

VISUALIZING METHOD FOR COMPARING MULTIPLE STOCKS

Daiki ISHIKAWA[†]

Takafumi SAITO^{††}

^{† ††}Tokyo University of Agriculture and Technology

ABSTRACT

The goal of this study is to propose a visualizing method for comparing multiple stock data. We suppose users of this method are general stock traders who want to analyze stock market. Though comparing multiple stock data is useful for stock trading, existing tools are unsuitable for comparison because they prioritize analyzing for one brand. Existing visualizing methods for large scale time-series data are unsuitable for stock data because there are characteristic that is trend of change of stock price is more important than amount of price in analyzing stock data. In this study, we propose visualizing method by combining line graph, heat map, and Similarity-Category Map (SCM) and aim to visualizing stock market, about 1800 brands in Tokyo Stock Exchange First Section.

1. INTRODUCTION

There is much information for making a decision in stock trading, such as change of stock price, financial information of companies, multiple indexes and so on. Comparing multiple stocks is also useful for stock trading. It provides traders information of trend of market, trend of industries and discovery of characteristic brands. However existing visualizing methods are only suitable for analyzing one stock and not suitable for comparing stocks because of their complex form.

2. RELATED STUDY

There are many preceding study about visualizing stock price and large amount of time series data. For example, Wattenberg at el. visualized stock market by improved TreeMap[1], Ziegler at el. used line graph and clustering by similarity of value change to analyze market[2], Simunić used hierarchical clustering and he decide representative stock by self-organizing Kohonen feature map[3], and Yagi at el proposed method to overcome weakness of line graphss[4].

2.1. Visualizing stock market by Treemap

Wattenberg et al. proposed to visualize stock market by improved Treemap. All brands are hierarchically clustered by using their similarity of stock prices. The

area of each rectangle indicates market capitalization, and color shows comparison with previous day, and their distance shows similarity of brands. This method shows similarity of brands and trend of market clearly. Moreover it can visualize 500 over brands. However time series information of stock price is missed.

2.2. Line graph

Line graph is suitable method for visualizing time series data, but it is difficult to understand graph when the number of data increase. Yagi et al. overcome this weakness by generating clusters and visualizing only representative of each cluster. This method can control level of detail of graph interactivity by controlling scale of cluster. However this method is not effective in comparing stock data because difference of each stock price is large, so cluster is not generated well. In addition, it is difficult to understand information of each brand.

2.3. Analysis of market sectors using cluster-based similarities of time series data.

Ziegler at el. Proposed two visualizing methods to analyze large amount of time series data. First method uses bar chart it expresses value of relative change from day to day by color. This method shows when value changed violently. Another method uses line graph and clustering by similarity of value change. The cluster centroids are visualized largely by line graph, and stocks belong to the clusters are visualized small by line graph under the cluster centroids. Besides, stocks are colored according to type of industry.

3. HIERARCHICAL CLUSTERING

Our proposed method has two steps. First step is to generate hierarchical clusters by dissimilarity of change of stock price. Then, stock data is visualized by SCM, heat map and line graph.

We don't apply clustering only to brands of same industry but to all brands in market. We consider it enables traders to analyze trend of market, similarity of industries and how much each industry is affected by economy and social event.

3.1. Dissimilarity

Preceding method proposed by Wattenberg et al. also use hierarchical clustering but they calculated dissimilarity from difference of stock prices. Generally, trend of stock price change is more important than absolute value in analyzing stock data, so we calculate dissimilarity from rate of change of stock price.

3.1.1. Logarithm of rate of change

Stock price change is usually evaluated as the percentage of its difference. However, it is not suitable for dissimilarity calculation, because rise and fall cannot be expressed symmetrically. In this paper, we defined symmetric rise and fall is that their rates of change is reciprocal. Figure 1 shows line graph of three data, symmetric rising stock A and falling stock C and no change stock B. Table 1 shows rate of change of three data. It is desirable that dissimilarity of A and B and dissimilarity of C and B are equal, so we took logarithm of rate of change and define sum of square of it during the period is dissimilarity.

3.1.2. Smoothing

Dissimilarity is sometimes undesirable value when the period is long. Desired dissimilarity is dissimilarity during whole period but calculated dissimilarity is sum of square of rate of change between two days. Therefore, stock price data is smoothed by one fifth days of period.

3.1.3 Ward method

Brands are clustered by ward method. Reason to use ward method is stability. This result is used to SCM display.

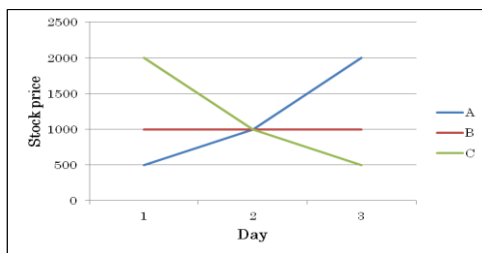


Figure.1 Symmetric rise and fall

4. VISUALIZING METHOD

In proposal method, three visualizing method, Similarity-Category Map, heat map and line graph, is combined and improved to compare multiple stock data.

4.1. Similarity-Category Map

We propose SCM. It has two dimensions, cluster and category, and brands are placed in according to them. We express brands as rectangles. Brands are clustered by similarity of stock price change and each cluster placed vertically. The order of clusters follows the order of dendrogram. In addition, categories, type of industry, are placed horizontally. In this work, the order of categories follows number of type of industry. Brands in same block are same cluster and same type of industry. Brands are places in each block by top alignment. The order of placing brands also follows the order of dendrogram. The number of columns of each type of industry is different. These are minimum values make height of each type of industry shorter than the specified height. At the end, coordinate of each block are moved as height of blocks of same cluster is as equal as possible. Then, this method shows trend of each cluster and each type of industry clearly.

Vertical length of each rectangle is proportional to variation amount of stock price. The more violently stock price change, the more noticeable due to this. Width of each rectangle is fixed for the sake of comparing stock price. Figure 2 is an example of SCM display

Table.1 Rate of change of Figure.1

	Rate of change		Logarithm of rate of change	
	Day 1-2	Day 2-3	Day 1-2	Day 2-3
A	+100%	+100%	log2	log2
B	±0%	±0%	0	0
C	-50%	-50%	-log2	-log2

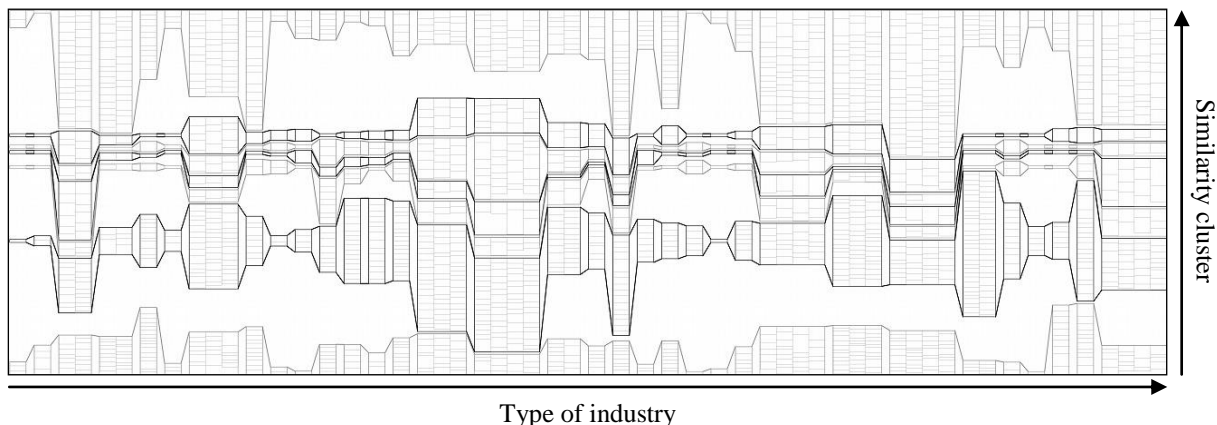


Figure.2 Example of SCM display

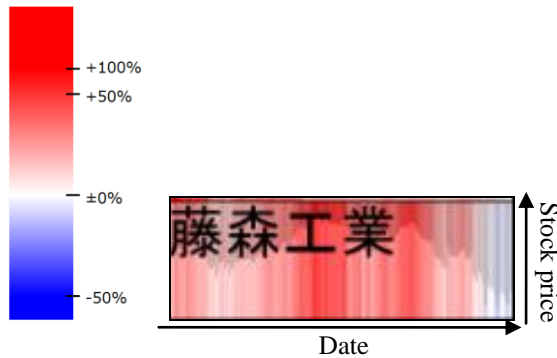


Figure.3 Color scale Figure.4 Example of each brand

4.2. Heat map

Heat map can show rough change of stock price. Method proposed by Wattenberg et al. define comparison with the previous day is expressed by color of each rectangle, but it is expressed by one color, so time series information of stock price is missed. Our proposal method expresses change of stock price by change of color. Each stock price data is normalized by closing price and color is decided according to color scale, Figure 3.

4.3. Line graph

Line graph is suitable for visualizing time scale data and can show time scale data in more detail than heat map. In proposal method, line graph is drawn inside of rectangles. On the upper side of heat map, we lower the value of color than the lower side of heat map so that shape of the border of color becomes line graph. This method is easier to understand time scale information than draw a line. Figure 4 is an example of it

5. RESULT AND STUDY

We applied proposal method to two datasets. One is 252 brands of vehicle and chemical between 12 workdays during appreciation of the yen. Another is 1764 brands in Tokyo Stock Exchange First Section between 833 workdays, about 3years.

5.1. Vehicle and chemical during appreciation of the yen

Vehicle and chemical is almost export industry and import industry. In general, export industry is in bad condition and import industry is in good condition during appreciation of the yen. However as Figure 5 shows stock prices of brands of two industry types change similarly. Pulp and paper is especially similar to vehicle. In this dataset, proposal method can show partly information to compare brands.

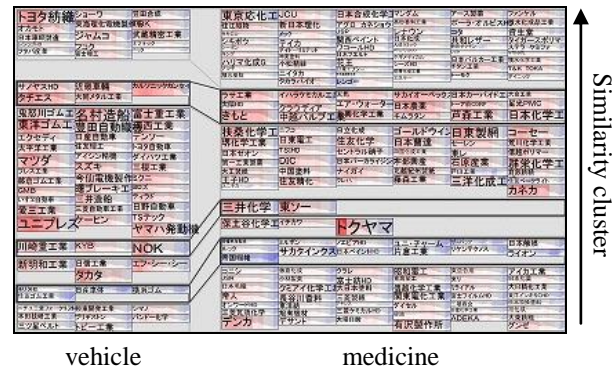


Figure.5 Vehicle and medicine

5.2. Tokyo Stock Exchange First Section

This dataset is one of the goals in this study. However Figure 6 shows it is difficult to understand this visualizing result. Each rectangle is too small and color cannot express change of each stock price. We can understand only rough trends of clusters, but this is not enough information to compare brands.

6. IMPROVE VISIBILITY

Figure 6 shows if all of brands are visualized, it is difficult to read graph. Therefore we integrate rectangles to improve visibility and if user wants to analyze in detail, chosen rectangles are expanded partly.

6.1. Integrating rectangles

If the number of rectangles included a block, divided space by cluster and type of industry, is N ($N > 1$), we integrate rectangles as far as the number of rectangles becomes $\log_2 N$ (truncated after the decimal point). Integrated rectangles are most similar two brands in the block. Figure 7 is result of visualizing whole of stock market.

6.2. Expanding rectangles partly

Integrating rectangles improve visibility, but information of each brand is lost. Therefore if user wants information of each brand, user can expand rectangles partly. Figure 8 is result of expanding one cluster, and Figure 9 is result of expanding two type of industry.

7. CONCLUSION

We propose visualizing method to compare multiple brands. This method contains combined and SCM, heat map, line graph. We applied this method two datasets. In the case of visualizing 252 brands, this method enable user to compare brands and presents some information to analyze. However, in the case of visualizing 1764 brands, it is difficult to understand visualizing result.

Proposal method is not suitable for comparing a large amount of brands from the above result. Therefore we integrate rectangles to improve visibility and enable user to expand rectangles partly. We consider future subject is to evaluate this method by comparing existing method.

8. REFERENCES

[1] Martin Wattenberg, Dow Jones, "Visualizing the Stock Market", CHI '99 Extended Abstracts on Conference on Human Factors in Computing Systems, pp.188-189, May 1999

[2] Hartmut Ziegler, Marco Jenny, Tino Gruse, Daniel A. Keim, "Visual Market Sector Analysis for Financial Time Series Data", 2010 IEEE Symposium on Visual Analytics Science and Technology, pp.83-90. October 2010.
[3] Krešimir Šimunić, "visualization of Stock Market Charts", WSCG '2003: Posters: The 11-th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision 2003, pp.129-132, February 2003.
[4] Sayaka Yagi, Yumiko Uchida, Takayuki Itoh "A Polyline-based Visualization Technique for Tagged Time-varying Data", 2012 16th International Conference on Information Visualisation, pp.106-111. July 2012.

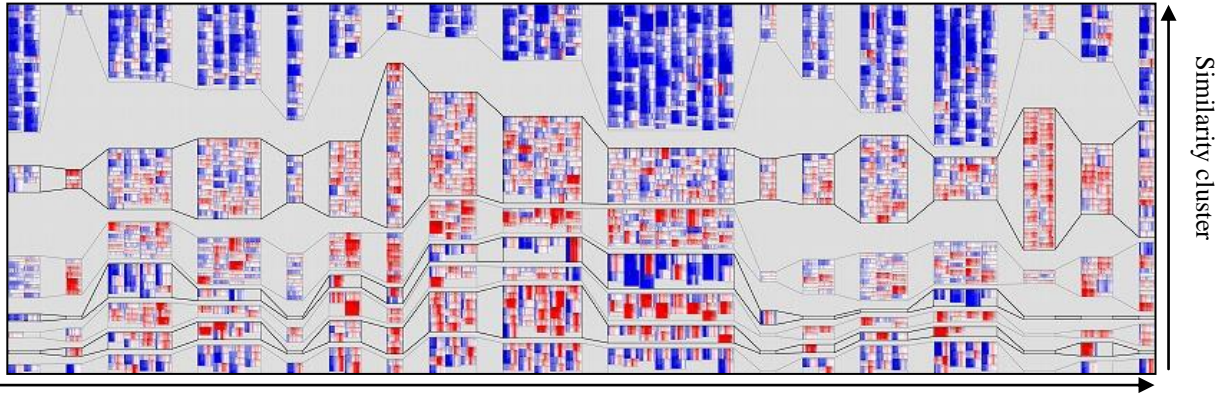


Figure.6 Visualizing whole of stock market

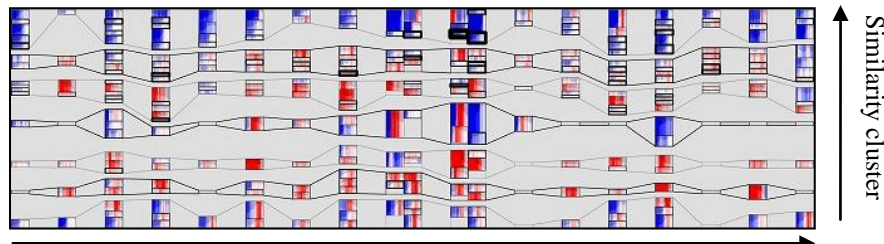


Figure.7 Integrating rectangles

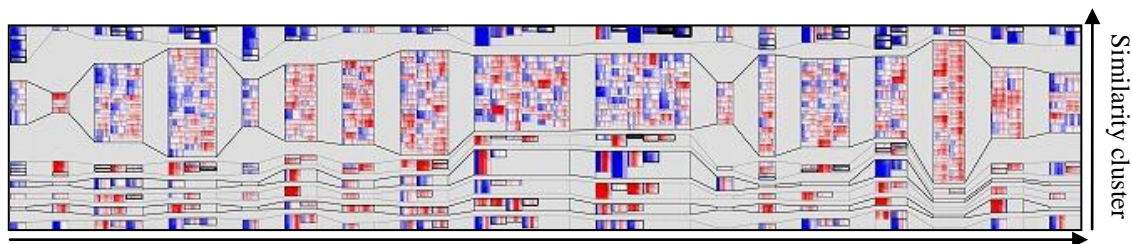


Figure.8 Expanding one cluster

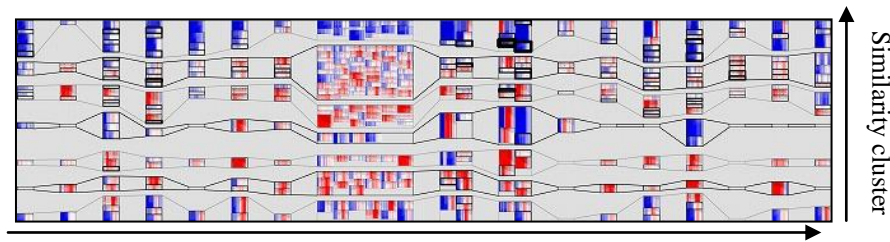


Figure.9 Expanding two types of industry