DataWear: Revealing Trends of Dynamic Data in Visualizations

Chris North, Umer Farooq, and Dilshad Akhter
Department of Computer Science
Virginia Tech
Blacksburg, VA 24061
{north, ufarooq, dakhter}@vt.edu

Abstract

DataWear enables the visualization of trends of change in dynamic data in a visualization without sacrificing a visual axis to represent time. DataWear represents data changes, and hence changes in the visualization of the current state of data, as wear marks in the visualization background. Worn paths represent common changes in data, and fading represents recentness of changes. The DataWear concept is demonstrated with a scatterplot visualization tool showing dynamic stock market data.

1. Introduction

Dynamic data is data that incurs changes or updates over time. For example, a database of stock market data might update each day to capture new stock prices and numbers of transactions. Since previous states of data may be important, databases can be constructed to capture changes while maintaining previous values for future reference. For example, the stock market database might include an additional attribute for date.

Different approaches can be taken to visualize dynamic data depending on whether the temporal dimension (dynamics) of the data is important to the user's task. If data dynamics is not important, then the user is mainly interested in maximizing the visualization of the current state (or any single previous state) of the database at the expense of representing the time dimension. For example, the current state of the stock market might be visualized in a Spotfire [1] scatterplot of stocks as price vs. number of transactions. Most visualizations fall into this static-data category.

If data dynamics is important, then the visualization must be altered to explicitly represent time as a major axis. For example, a time series chart represents time on the x-axis (e.g. [2]). Users can then examine relationships between time and another attribute such as stock price on the y-axis. The drawback in this case is that time consumes a valuable dimension of the visualization space at the expense of other attributes in the data.

This paper is concerned with a task set that lies between these two cases, in which visualizing current state is the user's primary focus and data dynamics is of secondary importance. One approach to accomplish this is to animate the primary visualization to show an instant replay of the data changes. The drawback is that it is difficult for users to cognitively integrate the replay into a coherent understanding of data trends over time.

DataWear is a new method for visually revealing the trends of dynamic data within the context of the primary visualization of current state. The DataWear concept can be applied to many types of static-data visualizations to enable them to reveal dynamic-data trends.

2. Related Work

There has been some work on the display of paths of objects that move with time [3].

DataWear derives its concept from the Information Mural [4]. Information Murals enable visual overviews of large information spaces, even when the number of informational elements greatly outnumbers the available pixels. It creates a miniature version of the information space using grayscale shading and anti-aliased compression techniques.

The idea of computational wear in the domain of collaborative document processing has been described in [5]. *Edit wear* represents the edit history of the document while *read wear* represents its readership history. The wear is represented on the document's scroll bars. The geometry of the scroll bar is used to interpret the wear marks in relation to the structure of the document. This display technique reuses precious screen space and collocates information display with navigation control.

3. The DataWear Concept

The goal of DataWear is to reveal how the data has changed over time, but also to preserve the view of the primary visualization of the current state of the data. That is, users can maintain focus on the visualization of current state, and also see trends of past changes within the context of the same visualization. The visualization does not need to sacrifice a visual axis to represent the time dimension. Many standard static-data visualizations (e.g. Spotfire) can be augmented to show the effects of dynamic data by including DataWear in the background of the visualization.

DataWear introduces the concept of representing time in dynamic data as *wear* in a visualization. As time progresses, changes to the data cause data items to move within the visualization. These movements leave worn paths in the background of the visualization, like translucent shadows or ghosts. Hence, patterns of wear reveal dynamic data trends.

The metaphor is worn paths in grass. As people traverse a green grassy field, the surface tends to wear off gradually and reveals the soil on frequently traveled paths. But as time increases, the grass grows back to retain its original integrity.

DataWear reveals two major components of dynamics:

- Recentness, and
- Frequency (or accumulation).

More recent and more frequently traveled paths are considered to be more relevant or more important.

Recentness is revealed by varying color saturation: the older the data are, the less saturated shade is used to represent them. At the first instance of time, the data is represented with a high level of grayscale intensity, near white. As time increases and data points move to new locations, their paths are highlighted using different levels of grayscale – the latest paths being highlighted with higher grayscale intensities and previous paths with decreasing intensities. Therefore, trends in the data can be inferred by the relative levels of grayscale. The oldest data (starting from the first instance of time) blends in with the black background after a certain threshold time. This technique avoids obscurity with an increased range of time, while maintaining a worn away effect of the data. By using this technique, more recent data, which represents more relevant data, is obvious.

DataWear also addresses the problem of intersecting data paths. Intersecting paths are important because they represent the areas that are traveled most frequently. The intersected paths are distinguished from normal paths by higher levels of grayscale intensities. More frequently traveled areas accumulate brighter grayscales.

The DataWear concept is effective for visualizing dynamic data trends at two levels of scale:

- Micro level: movement pattern of an individual data point. Like a comet tail, the most recent portion of the path is the brightest. Older portions of the path wear away as the comet distances with time.
- Macro level: movement pattern of many data points over a long period of time. Like worn paths in a grassy field, the most popular routes and areas are evident. This effect is important for large scale data.

4. Visualization Examples

The implemented visualization is a simple scatterplot tool augmented with the DataWear concept. This example demonstrates how similar visualizations such as Spotfire can be augmented with the DataWear concept, and hints at how other types of visualizations might be similarly modified.

The visualization is based upon fabricated stock market data of 100 stocks consisting of four attributes: Share ID, share price, number of transactions (number of shares bought or sold), and time ranging from 1 to 50 days. Using our visualization tool on this sample of dynamic data, we attempt to reveal trends and show the effectiveness of the DataWear concept.

The initial visualization is a scatterplot with a black background, and options on the right side of the applet. It displays the first day's data by default. 100 spots appear in the scatterplot corresponding to 100 shares, with x-axis as price and y-axis as number of transactions. Figure 1 shows the state on day 15 along with the DataWear for days 1-15. The red spots represent the current state of the data points (day 15). The DataWear effect can be noticed since the previous paths of each data point is colored with low grayscale intensities and the end paths (near the red spot) are almost white.

Figure 1 reveals, for example, that number of transactions is more stable than price because the trend of dynamics tends to be horizontal in the scatterplot. Outliers that have moved more in the vertical direction can be seen, and there are obvious clear areas where no stock has tread.

The visualization tool provides other interactive features. If the user is interested in particular shares (the share ID is known), specifying the share ID will highlight those shares in green. Figure 1 shows shares 73, 37, and 50 highlighted. The green highlight color also varies in intensity from low to high corresponding to old and new data respectively. Moreover, the labels for share IDs are indicated on top of the starting point of the share on the first day, so the trend can be observed over the whole range of days.

Alternatively, if the user wishes to know the share ID of a particular share from the visualization, moving the mouse over a data point displays the respective share ID. Potentially, other detail information about the share could be displayed as well. For finding out specific values of shares, vertical and horizontal lines are drawn corresponding to values on the x and y-axes.

The effectiveness of this visualization is realized with a large increase in the number of days. Figure 2 depicts the visualization for day 46. The low grayscale intensity paths wear away and blend in with the black background, since these paths are less significant than the recent data. Different levels of grayscale intensities can be seen, and trends identified. For example, in the price range from

$81 to $99 and number of transactions from 10,500 to 13,375 shares (see top-right yellow bounding box in figure 2), it can be inferred that approximately eight shares are represented (eight red spots) in this range. Moreover, a lot of fluctuation has occurred in this range, which from a shareholder's point of view, might imply that the lower bound for price is approximately $81. That is, while these eight shares fluctuate between the prices of $81 and $99, they tend not to go below $81. Assuming that those shares do not incur heavy losses for that particular shareholder (since the lower bound of $81 is presumably bearable), these shares are not to be sold off.

Consider a novice speculator who wants to enter the share trading business. He wants to experiment with shares at a low risk, and thus wants to invest in shares whose prices are fluctuating less. Assuming that day 46 is the current instance of time (Figure 2), it is noticed that many transactions (buying and selling) are taking place with the share in the price range from $15 to $27 and approximately 10,000 shares transactions (see upper-left yellow bounding box in figure 2). Noticing that the share in this price range is being traded frequently and maintaining a stable price range, the novice spectator resolves to invest his money by buying that share.

Another feature of the visualization tool is the simulation capability that shows the gradual wearing of data. By specifying start and end day, the simulation will animate the data, first displaying the start day's data and then gradually moving on to the next day's data until end day is reached. By playing the simulation from day 1 to 46, users can see the wear patterns gradually grow. For example, in figure 2, simulation revealed that the shares in the price range from $99 to $105 and number of transactions from 5325 to 7050 (see lower-right yellow bounding box in figure 2) were originally in a lower price range, and with the increase in time from day 1 to 46, the price has increased with almost no change in the number of transactions. This may be vital information to the shareholder. Even though the prices are rising, the numbers of share transactions are stable, possibly indicating that shareholders are not selling the shares since the prices are expected to go even higher in future.

5. Discussion and Future Work

The visualization tool provides an initial framework for exemplifying the concept of DataWear. With its current features and advantages, there are a number of drawbacks and other important interactive features that need to be incorporated.

There may be a conflict in the DataWear representation of recentness and frequency. Since both are represented with higher levels of grayscale intensity, it could potentially create misinterpretations. One solution to this is to use two different colors for representing normal and intersecting paths. For example, normal paths could be represented in red intensities, and intersecting paths with blue intensities. As the intersection increases on a specific path, an increased blue intensity will result in a mixture with the red color, thereby giving a purplish effect. This would give recentness and frequency distinct color representations.

The time before a path wears away completely is currently fixed. A customizable wear threshold should be provided to the user for specifying the time duration for a normal path to completely wear away. A zoomable user interface [6] is also suggested that would give accurate and detailed information about the individual data points.

The visualization tool should also provide an interface for visualizing any sample of dynamic data using the DataWear concept. The user would connect to a database and provide the appropriate fields to be displayed, resulting in a customized visualization of the dynamic data trends.

6. Conclusion

This paper presents the concept of visualizing trends in dynamic data by using the wear effect within static-data visualizations. With the passage of time, more *recent* data and *frequently* traveled paths are highlighted using higher levels of grayscale intensities and older data faded away using lower levels that gradually blend in with the background. The presented example visualization tool is a demonstration of the DataWear concept in a scatterplot visualization. The concept can be used to augment many standard static-data visualizations for revealing the trends of dynamic data by showing the wear patterns in the background of the visualization.

7. References

[1] Ahlberg, C., Wistrand, E., "IVEE: An Information Visualization and Exploration Environment", *Proc. IEEE Information Visualization '95*, pp. 66-73, (1995).

[2] Chi, E., Pitkow, J., Mackinlay, J., Pirolli, P., Gossweiler, R., Card, S., "Visualizing the Evolution of Web Ecologies", *Proc. ACM CHI '98*, pp. 400-407, (1998).

[3] Andrienko, N. Andrienko, G., Gatalsky, P., "Supporting Visual Exploration of Object Movement", *Proc. AVI 2000*, pp. 217-220, (2000).

[4] Jerding, D. Stasko, J. "The Information Mural: A Technique for Displaying and Navigating Large Information Spaces", *Proc. IEEE Information Visualization '95*, pp. 43-50, (1995).

[5] Hill, W. C., Hollan, J. D., "Edit Wear and Read Wear", *Proc. ACM CHI '92*, pp. 3-9, (1992).

[6] Bederson, B., Meyer, J., Good, L., "Jazz: An Extensible Zoomable User Interface Graphics Toolkit in Java", *Proc. ACM UIST 2000*, pp. 171-180, (2000).

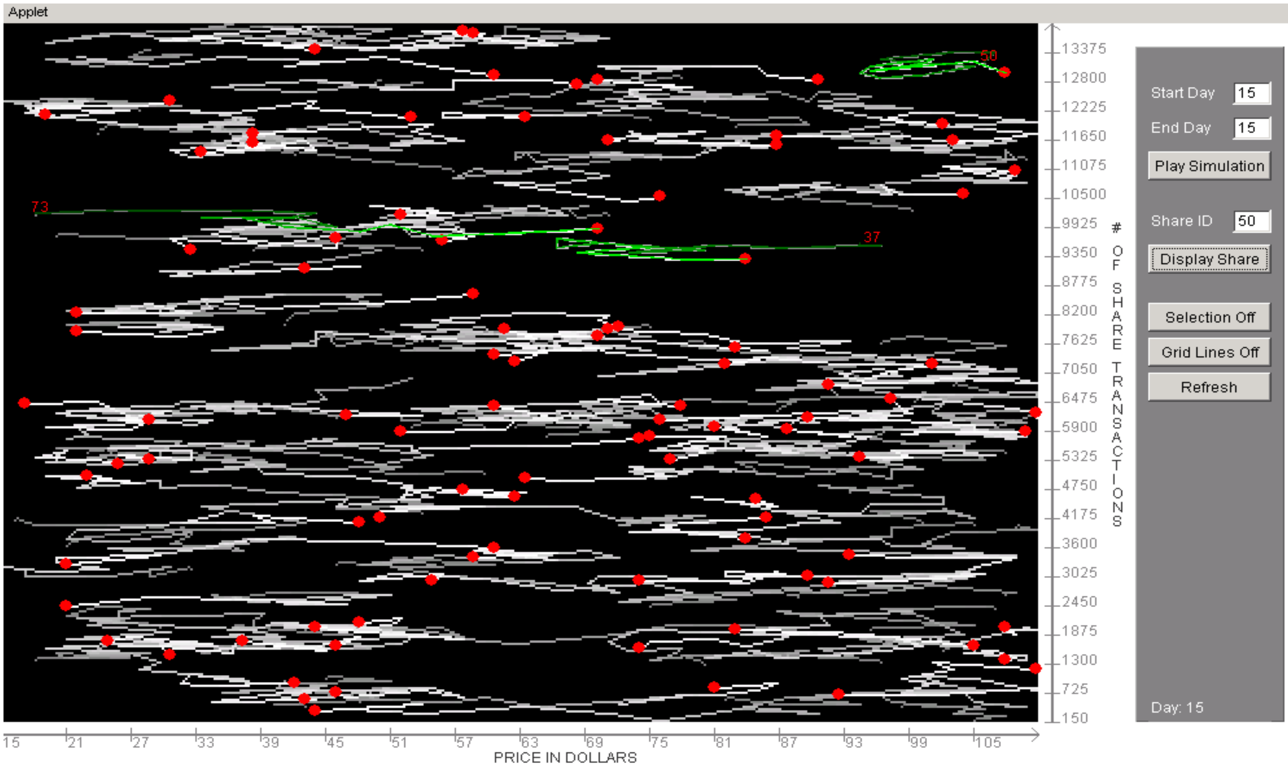

Figure 1. The DataWear effect of stock market data on day 15, with shares 73, 37, and 50 highlighted

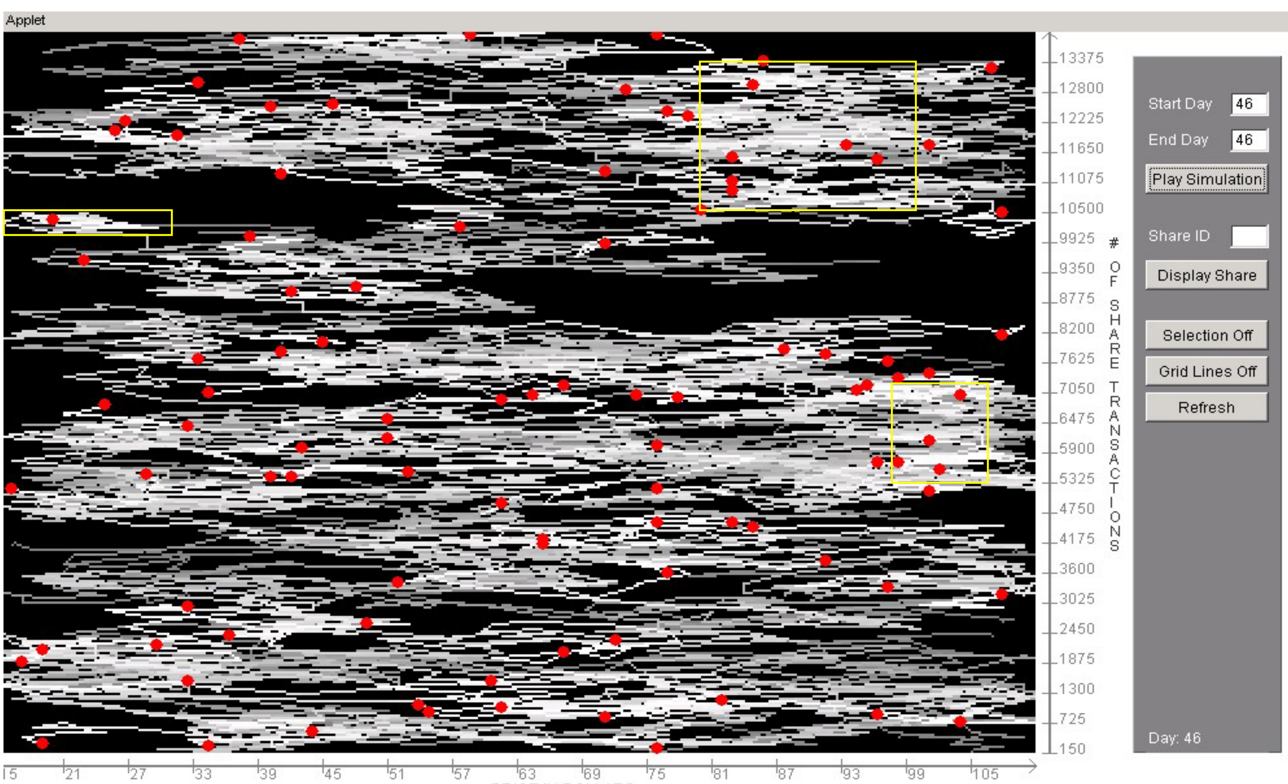

Figure 2. Trends revealed for stock market data on day 46