A Visual Analytics System for Stock Data

Shenghui Cheng[1,2], Zhifang Jiang[2], Zhiyuan Zhang[1] and Klaus Mueller[1]

[1] Visual Analytics and Imaging Lab, Computer Science Department, Stony Brook University, NY, USA and SUNY Korea, Songdo, Korea

[2] Human-Computer Interaction and Virtual Reality Lab, School of Software, Shandong University, China

ABSTRACT

Despite the occasional downfall, stocks have proven to be good investments. However, making money in stock requires a great deal of research. With just under 7,000 stocks listed in the US alone, picking the right stock at the right time can be a daunting task. There are many metrics by which to rate the prospects of a given stock and seasoned investors typically weight these by their own experience to decide on an acquisition. In this poster we describe a visual analytics system purposed to aid in this task.

1 INTRODUCTION

Stock data analysis has been a hot topic for many years. Inspired by the prospects of lucrative gains, much theory and technology has been developed to determine influencing factors and future trends about stocks. A wealth of metrics has been defined to rate the prospects of a stock to flourish in the near future. Some of these are earning-per-share, dividend yield, return-on-equity, MACD (moving average convergence divergence), EBITDA, and many others. Some of these can be derived from a company's report to shareholders, while others can be retrieved from popular stock market sites such as Yahoo! Finance. The most successful investor is one who looks at all of these metrics, balances them with their own (secret) formula of success and based on that informed assessment acquires or sells a given equity [2].

These valuation metrics essentially form a high-dimensional space of stocks and one of the challenges is to visualize and explore it. While visual tools for financial data do exist, they typically use quite traditional paradigms, such as scatterplots, bar charts, pie charts and heat maps – popular are also candlestick plots and tree maps. While these visualizations are clearly helpful, none of them can truly convey – and let users explore – the interactions that exist in high dimensional metrics space. In our work we strived to devise a visual analytics tool that would be similar to existing tools used in the financial domain, such as scatterplots in which each stock forms a plotted point. This ruled out parallel coordinates where a stock forms a polyline. Conversely, both our biplot and our GBC interface use point plots.

2 OUR HIGH DIMENSIONAL VISUAL ANALYTICS INTERFACE

A scoring function is typically used to judge and rank a given stock in relation to other stocks. But a score is just a single cumulative number. Warren Buffet, one of the most successful investors of our times has often said that knowing the reasons for a score is just as important as the score itself [1]. In this spirit, our visual interface aims to 'explain' a stock's score by ways of the location of its representative point in the depiction of the high-D space. Finally, as mentioned above, individual investors typically give different weights to the various metrics to arrive at the cumulative score. Hence, the score function in our interface can be determined according to the weight of each metric. Sliders are available that allow users to tune these weights and interactively see the effect of this action on the distribution of valuated stocks. In the following we show both our biplot and our GBC interface.

[1]Email:{shenghui.cheng, zhiyuan.zhang, klaus.mueller}@stonybrook.edu [2]E-mail:zfjiang@sdu.edu.cn

2.1 Biplot

The biplot visualization (see Fig. 1) maps metric axes and data points into a 2D space spanned by the two main Principal Component Analysis (PCA) axes.

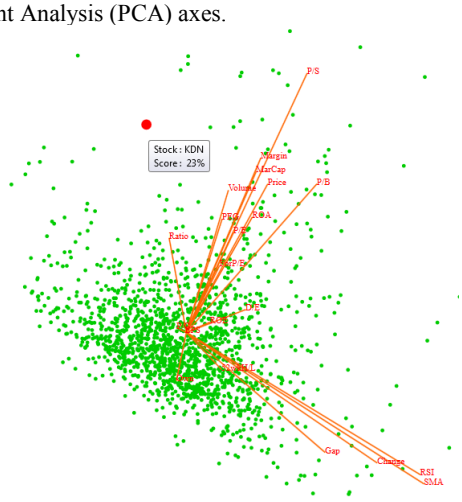

Figure 1: Data with all metric axes visualized in a biplot.

In Fig.1, the length of an axis line approximates the variance of the stocks within the respective metric, while the cosine of the angle between two axis lines approximates the correlation between the metrics they represent. The projection of a specific point onto an axis line approximates the value of that observation on the metric that the line represents, and the distance between two points approximates the Euclidean distance between two stocks in the multivariate space. In our interface, when the mouse moves over one point, it is enlarged and colored red and the label shows the stock's name and score. In Fig. 1, we labeled a stock with symbol KDN with a score of 23% of the maximum– it is an outlier stock that particularly excels in terms of the Ratio metric, but also has merits in the metrics signified by the axes nearby.

While the biplot does a good job in conveying the distribution of the stock data and the relationships between the metrics, the overplotting of axes leads to clutter which lowers the explanatory value of the plot with regards to a score's reason. For this purpose, we designed the GBC display as a complementary tool.

2.2 Generalized Barycentric Coordinates (GBC) Plot

Our GBC (see Fig. 2) interface consists of an equilateral polygon where the vertices correspond to the metrics and are rendered as filled red circles. It is somewhat similar to RadViz [1], but in place of springs it uses generalized barycentric coordinates [3] to determine the point locations. The order of the vertices is arranged through an approximate Traveling Salesman Problem solver that operates on the matrix of pairwise correlations among the metrics [5]. Arranging higher correlated vertices close by ensures a more accurate GBC mapping [4]. We see in Fig. 2 that the GBC plot's vertices 'pull' the stocks to the metrics that dominate their scores. We again highlighted a stock – here with symbol KDN and score 23%. Compared to the biplot it is much easier to determine the stock's strongest metrics as the vertices are well separated.

However, we still notice an overplotting of the stocks – most of the points are close to each other. This makes it difficult for users to label the points and so track individual stocks. In order to solve this problem, we apply a nonlinear mapping to the interior layout. The gray circles in the Fig. 2 show the contours of the mapping.

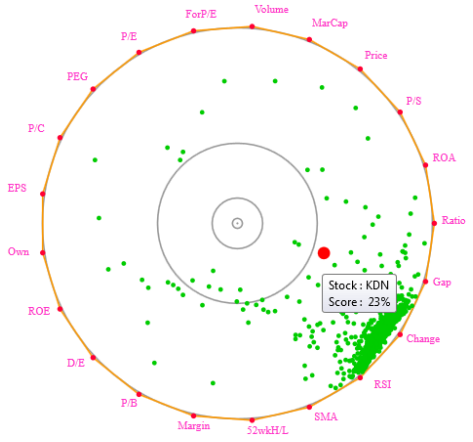

Figure 2: Data with all metrics plot in the GBC plot.

3 Score Function

As mentioned, scoring functions are often used to rate and rank stocks. Our interface supports this important functionality.

3.1 Building the Scoring Function

Each metric has a specific evaluation criterion [2] and we use the following function to shape the scoring function.

$$s = e^{-\alpha(x-0.5)^2} \quad (1)$$

$$s = \frac{\arctan\left(\beta\left(\pi x - \frac{\pi}{2}\right)\right) + \frac{\pi}{2}}{\pi} \quad (2)$$

where x is the metric value normalized to [0,1], α and β are parameters controlling the curve's trend according to the metric's sharpness, and s is the resulting score. Fig. 3 presents examples.

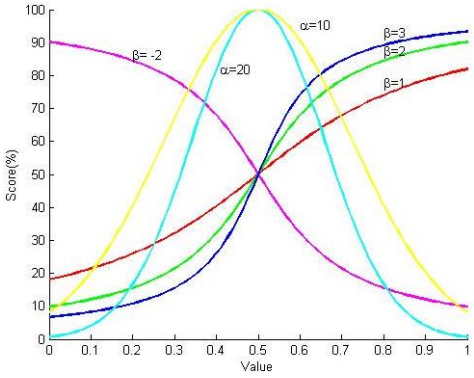

Figure 3: Score functions for a set of metrics

Investors have different preferences for these numerous metrics. We obtain the total score S_i for the ith stock by a weighted average of scores from all metrics as follows:

$$S_i = \frac{\sum_{j=1}^{n} w_j s_{ij}}{\sum_{j=1}^{n} w_j} \quad (2)$$

where w_j is the weight of metric j and s_{ij} is the score of stock i.

3.2 Selecting the Top Stocks

To give users insight into the distribution of the stock scores we provide a probability density function (shown below the GBC plots in Figs. 4, 5). Users can filter out the lower-scoring stocks by moving the vertical bar to the right – the number in the center is the probability of the yellow area of top-scoring stocks. Fig. 4 shows the top 25.33% of stocks. We typically show both biplot and GBC plots together for their complementary value.

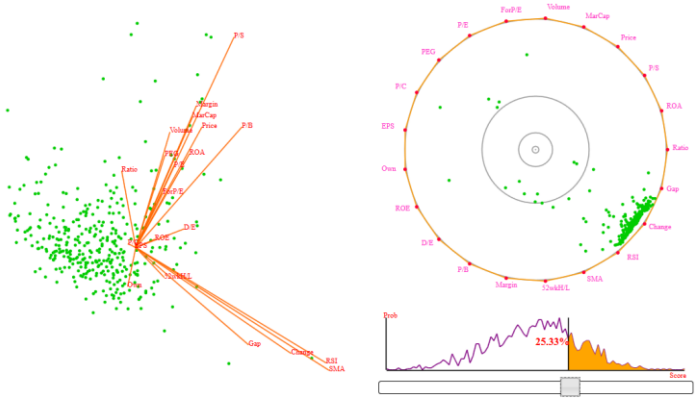

Figure 4: Selecting the top percent stock.

3.3 Adjust the weights

Our interface that allows users to tune the weights of the metrics is shown in Fig. 5. The sizes of the vertices correspond to the weights of the metrics for the overall scoring function.

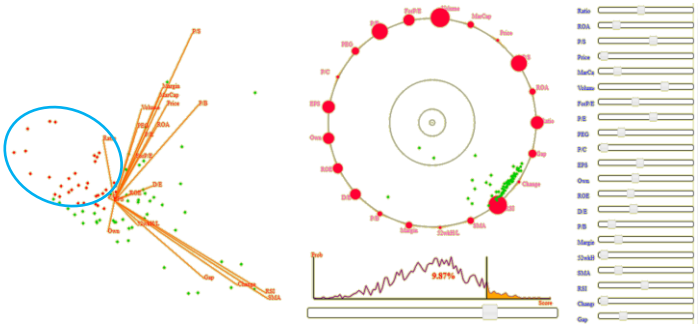

Figure 5: Adjusting the weights.

This figure also highlights the different strengths of biplot and GBC. While the GBC can show which metrics have particular strengths – here the majority of top stocks tend to be biased towards good scores in RSI, Change, and Gap –, the biplot can show the actual values of these metrics – here the values are not that high, especially not for those stocks in the blue circle. (Please note that the origin of the coordinate system has been moved to the center of the display – no metric produces negative values).

References

[1] C. Brunsdon, A. Fotheringham, M. Charlton., "An Investigation of Methods for Visualising Highly Multivariate Datasets," *Case Studies of Visualization in the Social Sciences*. pp. 55–80, 1998

[2] J. Kelly. *The Neatest Little Guide to Stock Market Investing.* Plume, ISBN-13: 978-0452298620, December 24, 2012.

[3] M. Meyer, H. Lee, A. Barr, M. Desbrun, "Generalized Barycentric Coordinates on Irregular Polygons", *Graph. Tools*, 1086-7651, 2002.

[4] J. Nam, K. Mueller, "TripAdvisor^{N-D}: A Tourism-Inspired High-Dimensional Space Exploration Framework with Overview and Detail," *IEEE Trans. Vis. Comput. Graph*. 19(2): 291-305, 2013.

[5] Z. Zhang, K. McDonnell, K. Mueller, "A Network-Based Interface for the Exploration of High-Dimensional Data Spaces." *IEEE Pacific Visualization Symposium,* 17-24, 2012.