BubbleUp: Toward Better Analysis for the Temporal Event Data

Wenjun Guo*
Ajou University

Seungwook Kim†
Seoul National University

Seongmin Mun‡
Ajou University

Kyungwon Lee§
Ajou University

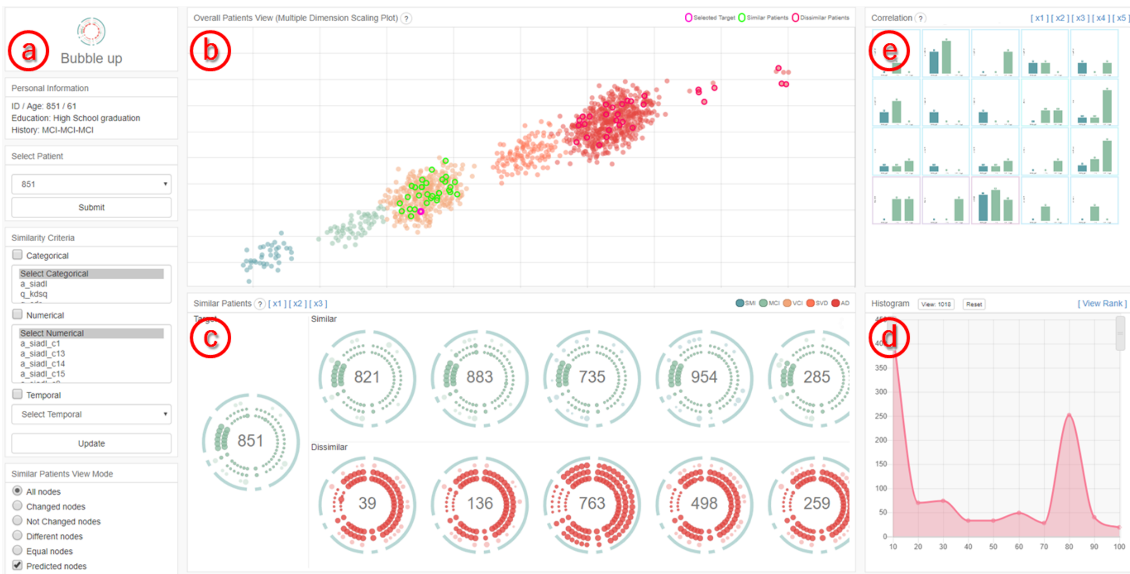

Figure 1: The overview of BubbleUp visualization system shows the diagnosis of multiple dementia patients. The overview on the left side (a) lets the user select the dementia patient by filtering methods. (b) shows the distribution and clustering of the whole data. (c) displays the details of the selected data. (d) shows similarity between selected data and other data. (e) predicts and shows the change of temporal event data.

Abstract

Temporal event data such as diagnosis records of multiple dementia patients include both multivariate data features and temporal changes. Analyzing temporal event data can reveal changing patterns by time or by correlations between target data and others. It is challenging but to explore visually both of multivariate data features and temporal changes in one view. In this study, we present a novel visualization system named BubbleUp which can explore temporal event data based on machine learning methods to confirm the correlations among the data and visualizes the changing patterns by time. The usage of BubbleUp visualization system can be divided into four steps; Overall distribution, detail view, correlation base on similarity, and prediction. We evaluate the usage and effectiveness of BubbleUp visualization system through the usage scenarios.

Keywords: Temporal event data, Data Clustering, Data Filtering, Interaction Design, Multiple Views, Task and Requirements Analysis

Index Terms: Computer Graphics [I.3.8]: Applications—Interactive Visualization ; Information interfaces and presentation (e.g.,HCI) [H.5.2]: User Interfaces—Web-based Interaction

*e-mail: xiaoxiaozhu524@ajou.ac.kr
†e-mail: ksw29zz@snu.ac.kr
‡e-mail: stat34@ajou.ac.kr
§e-mail: kwlee@ajou.ac.kr

1 Introduction

The main characteristic of temporal event data is that it is constantly generated and accumulated over the long period of time so that, the amount is very large and involves the pattern of changes over the time. And this changing pattern is very valuable as comparative analysis for the characteristic of the data base on similarity between target data and others. However, previous researches, which is using temporal event data, just have focused on the changing pattern for the characteristic of data and have expressed it, but have not provided similarities between data [3–5]. In this paper, we propose a visualization system for temporal event data to express the correlation data and to visualize the changing patterns by time. With our visualization, users can easily understand the distribution of whole data, can see the details of temporal event data, can analyze similarities between data, and can predict changes of the data.

2 Data Processing

In this section, we present a data processing structure in order to extract information from temporal event data. The data, which are used in the analysis, are the CREDOS (Clinical Research Center for Dementia of South Korea), a diagnostic data cohort of patients with dementia collected from 30 hospitals in Korea from 2005 to 2013 [2]. This data include 21,094 diagnostic records, which are the result of collection of diagnostic records of about 14,917 patients. Some of these data are diagnostic data of patients who visited a certain time period. The CREDOS dataset consists of demographic, basic characteristics, caregiver and patient information. This information dataset included 14 diagnostic categories which are covering 486 criteria [1]. Fig. 2 summarizes the architecture of our data processing, which is described in detail below.

Preprocessing The preprocessing step can be divided into three

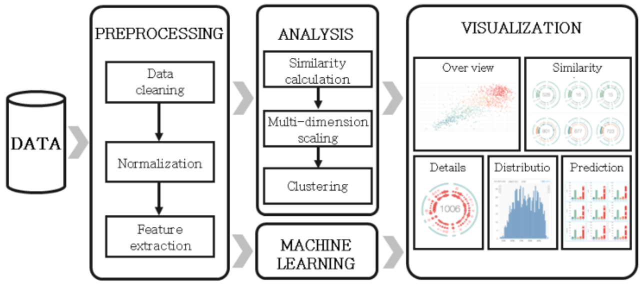

Figure 2: Data processing structure. Framework for the information acquisition from temporal event data.

steps; data cleaning, normalization, and feature extraction. Outliers were detected by the data cleaning. The missing values were handled. And then, we normalized the scale of the data and we used it for the analysis.

Analysis and Machine learning Through the above procedure, we obtained cleaned data from CREDOS. We analyzed these data by three different steps; similarity calculation, multi-dimension scaling, and clustering. In addition, we used a machine learning method; gradient boost, MLP, and SVM algorithms to predict the change of data.

Visualization We thus extracted correct information from CREDOS, one of temporal event data. The usage method for our visualization can be decided into four steps; overall distribution, detail view, correlation based on similarity, and prediction. We introduce how to use the BubbleUp visualization system through a usage scenario in the following session.

3 Usage Scenarios

We conducted usage scenarios to understand whether and how our visualization system was helpful to explore the temporal event data. Also we evaluate its effectiveness and usability. For the evaluation, we selected No.851 patient in the actual data. The following sections describe the requirements of the patient and the ways how to address these requirements with BubbleUp visualizations.

- **Question1: Can I identify my position within the whole group? Can I also identify the patient groups with similar tendencies and those with different tendencies?**

  Fig. 1 (b) shows the clustered group through K-means clustering to the user with MDS visualization. Through this view, the user can identify his or her location and identify the patient groups that are similar to him or her and those that are different.

- **Question2: Can I confirm the details of my diagnostic result and the temporal changes?**

  Fig. 3 shows the left part of Fig. 1 (c) in detail. The radius of the circle shows the diagnostic result according to time, and each node constituting the circle means the variable which is used in the test. Through this visualization, users can confirm the changes of disease and the difference according to the test variable over time.

- **Question3: Can I see the details of patients similar to me, the details of patients not similar to me, and their distribution?**

  On the upper-right side of Fig. 1 (c) shows the similar group with selected patient and on the bottom right side of Fig. 1 (c) shows the not similar group. Furthermore, through Fig. 1 (d), the user can identify the distribution of patients with a similar tendency to the selected patient.

- **Question4: Can I predict the test results I will receive in the next test?**

  In the right side visualization of Fig. 3, the final range of concentric circles represents the predicted value of the variable. If there is a change in the existing data, it is displayed in red. In Fig. 1 (e), the change in the size of the predicted data is shown in the bar graph. This allows users to check easily the predicted value of his or her test result.

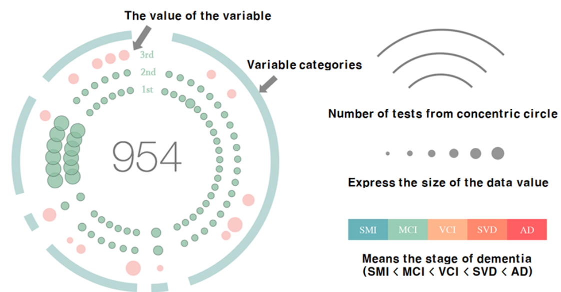

Figure 3: The visual expression shows both multivariate data features and temporal changes in one view.

4 Conclusion

In this study, BubbleUp visualization system is presented to express the characteristics of data that changes over time and similarity among data. The user can select the data what they want, confirm the distribution of whole data, see the details of the selected data, check the similarity between selected data and others, and predict the changes over time. The usage scenario presented previously explains the process by which a user searches for desired information through visualization. We have confirmed the effectiveness and usability of our visualization system through it. In the near future, we will conduct the expanded evaluation for the usability of the visualization and develop the visualization to be more useful.

Acknowledgments

This work was supported by the Ajou University research fund, the BK21 Plus project in 2018 and National Research Foundation of Korea (NRF-2015S1A5B6037107)

References

[1] S. Bang, S. Son, H. Roh, J. Lee, S. Bae, K. Lee, C. Hong, and H. Shin. Quad-phased data mining modeling for dementia diagnosis. *BMC Medical Informatics and Decision Making*, 17(1):60, May 2017. doi: 10.1186/s12911-017-0451-3

[2] S. H. Choi, S. Kim, S.-H. Han, D. L. Na, D.-K. Kim, H.-K. Cheong, J.-H. Lee, S. Y. Kim, C. H. Hong, S. Y. Moon, J. C. Kwon, J. E. Kim, J. H. Jeong, H. R. Na, K. R. Cha, S. W. Seo, Y. S. Shim, J.-Y. Lee, and K. W. Park. Neurologic signs in relation to cognitive function in subcortical ischemic vascular dementia: a credos (clinical research center for dementia of south korea) study. *Neurological Sciences*, 33(4):839–846, Aug 2012. doi: 10.1007/s10072-011-0845-4

[3] N. S. B. S. Fan Du, Catherine Plaisant. Eventaction: Visual analytics for temporal event sequence recommendation. *2016 IEEE Conference on Visual Analytics Science and Technology (VAST)*, 2016. doi: DOI: 10.1109/VAST.2016.7883512

[4] R. Z. D. G. H. Z. N. C. Shunan Guo, Ke Xu. Eventthread: Visual summarization and stage analysis of event sequence data. *IEEE Transactions on Visualization and Computer Graphics*, 24:56–65, Jan. 2018. doi: 10.1109/TVCG.2017.2745320

[5] J. Zhao, Z. Liu, M. Dontcheva, A. Hertzmann, and A. Wilson. Matrixwave: Visual comparison of event sequence data. In *CHI*, 2015.