Proceedings of the Twenty-Sixth International Florida Artificial Intelligence Research Society Conference

Visualizing Stock Market Data with Self-Organizing Map

Joel Joseph and Indratmo
Department of Computer Science, Grant MacEwan University
JosephJ2@mymail.macewan.ca, indratmo@macewan.ca

Abstract
Finding useful patterns in stock market data requires tremendous analytical skills and effort. To help investors manage their portfolios, we developed a tool for clustering and visualizing stock market data using an unsupervised learning algorithm called Self-Organizing Map. Our tool is intended to assist users in identifying groups of stocks that have similar price movement patterns over a period of time. We performed a visual analysis by comparing the resulting visualization with Yahoo Finance charts. Overall, we found that the Self-Organizing Map algorithm can analyze and cluster the stock market data reasonably.

Introduction

The stock market contains a huge amount of data that vary over time. The stock price of a company is determined by various factors, from the performance of the company itself to the condition of the economics in general. To manage portfolios well, fund managers and investors must analyze stock market data regularly to identify undervalued companies or "hot" stocks depending on their investment goals and time frames. Finding useful information in such complex data, however, requires high analytical skills and effort.

The complexity of stock market data and the tasks involved in analyzing this data call for a tool that can amplify human cognition. To help people gain insight into stock market data, we developed a visual analytical tool that analyzes the historical price movements of publicly traded companies, that clusters similar data together, and that visualizes the data onto a two-dimensional space using a learning algorithm called Self-Organizing Map (SOM) (Kohonen 1990). We combined a learning algorithm and information visualization to exploit human perceptual ability to recognize patterns and derive a lot of information from visualization with little effort (Ware 2000).

Copyright © 2013, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

From a practical perspective, our tool can assist people in analyzing stocks and portfolios. Our SOM implementation learns about stock market data and clusters companies in the dataset onto a two-dimensional grid where similar companies appear close to one another. This map can reveal useful patterns in the data and help with portfolio management. For example, if most stocks in an investment portfolio appear in the same region of the map, the portfolio may lack diversification, as the stocks have similar trading patterns. The investment manager can then follow up on such visual analysis by examining the details of the portfolio.

We evaluated the resulting visualization by comparing it visually to Yahoo Finance line charts. We plotted stock prices of companies that were put in the same group by our tool and those that were put in different groups. Overall, we found that stocks in the same group have a more similar trading pattern than those in different groups.

Related Work

Prior work on stock market visualization includes FolioMap (Jungmeister and Turo 1992), Smartmoney's Map of the Market (Wattenberg 1999), and FundExplorer (Csallner et al. 2003). To provide an overview of stock-related data, these visualizations organize the datasets hierarchically and then construct tree-map visualizations (Shneiderman 1992).

FolioMap is intended to help fund managers manage portfolios in large corporate settings (Jungmeister and Turo 1992). Its features and visualizations include how to help users decide when to buy or sell stocks, assess the performance of their portfolios, and identify the general market interest in a particular stock. FolioMap also enables upper-level managers to review portfolios managed by their subordinates.

FundExplorer has a similar purpose to FolioMap's: helping users manage their portfolios (Csallner et al. 2003). In particular, FundExplorer is designed to assist users in exploring mutual funds and developing diversified portfolios.

Rather than focusing on specific investment portfolios, Smartmoney's Map of the Market is concerned with the general market conditions and displays the stock prices of more than 500 stocks in a single view (Wattenberg 1999). The map organizes companies by their business sectors and visualizes them as rectangles. The size and color of a rectangle represents a company's market capitalization and its stock price performance relative to that on the previous trading day. This visualization essentially shows the mood of the stock market on a daily basis.

Another related work is Simunic's (2003) stock-market-chart visualization. Simunic uses the SOM algorithm to cluster stock-market-chart shapes and produce representative charts for groups of similar chart shapes. Unlike our work that analyzes historical stock prices, Simunic focuses on the chart shapes and the percentage increase or decrease of stock prices over 30-trading days. Such visualization of chart shapes is particularly useful for supporting technical analysis of the stock market.

Self-Organizing Map Implementation

The Self-Organizing Map (SOM) is an unsupervised learning algorithm that can reduce the dimensions of high-dimensional data and organize the data spatially onto a usually two-dimensional grid (Kohonen 1990; Kohonen and Honkela 2007). We chose the SOM algorithm because it has been found useful for analyzing and visualizing datasets in various domains, including financial data (Deboeck and Kohonen 2010).

The SOM algorithm works by initially assigning random weights on spatially organized nodes. In each iteration, the algorithm selects a random sample of data, then chooses the best matching unit (BMU)—the node that best matches the randomly selected sample within a neighboring distance—and changes the weights of neighboring nodes of the BMU to be more similar to the weights of the random sample based on their distances to the random sample. The SOM algorithm starts with a high learning rate and neighboring distance, and then these parameters are reduced slightly in next iterations.

Our SOM implementation uses the Euclidean distance to compute the best matching unit. Calculating the distance between two stocks, A and B, over a period of time involves the following steps. First, the daily closing prices of each stock are divided by its highest price during that period, resulting in $A_1 \ldots A_n$ and $B_1 \ldots B_n$, where A_i and B_i represent the closing prices of these stocks on a particular day divided by their highest prices. Then, the distance between these stocks is calculated using this formula:

$$d(A,B) = \sqrt{(B_1 - A_1)^2 + (B_2 - A_2)^2 + \ldots + (B_n - A_n)^2}$$

Based on this formula, the SOM algorithm determines the best matching unit by selecting the node that has the minimum distance to a selected sample.

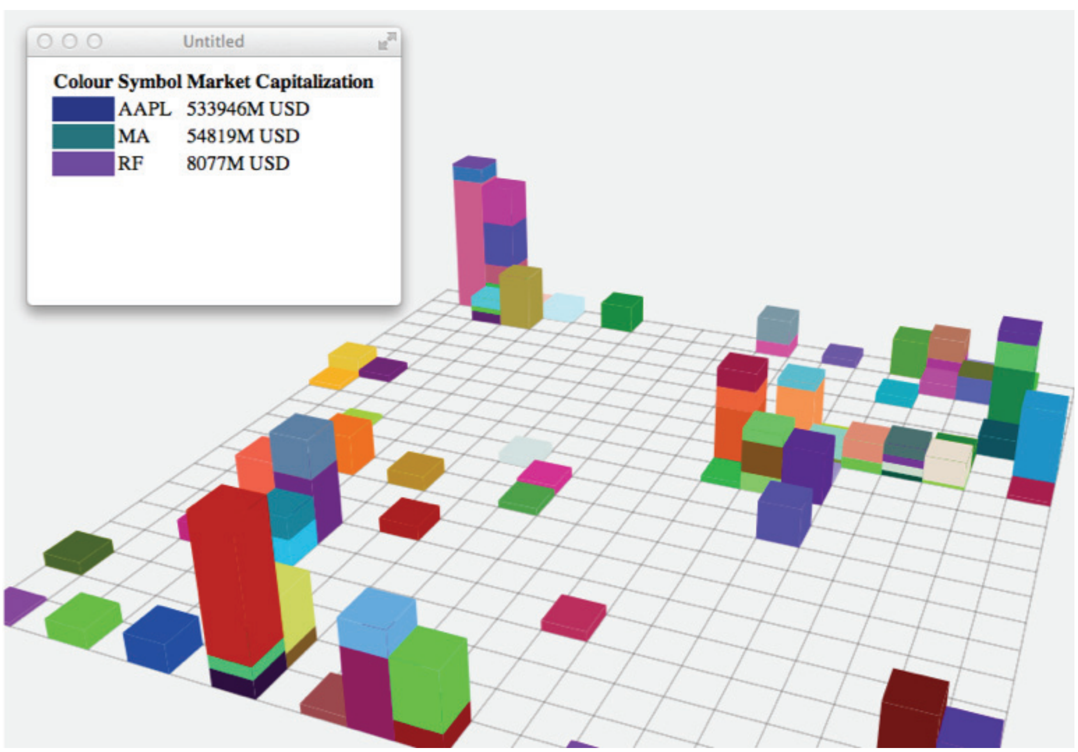

Figure 1: A Visualization of S&P 100 Stocks (Jan 1, 2012 – Jul 30, 2012)

The Visualization Tool

Visualization Design

Figure 1 shows a screenshot of our visualization tool showing the resulting map of S&P 100 stock data from Jan 1, 2012 to Jul 30, 2012. Our prototype uses a 20 x 20 grid to cluster stock market data spatially. A box in the visualization represents a company's stock where the volume of the box is proportional to the company's market capitalization. Each company is assigned a different color to help users recognize different companies. In choosing graphical representations of data, we followed the guidelines as outlined by Mackinlay (1986).

A stack of companies indicates that these companies share a similar price movement pattern over a certain period of time. Users can click on a stack, and the tool shows a list of companies on that stack. For example, when users click on the top-left corner, a popup window appears and shows a list of companies in that stack: APPL (Apple), MA (MasterCard), and RF (Regions Financial). Users can also rotate this visualization 360 degree so that they can focus on a different region of the map.

In designing this visualization, we used the following design principle in information visualization as a basis: "overview first, zoom and filter, then details-on-demand" (Shneiderman 1996). This design principle suggests that information visualization should start by providing an overview of the whole dataset, allow users to interact with the visualization, and display the details of an information item on demand. We realized, however, that at this stage, our prototype still supports a limited number of user interactions.

Software Prototype

Our prototype consists of three main parts. The first part is responsible for gathering stock market data from Yahoo Finance (http://finance.yahoo.com/). The program simply retrieves all the available data on the site (i.e., market capitalizations, dates and closing prices of publicly traded companies) and saves them into a local database.

The second part contains an implementation of the SOM algorithm. It is possible to modify the program to specify the time range of data to analyze (e.g., Jan 1, 2012 to Jul 30, 2012), which stocks to include in the analysis (e.g., the S&P 100 index), and a few other parameters that determine how the algorithm learns about the data. This program uses a data-interchange format (text-based) named JavaScript Object Notation or JSON (http://www.json.org/) to express the clustering results.

To help users explore the clustering results, the third part of our prototype parses the JSON document and creates a grid-based visualization as shown in figure 1.

Results and Discussion

Visual Comparison

To assess how well our visualization tool clusters stock market data, we performed a visual analysis of sets of stocks in the same groups (Figure 2) and those in different groups (Figure 3). We used Yahoo Finance line charts to plot stock price movements of these stocks over the same period of time used in our visualization (i.e., Jan 1, 2012 to Jul 30, 2012). Figures 2 and 3 show examples of these

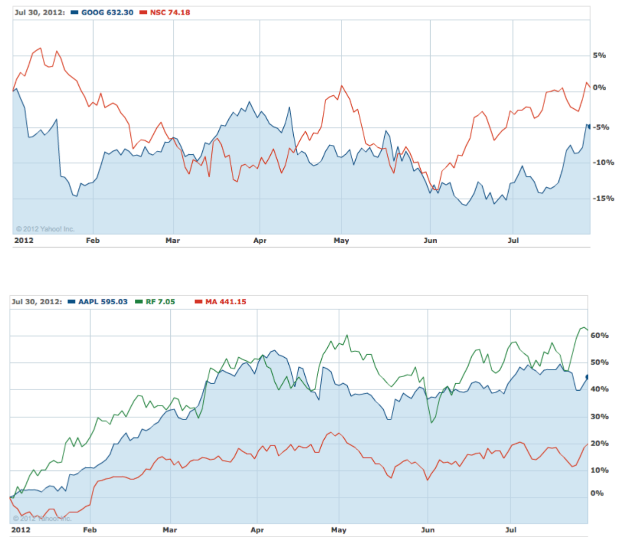

Figure 2: Visual comparison of stocks in the same group (above: GOOG and NSC, below: AAPL, RF, and MA)

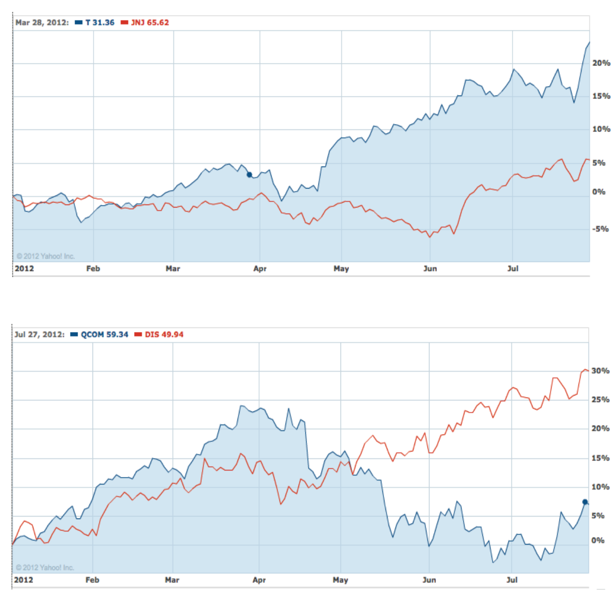

Figure 3: Visual comparison of stocks in different groups (above: T and JNJ, below: QCOM and DIS)

charts. Based on such comparisons, we found that, in general, stocks in the same cluster have more similar price movement patterns than those in different clusters. While a visual comparison lacks of mathematical rigor, it served our tool's purpose of providing a starting point for users to do further analysis of particular stocks. After all, no one should rely solely on an automatic tool for making investment decisions.

Potential Uses

Our visualization tool can support exploratory search and facilitate serendipitous discovery. For example, since stocks are clustered based on their historical price movement patterns, stocks in different industries may be grouped together (e.g., APPL and RF). Investors who are specialized in the technology sector may not be aware that a relatively small company such as RF (Regions Financial) had a very similar trading pattern to AAPL (Apple) from Jan 1 to Jul 30, 2012. Such discovery can help traders identify, for example, a set of hot stocks.

Besides revealing groups of similar stocks, our visualization can be used to find stocks that show different trading patterns. Instead of chasing hot stocks, some investors may prefer to buy undervalued or oversold stocks. To find such stocks, they can explore clusters of stocks that are away from a cluster of hot stocks. This strategy may help reduce the volatility of a portfolio and assist in decision-making process of buying and selling stocks.

Conclusion

We used the SOM algorithm to analyze S&P 100 stock price data from Jan 1 to Jul 30, 2012 and visualize the clustering results in a two-dimensional grid. We performed a visual comparison of the resulting visualization with Yahoo Finance line charts of the same dataset and found that the SOM worked reasonably well in clustering stock market data.

Users can use our visualization tool in ways that fit their investment styles and goals. Our visualization is not intended to automate portfolio management. Users are still responsible for doing due diligence and making investment decisions. However, our visualization can support exploration of stock market data, facilitate serendipitous discovery, and provide a starting point for users to do further analysis.

There are several directions to follow up on this initial project. First, more interaction features can be added to our prototype to support both specific and exploratory search. For example, users should be able to highlight selected data items (Becker and Cleveland 1987), focus on a subset of data (Shneiderman 1996), and search for a specific stock.

Second, instead of focusing on price movements, we can analyze fundamental data of companies, such as earnings per share, return on equity, price-earnings ratio, price-to-book ratio, dividend yield, and dividend payout ratio. Such analysis and visualization can be invaluable to investors who prefer fundamental analysis to technical analysis.

Third, a rigorous analysis can be conducted to study our SOM implementation and clustering results in detail. Such analysis may include examining the scalability and limitations of the SOM algorithm and comparing it with other competing methods.

Acknowledgments

This project was supported by MacEwan Faculty Research, Scholarly Activity and Creative Achievements Fund.

References

Becker, R. A. and Cleveland, W. S. 1987. Brushing Scatterplots. *Technometrics*, 29 (2), 127–142.

Csallner, C., Handte, M., Lehmann, O., and Stasko, J. 2003. FundExplorer: Supporting the Diversification of Mutual Fund Portfolios Using Context Treemaps. In Proceedings of the 9th Annual IEEE Conference on Information Visualization, 203–208.

Deboeck, G. J. and Kohonen, T. eds. 2010. *Visual Explorations in Finance with Self-Organizing Maps*. Springer-Verlag.

Jungmeister, W-A. and Turo, D. 1992. Adapting Treemaps to Stock Portfolio Visualization. Technical Report, CAR-TR-648, University of Maryland, College Park, MD.

Kohonen, T. 1990. The Self-Organizing Map. In Proceedings of the IEEE, 78 (9), 1464–1480.

Kohonen, T. and Honkela, T. 2007. Kohonen Network. Scholarpedia, 2(1): 1568.

Mackinlay, J. 1986. Automating the Design of Graphical Presentations of Relational Information. *ACM Transactions on Graphics*, 5 (2), 110–141.

Shneiderman, B. 1992. Tree Visualization with Tree-Maps: 2-D Space-Filling Approach. *ACM Transactions on Graphics*, 11(1): 92–99.

Shneiderman, B. 1996. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In Proceedings of the IEEE Symposium on Visual Languages, 336–343.

Simunic, K. 2003. Visualization of Stock Market Charts. In Proceedings of the 11th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision.

Ware, C. 2000. *Information Visualization: Perception for Design*. Morgan Kaufmann Publishers.

Wattenberg, M. 1999. Visualizing the Stock Market. In CHI '99 Extended Abstracts on Human Factors in Computing Systems, 188–189.