3D Time-Varying Data Visualization Technique Featuring Symbolic Aggregate Approximation Method

Maiko Imoto*
Ochanomizu University

Takayuki Itoh†
Ochanomizu University

1 INTRODUCTION

Observation and analysis of time-varying data is important in various scientific, technical, and social fields. We believe that information visualization techniques to observe and analyze such data should be useful. There have been several interactive techniques to easily extract interesting polylines from the large time-varying data. For example, Uchida et al. reported a technique as one of our approaches, which clusters the polylines and renders only representative polylines for each cluster [1].

We have presented a 3D time-varying data visualization technique [2]. The technique draws the time-varying data for all-in-one display and extracts similar values. It firstly orders the time-varying data based on their similarity, and places the polylines in the order. Then, it provides two views. The first view briefly represents the whole time-varying data, and the second view represents the partial data extracted by click operations. Users can look over the time-varying data, extract groups which have interesting similar values, and carefully observe the extracted values.

This poster presents an extension of our technique, which detects frequent or outlier patterns by applying SAX(Symbolic Aggregate approXimation) and indicates them. SAX is a useful technique for pattern discovery from time-varying data [3][4]. The technique divides time-varying values into a set of segments, replaces the segments into symbolic characters. Integrating SAX to our visualization technique makes users easily select or observe groups of the polylines which have typical frequent or outlier patterns.

2 3D TIME-VARYING DATA VISUALIZATION

2.1 Overview

Our visualization technique [2] places them onto 3D spaces (xyz-space). Figure 1(Left) shows the overview of data placement by the technique. The technique firstly orders the polylines based on the similarity of their feature values, and then assigns z-coordinate values to the polylines according to the order. The placement is useful to extract groups of similar polylines. It then provides two views, while ”Viewpoint 1” has a ray perpendicular to xz-plane, and ”Viewpoint 2” has a ray perpendicular to xy-plane. The first view provides an overview of the whole time-varying data and a user interface to extract interesting polylines, and the second view provides zoomed display of the extracted polylines.

The technique consists of the following three components: 1) order the polylines, 2) overview of the whole time-varying data in the left-side of the window, 3) display of the extracted polylines in the right-side of the window.

2.2 Ordering of the polylines

This section describes the j-th value at the i-th time as a_{ij}, and the j-th time series values as $A_j = \{a_{i1}, .., a_{in}\}$, while the data has n time steps. Here, the technique calculates distances between arbitrary pairs of polylines, and constructs a dendrogram of the polyline to order them. Current our implementation prepares several distance calculation schemes so that users can select one of them according to their needs. The poster introduces a scheme that calculates distances based on positions of polyline-vertices. The scheme calculates the distance between the j-th and k-th polylines $D(j,k)$.

Then, the technique constructs a dendrogram. Initially it treats a polyline as a cluster, and merges the closest clusters into a single cluster. The merge process is repeated until all the polylines are connected by a single dendrogram. Finally, it orders all the polylines by using the dendrogram.

2.3 Visualization

The technique prepares two viewpoints, which are used to global and local visualization of the time-varying data, as shown in Figure 1(Left), The technique firstly displays the whole data from ”Viewpoint 1”, in the left side of the window, as shown in Figure 1(Center)(1). Here all the polylines are drawn as ribbons, while they are shaded after setting several light sources. The polylines are colored according to y-coordinate values of the polyline-vertices: current our implementation assigns warm colors (e.g. red) to higher y-coordinate values, and cool colors (e.g. blue) to lower y-coordinate values. The technique initially adjusts ”Viewpoint 1” to display the whole data in one space; however, users can control the viewpoint and freely focus on specific parts of the data.

2.4 Extraction of similar polylines

The technique also provides a user interface to extract specific similar polylines on the left side of the window. Our implementation supports two-click operation to specify a range of being extracted polylines, as shown in Figure 1(Center)(2). When a user specifies a range, the technique renders the extracted polylines from ”Viewpoint 2” and displays on the right side of the window, as shown in Figure 1(Center)(3). Again, the technique initially adjusts the viewpoint to display the whole part of the extracted polylines; however, users can freely control the viewpoint to focus on specific parts of the extracted polylines.

3 INTEGRATION WITH SAX

The technique is generally useful to extract a set of totally similar time-varying values in large-scale datasets, because the polylines are ordered based on their global similarity. On the other hand, we are often interested in extraction of partially characteristic values. We apply SAX for symbolic character representation of time-varying values, which is useful for various purposes, for example, discovery of frequent or outlier patterns, and query of specific patterns. By representing the time-varying data in a string, we apply the string manipulation and language analysis in the field of natural language processing.

Figure 1(Right) is an illustration of our implementation of SAX. It assigns A, B, C, D, E, and F to the illustrated ranges, then divides each interval into smaller pieces by a smaller interval, calculates average values of the piece, and finally assigns characters to the pieces. Figure 1(Right) shows that our implementation divides the interval into six pieces and assigns characters CABCDE.

*e-mail: maiko_i@itolab.is.ocha.ac.jp
†e-mail:itot@is.ocha.ac.jp

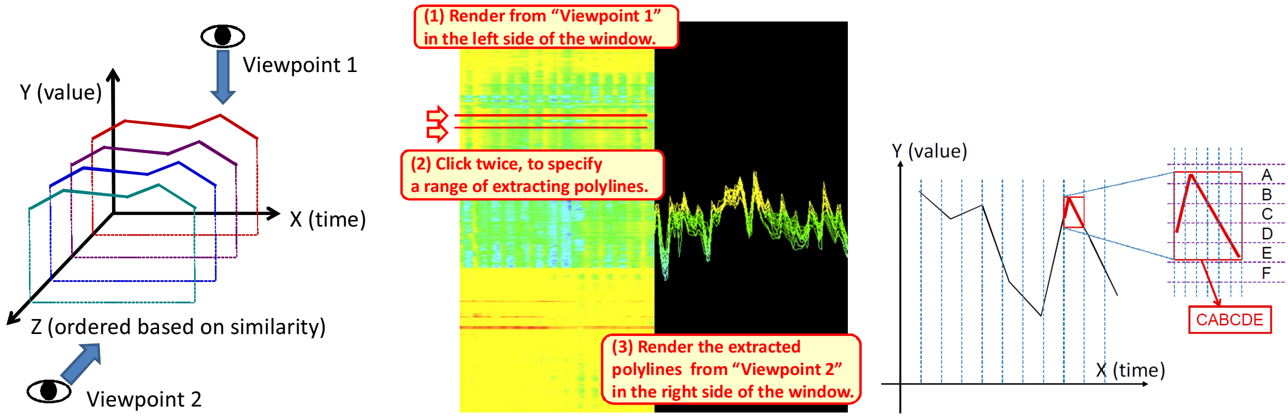

Figure 1: (Left) 3D placement of polylines. The technique orders the polylines based on their similarity, and assigns z-coordinate values according to the order. (Center) Processing flow of the presented technique. (Right) The pieces of time-varying values represent symbolic character by SAX.

We added a function to indicate user-specified patterns as rectangles in the left side of the window. Here, SAX may extract thousands of patterns from large time-varying data, and it is often difficult to extract interesting patterns from them. To solve the problem, our implementation provides an optional function to cluster patterns according to their similarity, and indicate the clusters as differently colored rectangle in the left side of the window.

4 Experiment

We applied the temperature data recorded by Japanese weather measurement system AMeDAS in January 2006 at 916 points in Japan. Figure 2 shows an example which indicates one of the frequent patterns by purple rectangles. We noticed that many polylines were enclosed by purple rectangles three times as marked by rectangles. So, we clicked two parts enclosed by blue and red rectangles in the left side of the window, and observed the polylines. We found that the polylines in blue did not range so much in the period marked by the first and third rectangle, though the polylines in red ranged so much at that time. This example donates that we discovered two groups of partially similar polylines which are separately placed in the left side of the window, and understood how the two groups were similar and different by the specific pattern indication. Figure 3 shows another example which several outlier clusters of patterns are indicated as differently colored rectangles in the left side of the window. Here, purple rectangles are concentrated on a specific day enclosed by a red rectangle. We selected the polylines and visualized in the right side of the window. As a result, we found that the temperature did not decrease on that day comparing with other days.

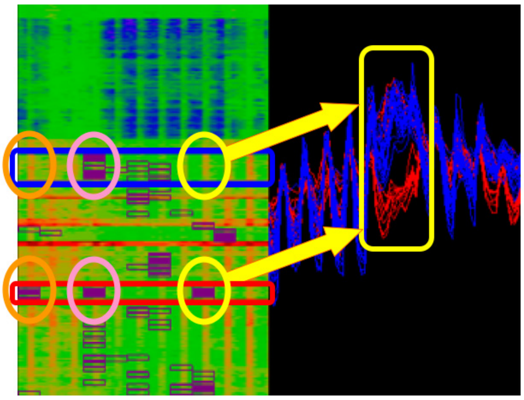

Figure 2: Example (1). Just one of the frequent patterns is indicated.

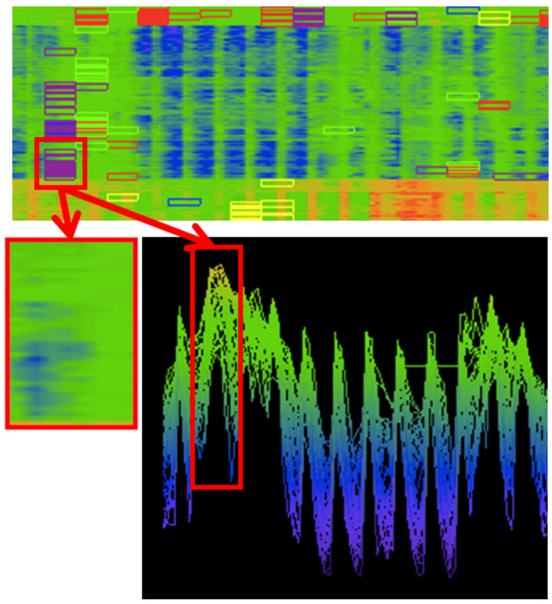

Figure 3: Example (2). Several colors of pattern clusters are indicated.

References

[1] Y. Uchida, T. Itoh, A Visualization and Level-of-Detail Control Technique for Large Scale Time Series Data, *13th International Conference on Information Visualisation (IV09)*, pp. 80-85, 2009.

[2] M. Imoto, T. Itoh, A 3D Visualization Technique for Large Scale Time-Varying Data, *14th International Conference on Information Visualisation (IV10)*, pp. 17-22, 2010.

[3] S. Gandhi, S. Nath, S .Suri, J .Lin, GAMPS: Compressing Multi Sensor Data by Grouping and Amplitude Scaling, *35th SIGMOD International Conference on Management of Data*, pp.771-784, 2009.

[4] J. Lin, E. Keogh, S. Lonardi, B. Chiu, A Symbolic Representation of Time Series, with Implications for Streaming Algorithms, *8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*, 2003.