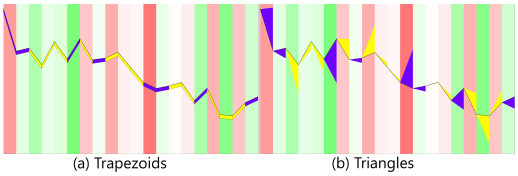

Figure 2: Two different visualizations (trapezoid and triangle slope polygons) for correlation analysis and pattern detection.

The correlation within a time stamp between the share price and the sector index is visualized as yellow or violet "slope polygons" which lie on top of the line graph. This design decision is taken because yellow and violet are complementary colors. A yellow polygon means that there is a positive correlation between the share and the sector index; in return violet means that there is a negative correlation between these two. Two types of slope polygons are implemented: trapezoid slope polygons and triangle slope polygons (see Figure 2). The triangles have the advantage of visualizing the quantitative differences between a share price and its sector index (the length of the polygon line from the start to the end of a time interval indicates the difference). However, when the change is small, the slope polygons can be too small to see. Hence the trapezoid slope polygons are also provided. The analyst can switch between the two slope polygon types to look for the information they need. For example, in Figure 2(b), one can see although the share price has dropped be-

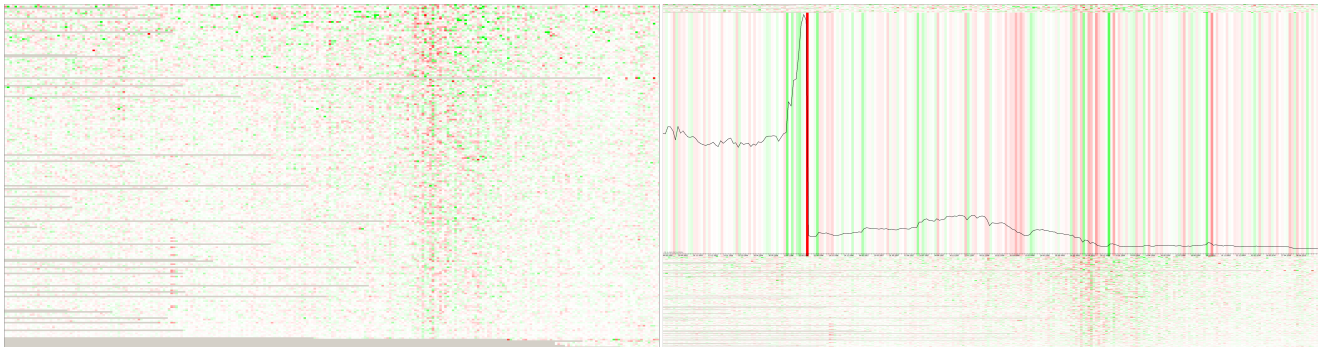

Figure 3: (left): Overview of 222 stocks over 5 years in the oil sector in the World-Datastream Oil and Gas producers Index with weekly intervals; (right): The line graph appears on top of the pixel based visualization on a mouse click, here with focus on "Oil & Gas Exploration & Production".

tween the 5th and the 6th time point (down trend in the line graph), the performance of the share is still better than the sector index (green background). From 6th to 7th time unit, there is a sharp increase in the share price whilst the sector index behaves very differently (big negative correlation).

Global and detailed view visualization: In global view visualization the performance of each share at each time point is shown as a colored pixel. The detailed values of each pixel can be seen in a tooltip with mouse over action. The line graphs of stock price time series are laid on top of the pixel based visualization. Although the line graphs are hidden in the global picture, the analyst can easily click on any of the pixels and see the details of the price series in the line graph (see Figure 3 (right)). The selected share takes the major part of the available space and the other shares are squeezed together, keeping their relative positions, so that they are still visible and can be considered for a comparison.

Filtering, zooming, and details-on-demand: A colormap slider is implemented as a data filter such that the analyst can select a performance range. The data that fall outside of the range will be removed from the visualization. While filtering, the number of colors in the colormap is reduced or the colors are renormalized according to the remaining data values. For example if the performance range is set from -50% to -10% by the colormap slider, a share with -10% performance measure has the "best" performance, hence it is colored green. This technique allows the analyst to see the relative performance of shares compared to its peers and amplifies the differences between data values such that some hidden patterns or outliers are discovered more easily. The system allows the analyst to change time intervals and specify time frames.

A horizontal zooming function is provided. The analyst can adjust the level of details in the line graph by zooming in and out of the view horizontally. This is especially useful when the analyst needs to study the correlations between the share price and the stock index and see the quantitative differences. The analyst can also select a time series in the visualization panel to expand a particular line graph. While more details of the selected share become visible in the expanded line graph and the other parts become smaller (see Figure 1), the visualization still provides the global picture of the whole dataset.

4. APPLICATIONS

4.1 Comparison in European banking sector

In this section we show the effectiveness of the system through its application in the analysis of two real world data sets. In the first use case we study the performance of the Deutsche Bank share which belongs to the European bank sector. The sector includes the shares of 54 big European banks (STOXX, see [9]). We have the Deutsche Bank share price time series over 5 years (from 01.08.2005 to 01.08.2010) on a daily basis. Figure 4 shows a screenshot of our system displaying the weekly aggregated data compared to the European bank index. There is an obvious turbulent pattern between the end of 2008 and the beginning of 2009 when the global crisis in the finance market took place. There were numerous transactions in the market during that period. On a daily basis the Deutsche Bank share sometimes developed very differently from the sector index.

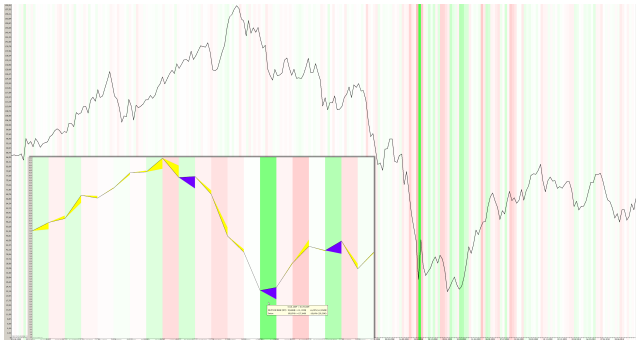

Figure 4: Deutsche Bank and STOXX Europe 600 Banks Index five years performance on a weekly aggregated basis from 01.08.2005 to 01.08.2010; (lower left): Performance of Deutsche Bank compared against this Index over the period aggregated quarterly: it has similar performance to the sector except for the first quarter of 2009: here the the stock records a +6.3% performance which means that the share outperformed the sector index by 25%.

The aggregation provided by the tool allows us to analyze the data at different levels by changing the time interval. Figure 4 (lower left) shows the line graph of the same data

on a quarterly basis. By changing from a small time interval to a bigger time interval, some price fluctuations in the small time interval are filtered out and thus the global trends can be perceived more easily. From the higher level performance graph we can see an interesting pattern which is hidden in the previous visualization: During the first quarter of 2009, Deutsche Bank performed much better than the sector average (see the strong green color in the background, which means the share performed well and the performance has a negative correlation to the sector index). The good performance of the Deutsche Bank share during the first quarter of year 2009 indicates that the bank had a rapid recovery from the financial crisis and the speed of recovery is faster than the sector average.

4.2 Explorative analysis of share performance in international oil and gas sector

In this use case study, we look at international oil and sector stock market data over 5 years (from 01.08.2005 to 02.08.2010) to see the shares with best and worst performance. To see the shares with bad performance, we set the displayed range to performances smaller than -10%. Figure 5 gives a global view of the filtering result. The performance values are renormalized after filtering; hence the performance of -10% is mapped to green and the maximal negative performance is mapped to red. As shown in the Figure 5, the share with the biggest loss (worst performance) stands out (the red pixel in blue circle in Figure 5). By clicking on the red pixel we get a details-on-demand view of the share (see Figure 3). We see that there was a sharp fall in the share price which caused the extreme negative performance value. The same filter function can be used for analyzing shares with best performance.

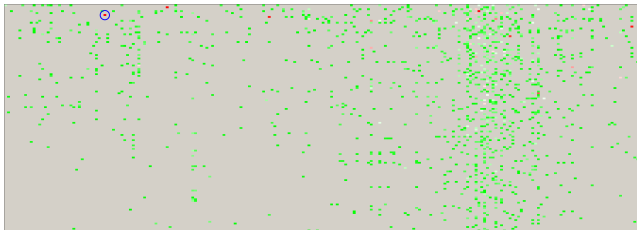

Figure 5: Limited to performances below 10% compared to the index. Shares with very bad performance emerge clearly, for example see the red pixel marked with a blue circle.

5. CONCLUSIONS AND FUTURE WORK

Visualizing large numbers of long time series faces great challenges in terms of display space, effective visual representations and interactive visual interfaces. In this paper we present a visualization system that combines different visual techniqes for historical stock data analysis. Two novel visual representations, triangle and trapezoid polygons, are proposed to visualize correlation between individual share performance and the sector index. A colormap slider is designed to help filtering data base on share performance. The two use cases analyzing the Deutsche Bank share and Oil and Gas sector data demonstrate that the system is able to successfully highlight some interesting patterns in the data and the visualization techniques scale well to large data sets. The current version of the system implements a basic sorting function which allows the analyst to see the shares ordered by their volatility. We want to further extend this analysis by implementing a similarity measure based on volatility values and use this measure to group shares that have similar performance patterns. This will be useful in pattern detection and time series prediction. An evaluation with real end users will be carried out to improve the design and functionalities of the system. We also want to integrate more statistics and data mining methods such as clustering, association rules and believe network models to support higher level data analysis tasks.

6. REFERENCES

[1] P. Buono, A. Aris, C. Plaisant, A. Khella, and B. Shneiderman. Interactive Pattern Search in Time Series. In *Proceedings of the Visualization and Data Analysis (VDA 2005)*, pages 175–186, January 2005.

[2] P. Buono and A. L. Simeone. Interactive shape specification for pattern search in time series. In *AVI '08: Proceedings of the working conference on Advanced visual interfaces*, pages 480–481, NY, USA, 2008. ACM.

[3] P. Groenen and P. H. Franses. Visualizing time-varying correlations across stock markets. *Journal of Empirical Finance*, 7(2):155–172, 2000.

[4] D. A. Keim. Pixel-oriented database visualizations. *SIGMOD Record, Special Issue on Information Visualization*, pages 35–39, December 1996.

[5] D. A. Keim, M. Ankerst, and H.-P. Kriegel. Recursive pattern: A technique for visualizing very large amounts of data. In *IEEE Visualization*, 1995.

[6] J. Lin, E. Keogh, S. Lonardi, J. P. Lankford, and D. M. Nystrom. Viztree: A tool for visually mining and monitoring massive time series databases. In *Proceedings of International Conference on Very Large Data Bases*, pages 1269–1272, 2004.

[7] T. Schreck, J. Bernard, T. Von Landesberger, and J. Kohlhammer. Visual cluster analysis of trajectory data with interactive kohonen maps. *Information Visualization*, 8:14–29, January 2009.

[8] B. Shneiderman. Treemaps for space-constrained visualization of hierarchies. Web document, http://www.cs.umd.edu/hcil/treemap-history/, 2009.

[9] STOXX. Stoxx europe 600 banks, 2010.

[10] M. Wattenberg. Visualizing the stock market. In *CHI '99 extended abstracts on Human factors in computing systems*, CHI '99, pages 188–189, NY, USA, 1999.

[11] M. Weber, M. Alexa, and W. Müller. Visualizing time-series on spirals. In *Proceedings of the IEEE Symposium on Information Visualization 2001 (INFOVIS'01)*, pages 7–, Washington, DC, USA, 2001. IEEE Computer Society.

[12] J. J. V. Wijk. Cluster and calendar based visualization of time series data. In *Proceedings of the 1999 IEEE Symposium on Information Visualization*, pages 4–9. IEEE Computer Society, 1999.

[13] H. Ziegler, M. Jenny, T. Gruse, and D. A. Keim. Visual market sector analysis for financial time series data. In *Visual Analytics Science and Technology (VAST)*, pages 83–90, 2010.