CHAPTER 7

A POLYLINE-BASED VISUALIZATION TECHNIQUE FOR TAGGED TIME-VARYING DATA

SAYAKA YAGI, YUMIKO UCHIDA,
TAKAYUKI ITOH

Abstract

We have various interesting time-varying data in our daily lives, such as weather data (e.g., temperature and air pressure) and stock prices. Such time-varying data is often associated with other information: for example, temperatures can be associated with weather, and stock prices can be associated with social or economic incidents. Meanwhile, we often draw large-scale time-varying data by multiple polylines in one space, to compare the time variation of multiple values. We think it should be interesting if such time-varying datasets are effectively visualized with their associated information. This chapter presents a technique for polyline-based visualization and level-of-detail control of tagged time-varying data. Supposing the associated information is attached as tags of the time-varying values, the technique generates clusters of the time-varying values, and selects representative values for each cluster, as a preprocessing. The technique then draws the representative values as polylines. It also provides a user interface so that users can interactively select interesting representatives, and explore the values which belong to the clusters of the representatives.

1. Introduction

"Big data" is a common term in the information technology industry in recent years. Various computer systems store huge information as various datasets: including computer system logs, sensor measurements, and transactions. These datasets expand from moment to moment, and therefore easily get huge, because they are time-varying. Information visualization is a useful and effective approach for understanding, analyzing, and monitoring of such time-varying datasets.

Polyline chart is one of the most popular representations for time-varying datasets. Newspapers and Web sites show various polyline charts: including stock prices, foreign currency, sports performance, and weather measurements such as temperatures. We also commonly draw multiple time-varying values in a single polyline chart space so that users can compare the time-varying values. For example, when we want to visually compare the temperatures at multiple locations, we would draw all of them in a single polyline chart space. On the other hand, we often deal with hundreds or even thousands of time-varying values in the above mentioned fields. We may want to visually compare the time variation of stock prices of many companies, or temperatures of several locations. It is usually difficult to understand if we draw hundreds or thousands of polylines in a single space.

Several recent works have addressed the visualization of such large-scale time-varying datasets. Wattenberg et al. presented a sketch-based query interface to search for specific shapes of polylines [1]. Hochheiser et al. presented Timeboxes and TimeSearcher [2], a gradient- and range-based query interface for polyline-based time-varying data visualization. Several works focused on similarity-based pattern and outlier discovery. Buono et al. presented a technique to interactively search for similar pattern [3] as an extension of TimeSearcher, and a similarity-based future pattern forecasting technique [4]. Lin et al. presented a technique to discover non-trivial patterns [5], by clustering a set of time-varying values and searching for outliers. Wang et al. presented a technique for important polyline selection [6]. Recently we presented two time-varying data visualization techniques, featuring sketch query on the clustered view [7], and pattern display on the heatmap [8].

Meanwhile, such time-varying data is often associated with other information: for example, temperatures can be associated with weather, and stock prices can be associated with social or economic incidents. Such information may be tightly correlated with time-varying values. For example, minimum temperature will be much lower in the morning of sunny days in the winter due to radiative cooling. On the other hand, difference between minimum and maximum temperature will be smaller during cloudy or rainy days. This kind of correlation between time-varying values and associated information can be useful for various purposes such as retrieval of frequent and outlier patterns. However, it is not always easy to determine the correlation between them. We believe it should be useful if time-varying data visualization techniques simultaneously display such associated information to assist the understanding of the correlation.

This chapter presents a new time-varying data visualization technique which assumes tags are assigned to the values of each time step of the time-varying data. We suppose that tags consist of a set of predefined terms: for example, sunny, cloudy, and rainy for weather data, or exercising, eating, or sleeping for health care data. As a preprocessing step, the technique clusters polylines based on their shapes and tags, and then selects representative polylines from the clusters. It realizes smooth level-of-detail control by interactively controlling the number of polylines to be displayed. The technique also features click and sketch interfaces so that users can interactively select particular polylines which are tagged by the user-interested terms.

This chapter presents the effectiveness of the presented technique with Japanese weather data recorded by AMeDAS (Automated Meteorological Data Acquisition System). The dataset consists of time-varying temperature values with weather tags including "Clear", "Sunny", "Cloudy", "Rainy", and "Snowy" at 376 observation points. We visually discovered different patterns of temperature variations from the visualization result.

2. Level-of-Detail Control for Time-Varying Data Visualization

This section briefly introduces a level-of-detail control and sketch interface [7] for time-varying data visualization previously we presented.
The technique supposes the following time series data, consisting of a set of values $\boldsymbol{P} = (p_1, p_2, ..., p_n)$ represented as n polylines. We describe the values of a polyline as $\boldsymbol{p_i} = (p_{i1}, p_{i2}, ..., p_{im})$; p_{ij} denotes the value at the j-th time of the i-th polyline. We draw the set of values as a polyline chart, while the horizontal axis denotes the 1st to the m-th time, and the vertical axis denote the magnitude of the values.
As a preprocessing, the technique temporarily quantizes polylines, generates clusters of them, and selects representative polylines from the clusters. While the quantization step, the technique generates a grid surrounding all polylines, and calculates intersections between the polylines and grid-lines. It then generates rough polylines by connecting the intersections, and uses them for the clustering. Number of clusters can

be controlled by the resolution of the grid as well as similarity threshold values, and our implementation prepares several clustering results so that the number of representative polylines smoothly varies.

The technique then initially displays representative polylines. Smoothly switches clustering results, it can seamlessly change the number of representative polylines to be displayed. The technique also provides a click interface, so that users can specify interesting representatives by directly clicking. It also provides a sketch interface, so that users can specify interesting representatives which have partial shapes similar to the sketched curves.

3. Extension to the Tagged Time-Varying Data Visualization

This section proposes an extended visualization technique for tagged time-varying data. This chapter extends the aforementioned time series data as follows: we describe the tags of the i-th polyline as $\mathbf{w}_i = (w_{i1}, w_{i2}, \ldots, w_{im})$; w_{ij} denotes the tag at the j-th time of the i-th polyline, as well as p_{ij} denotes the value at the j-th time of the i-th polyline.

3.1 Clustering and Representative Polyline Selection

The extended technique displays user-adjusted number of representative polylines to reduce the cluttering among the polylines and improve the comprehensibility. The technique clips the polylines by the intervals, and then generates clusters of clipped polylines for each interval, where clipped polylines in a cluster is similarly shaped and tagged.
The technique first generates a grid covering the drawing area, and then divides into $a \times b$ subspaces, as shown in Figure 1(a). Here this chapter formalizes the grid as follows: h_i is the i-th horizontal line of the grid $(0 \le i \le b)$, v_i is the i-th vertical line of the grid $(0 \le i \le a)$, t_i is the time at v_i, and b_i is the value at h_i.
The technique first samples P at t_0 to t_a, and temporarily quantizes the sampled values at b_0 to b_b. The technique then generates groups of polylines, if the polylines have the same quantized values both at t_{i-1} and t_i, as shown in Figure 1(b). It then clips polylines of a group by t_{i-1} and t_i, as shown in Figure 1(c), and generates clusters of the clipped polylines.

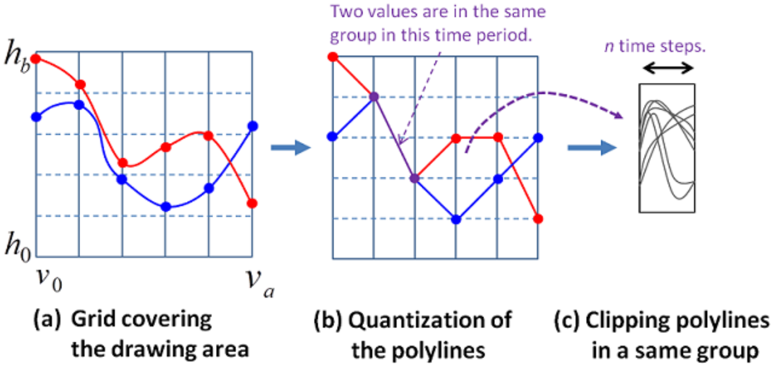

Fig. 1 Quantization and clipping of polylines.

The procedure for clustering of tagged polylines is as follows. Figure 2 illustrates the procedure, where the colors of the polylines denote their dominant tags. Here, the technique regards the clipped polylines as *n*-dimensional vectors, while they contain *n* time steps between t_{i-1} and t_i. This step firstly divides the clipped polylines according to tags, as shown in Figure 2 (b), using a dendrogram from the polylines constructed according to similarities of their *n*-dimensional vectors $(w_{jt_{i-1}}, \ldots, w_{jt_i})$. The technique then applies a non-hierarchical clustering (e.g. k-means) to the polylines in each cluster, using their *n*-dimensional vectors $(p_{jt_{i-1}}, \ldots, p_{jt_i})$. Here, the number of clusters is specified by users. Consequently, it generates clusters consisting of similarly tagged and shaped fragments of polylines, as shown in Figure 2(c).

Next, the technique selects representative polylines, for each cluster, as shown in Figure 2(d). Our current implementation simply extracts a polyline as the representative, which is the closest to the center of a cluster in a *n*-dimensional vector space. This strategy is basically effective because it selects average polylines. Here, if one or more polylines in a cluster have been already selected as the representative polylines of other clusters, the technique does not select any new representatives from the current cluster, so that we can reduce the total number of representative polylines.

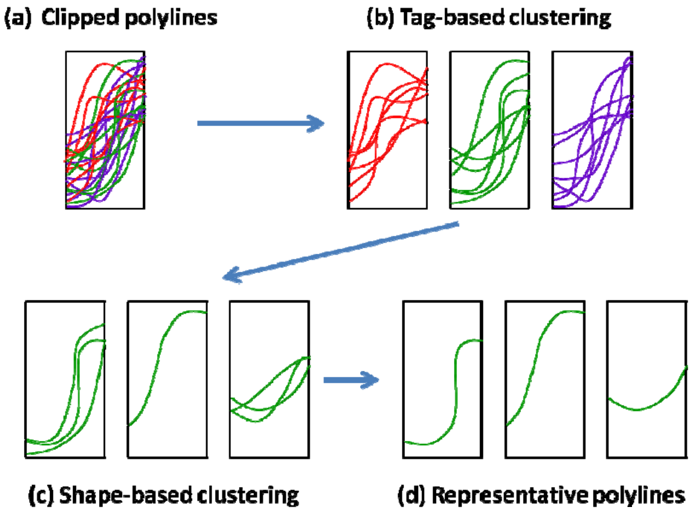

Fig. 2 Clustering of clipped polylines. It firstly clusters them based on their tags, and then clusters the polylines for each tag group based on their shapes. Finally, it selects representative polylines for each cluster.

3.2 Interactive Visualization

The extended technique represents the time-varying data as colored polylines. It assigns colors to the tags (e.g. blue to rainy, and green to cloudy, to the tags of weather data), and draws the polyline in the assigned colors. If the vertices of a segment of a polyline have different tags, it interpolates the colors along the segments. The technique also features selective polyline display based on the tags. Users can select particular tags so that the technique can draw only corresponding parts of the polylines. They can also select particular tags to be filtered from the display of the polylines.

Initially our technique draws only the representative polylines. Our current implementation generates several clustering results, with several configurations of the grid and the clustering process. Smoothly replacing the clustering results, our technique seamlessly displays several levels of numbers of representatives.

Here, the technique adjusts transparency of representative polylines in each time step. Here, a representative polyline is actually selected as a representative in several time steps, while it is not selected in other time steps. For example, Figure 3 shows that the polylines A, C, and D are selected as representatives, while A and C are actually selected as representative in the former two time steps, and A and D are selected in the last time step. In such time steps which a polyline is feasible to be selected as a representative, the technique draws the polyline in relatively lower transparency based on the number of original polylines in the cluster. Meanwhile, it draws a polyline in quite high transparency in the other time steps. Through this process, the technique can reduce the visual clutters while displaying consecutive polylines. Users can also observe how many original polylines are represented as the drawn polylines.

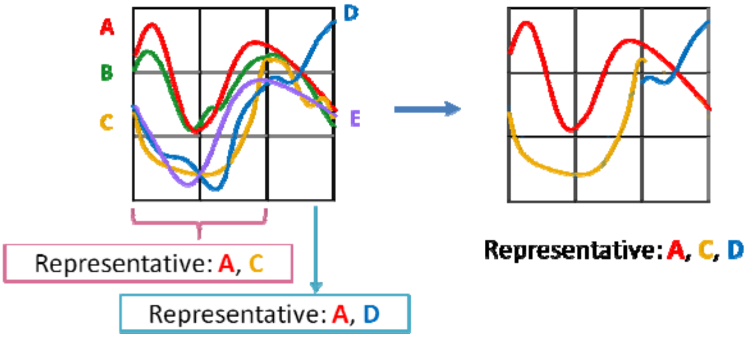

Fig. 3 The algorithm selects representative polylines for each interval. In this illustration, polylines A and C are selected at the former two intervals, while polylines A and D are selected at the last interval. Consequently, three polylines A, C, and D are selected as representatives.

The technique provides click and sketch interfaces, so that users can specify interesting representatives by directly clicking particular points or sketching particular shapes. When a user clicks a point on the display, the technique calculates distances between the point and all segments of the drawn polylines. If at least one of the segments of a polyline is enough close to the clicked point, the technique highlights the polyline. When a user draws a curve on the display, as shown in Figure 4, the technique samples several points on the curve, and calculates distances between the sampled points and all segments of the drawn polylines. If at least one of the segments of a polyline is enough close to each of the sampled points, the technique highlights the polyline.

While polyline reduction in our technique improves the comprehensibility of the data, users may want to look all the polylines that have the interested features. To satisfy such requirement, the technique can

reactivate the non-representative polylines, which belong to the clusters of the representative polylines specified by click/sketch operations.
Users can specify particular tags to be extracted by the above query operations. It can highlight only the parts of the polylines corresponding to the specified tags while the click or sketch operations. It can also reactivate only the parts of the non-representative polylines corresponding to the specified tags.

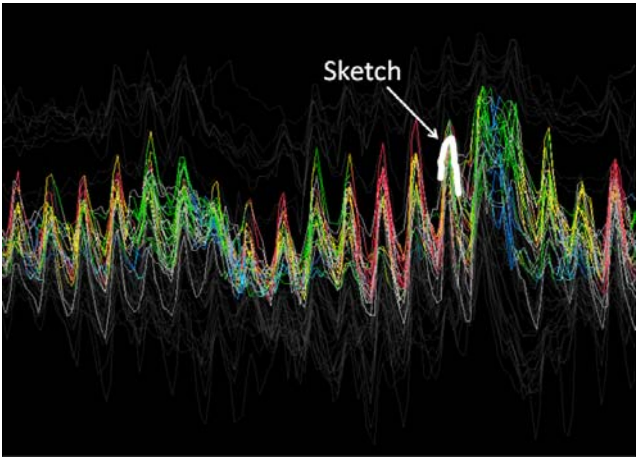

Fig. 4 This technique features a sketch interface. When a user sketches a curve on the screen, the technique highlights polylines close to the sketched curve. The interface can specify tags of polylines to be highlighted.

4. Examples

We developed the presented technique with JDK (Java Development Kit) 1.6, and executed on a personal computer (CPU 2.67GHz Dual Core, RAM 2.0GB) with Windows Vista (32bit). Processing time of the technique is estimated as follows. The time for level-of-detail control is $O(n(a+s))$; for quantization and clipping is $O(na)$ and for clustering is $O(ns)$, the time for sketch interface is $O(rq)$. Here, we formalize each variable as follows:

- n is the total number of polylines
- a is the number of sampled time steps
- s is the total number of time steps
- r is the number of representative polylines

- q is the number of sketched time steps

In our measurement, the average time for level-of-detail control was 177 milliseconds, and the average of sketch process was less than 1 millisecond, which n=376, a=10, s=240, r=249, and q=8. This result denotes that the technique is sufficiently fast for the interactive visualization.

We applied Japanese weather data recorded by AMeDAS (Automated Meteorological Data Acquisition System) to the presented technique. We extracted time-varying temperature data observed at 376 points around Japan in every 3 hours. We then assigned weather tags including "Clear", "Sunny", "Cloudy", "Rainy", and "Snowy" to temperature values at each timestamp.

Our implementation draws the segments of "Clear" polylines in red, "Sunny" in yellow, "Cloudy" in green, "Rainy" in blue, and "Snowy" in cyan, interpolating the colors if tags of two ends of a segment are different. Exceptionally it may draw the segments in gray if we cannot obtain the weather tags. While using click or sketch interfaces, our implementation draws selected polylines brightly, and others in gray.

4.1 Observation of Long-Term Trend

We visualized a dataset comprising temperature values during 1 to 31, July, in 2009. Figure 5 shows an overview of the temperature data. Figure 5(Upper) shows an overview without tags, and Figure 5(Center) shows an overview with tags. Here, the both show the results including all observation points before applying level-of-detail control. While Figure 5(Upper) displays just major and outlier variations of temperature, Figure 5(Center) looks much more informative. Colors in the dense parts denote major weather of the days. Blue ("Rainy") is observed in several consecutive days and it appears several times during the period. This result well demonstrates the effectiveness of the visualization of tagged time-varying data, while it is difficult to obtain such knowledge from Figure 5(Upper).

To observe the detailed variation, we applied level-of-detail control to the original polylines including all observation points. Figure 5(Lower) shows the results. In this figure, severe overlaps of polylines are reduced, especially non-tagged ones colored in gray. As mentioned above, we can observe that rain continued in many regions. We can also find only a few red or yellow ("Clear" or "Sunny") days, while there are about 10 to 20 sunny days in an average year as shown in Table 1. This result demonstrates that the presented technique is useful for observation of long term (e.g. one month) time variation of values and tags.

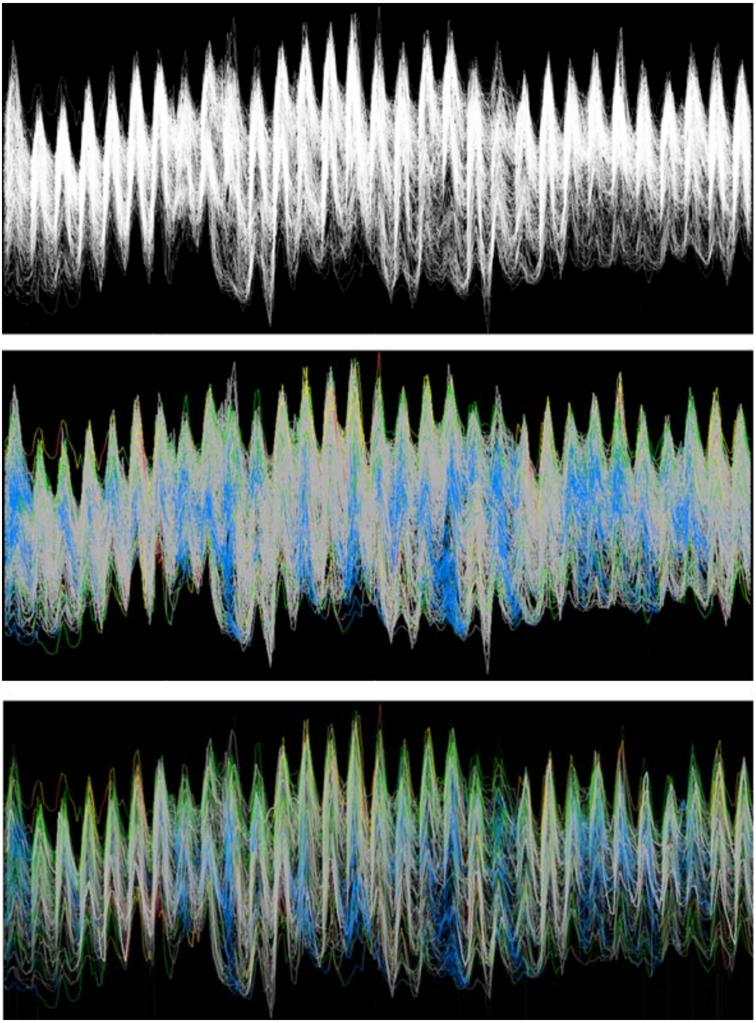

Fig. 5 (Upper) Overview without tags; (Center) Overview with tags (before applying level-of-detail control); (Lower) Overview with tags (after applying level-of-detail control).

Table. 1 Average number of sunny days in July (1981-2010 normal value).

Sapporo	Sendei	Niigata	Tokyo	Nagoya	Osaka
13.0	9.7	14.2	12.1	14.2	16.5

Hiroshima	Takamatsu	Fukuoka	Kagoshima	Naha
15.9	17.7	15.7	17.9	24.1

4.2 Observation with Interactive Extractions

We visualized a dataset which stores the weather and temperature during 1 to 31, December, 2009. Figure 6 shows a zoom up view of temperature variation during five days. Figure 6(a) shows the original view of the tagged temperature data before applying level-of-detail control, and Figure 6(b) shows after applying level-of-detail control with tag-based clustering. Many overlaps of polylines are observed in Figure 6(a), while Figure 6(b) represents the features of the data more clearly. In Figure 6(b), we can find that extremely lower temperature was often observed at "Clear" or "Sunny" points, which denotes typical radiative cooling. It is difficult to obtain such knowledge from the visualization results with shape-based clustering without considering tags, as shown in Figure 6(c). Here, representative polylines are selected based on only their shapes. In Figure 6(c), polylines which have extremely low temperature are drawn as gray, and missed important weather features as mentioned above. The result denotes that we can observe overall features without missing detailed features while reducing cluttering among polylines by the effort of level-of-detail control considering tags in the clustering step.
Figure 7 shows the visualization result applying the level-of-detail control with tag-based clustering. In the middle of the month indicated by a dotted white rectangle, we can observe that temperatures drawn as red or yellow polylines drastically decreased. While temperatures drawn as blue or cyan polylines slightly varied.

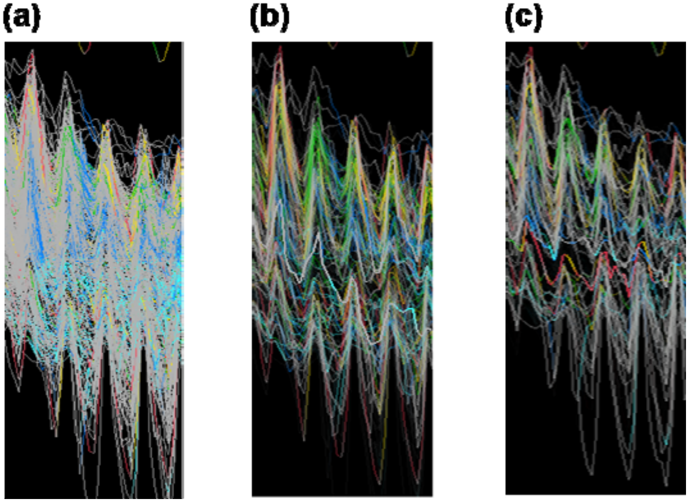

Fig. 6 Zoom up view: (a) without level-of-detail control; (b) with level-of-detail control (tag-based clustering); (c) with level-of-detail control (shape-based clustering).

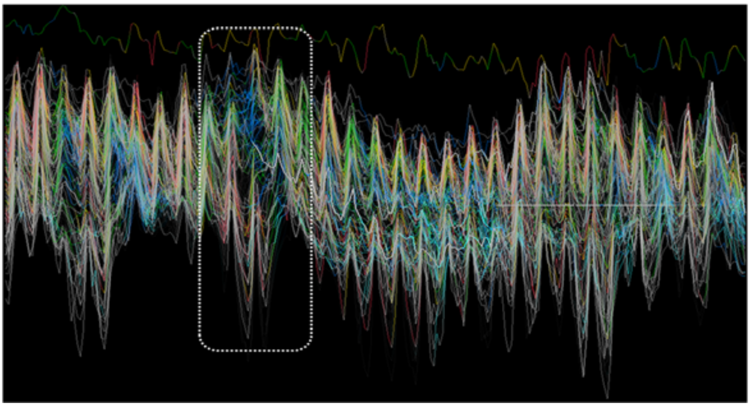

Fig. 7 Example applying the level-of-detail control with tag-based clustering.

To observe long-term variation of 2 patterns as mentioned above, we clicked the polylines. Figure 8(Upper) shows the result which we clicked relatively lower points while selecting two tags "Clear" and "Sunny"; consequently polylines tagged as "Clear" or "Sunny" at the clicked points were highlighted. It represents that range of temperature between the daytime and the night was relatively large. We can also find that extremely lower temperature was often observed at "Clear" or "Sunny" points. Figure 8(Lower) shows the result which we clicked relatively lower points while selecting two tags "Rainy" and "Snowy"; consequently polylines tagged as "Rainy" or "Snowy" at the clicked points were highlighted. It represents that range of temperature between the daytime and the night was relatively small, especially in regions tagged as “Rainy” or “Snowy” in the middle of the month. Also, it denotes that the average of temperature at “Rainy” or “Snowy” points was a little bit higher than temperatures of "Clear" and "Sunny". We can also observe that there were few sunny days through the periods.

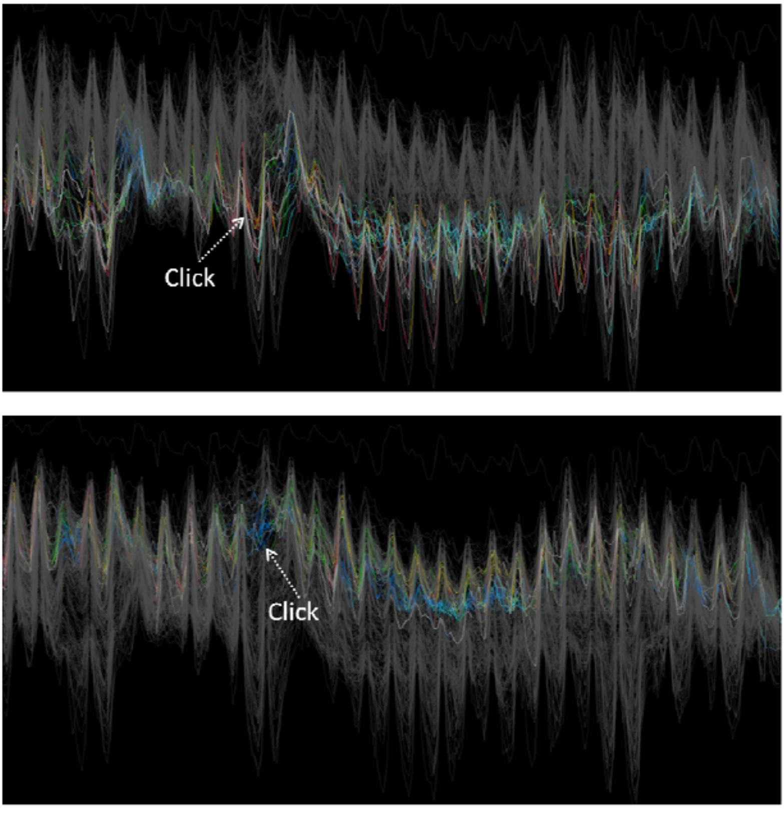

Fig. 8 Interactive selection: (Upper) Polylines tagged as "Clear" or "Sunny" at the clicked points are highlighted; (Lower) Polylines tagged as "Rainy" or "Snowy" at the clicked points are highlighted.

These results denote that we can discover interesting local features from large number of polylines, and observe the features by the effort of interactive extraction of polylines.

5. Conclusion

This chapter presented a polyline-based visualization technique for tagged time-varying data. The chapter first described the definition of the tagged time-varying data, and presented level-of-detail control and interactive polyline selection techniques on the top of the polyline-based time-varying data visualization technique. The chapter also demonstrated effectiveness of the technique by applying temperature data with weather tags. The level-of-detail control technique can adjust the number of displayed polylines without missing the important features of the datasets, and therefore it greatly improve the comprehensibility of the tagged time-varying datasets. Interactive polyline selection technique can assist users to focus on their interested portion of the datasets or filter unnecessary portions considering the tags. We believe that visualization of tagged time-varying data is still an open problem and this chapter presented a good solution for this problem.

Our potential future work includes the following:

[Many tags:] Since our current implementation represents tags as colors, it may be difficult to visually distinguish if we have many kinds (e.g. more than 10) of tags. We would like to discuss what kinds of visual metaphor can be more effective for the representation of many tags.

[Multiple tags at a point:] It is also difficult for our current implementation to represent if multiple tags are simultaneously assigned to a particular time of a particular polyline. Again, we would like to discuss what kinds of visual metaphor can be more effective for the representation of multiple tags.

[Observation of tag-change:] It is interesting for several kinds of data to observe how time-varying values vary when the assigned tags change. We would like to add features to the technique so that we can focus on time-varying value variations with particular patterns of tag changes.

[More applications and tests:] We would like to apply more various data to the technique, including medical measurement datasets, system measurement datasets, and stock price datasets. We would also like to have experiments for subjective and objective evaluations of the technique.

References

1. M. Wattenberg, D. Jones, Sketching a Graph to Query a Time-Series Database, SIGCHI Conference on Human Factors in Computing Systems Extended Abstract (CHI2001), 381-382, 2001.
2. H. Hochheiser, B. Shneiderman, Dynamic query tools for time series data sets: Timebox widgets for interactive exploration, Information Visualization, 3(1):1-18, 2004.
3. P. Buono, A. Aris, C. Plaisant, A. Khella, B. Shneiderman, Interactive Pattern Search in Time Series, Conference on Visualization and Data Analysis '05, 175-186, 2005.
4. P. Buono, C. Plaisant, A. Simeone, A. Aris, B. Shneiderman, G. Shmueli, W. Jank, Similarity-based forecasting with simultaneous previews: A river plot interface for time series forecasting, 11th International Conference on Information Visualisation, 191-196, 2007.
5. J. Lin, E. Keogh, S. Lonardi, Visualizing and Discovering Non-trivial Patterns in Large Time Series Databases, Information Visualization, 4(2), 61-82, 2005.
6. C. Wang, H. Yu, K.-L. Ma, Importance-driven time-varying data visualization, IEEE Transactions on Visualization and Computer Graphics, 14(6):1547-1554, 2008.
7. Y. Uchida, T. Itoh, A visualization and level-of-detail control technique for large scale time series data, 13th International Conference on Information Visualisation, 80-85, 2009.
8. M. Imoto, T. Itoh, A 3d visualization technique for large scale time-varying data, 14th International Conference on Information Visualisation, 17-22, 2010.

Index Entry

polyline
time-varying data
tagged time-varying data
clustering
level-of-detail control
representative polyline selection
click and sketch interfaces
AMeDAS