VisXplore: Flexible Visualization System for Analyzing Complex Clinical Datasets

Filip Dabek[1] and Jesus J Caban[2]

[1] University of Maryland, UMBC
[2] National Intrepid Center of Excellence (NICoE), Walter Reed Bethesda

Abstract:
Assessing, exploring, and analyzing large amount of heterogeneous clinical data are some of the key challenges faced by researchers and clinicians interested in identifying complex clinical patterns. Flexible visualization systems capable of loading different data types and performing individual statistical analysis have the potential of impacting the way clinical datasets are analyzed. This paper introduces VisXplore -- a clinical data visualization system that has been designed to perform group or single-subject analysis of multivariate tabular, hierarchical, or temporal clinical data. The visualization system has been developed as a robust, flexible, and easy-to-use application that can be used by either researchers or clinicians. The framework has been tested with different clinical datasets including a dataset of patients diagnosed with post-traumatic stress disorder (PTSD) and complex comorbid conditions.

I. Introduction

Today, the assessment, evaluation, and treatment of any disease generates a large amount of data that needs to be analyzed. The large number of clinical elements and the vast amount of unstructured information being collected can quickly become overwhelming for clinicians and researchers alike. The complexity of analyzing multiple clinical elements simultaneously often causes that only a small subset of clinical measurements are used to assess the condition of a patient and/or to determine the effectiveness of a particular treatment. In order to successfully analyze a large number of multi-modal clinical variables and integrate different evaluation protocols, new visual analytics techniques and applications are needed.

This paper addresses the clear gap between the acquisition of clinical measurements and the diagnostic step, by providing an intuitive, flexible, and customizable interactive data visualization framework. Current analysis and visualizations techniques do not allow the user to be able to easily navigate through the data. The proposed framework is designed to help clinicians obtain new insights about the underlying conditions of the patients, analyze complex hidden clinical patterns, partition the data into different subgroups, visually explore the correlations between individual assessment techniques, and quickly analyze a particular patient against a specific cohort.

II. Previous Work

Data visualization systems are becoming increasingly important in numerous disciplines for data exploration, data mining, information retrieval, and hypothesis generation [1,2]. The increasing popularity of visual analytic tools has been driven by their effectiveness to enable users to detect patterns, correlations, and anomalies not obvious with other forms of data representations.

Over the last several years, multiple high-dimensional data visualization techniques have been proposed to enable interactive exploration of complex datasets. Generally, three of the main purposes of visual

analytic tools are to find averages of where most of the data lies, to find unique complex patterns, and to better understand outliers. During the last ten years, many graphing and plotting software applications have successfully been used to analyze complex datasets including Excel, SAS, SPSS, Tableau, Spotfire, and QlikView [3,4,5]. However, most of these applications are limited when trying to illustrate the relationship between more than three variables and are not optimized for clinical and healthcare applications. In particular, most of these visualization systems are limited in the amount of data that can be displayed at one time because of the constraints of physical viewing space and the two-dimensional characteristics of the plots.

III. Approach

We recognized that clinical data comes in different types and sources, to guarantee flexibility and acceptance across multiple clinical disciplines, VisXplore supports three of the most widely used types: *tabular, hierarchical,* and *temporal*. Once the data has been loaded into the system, the users can perform group or single subject analysis by employing any of the 15+ data visualization techniques and analytical modules that have been implemented. The group analysis focuses on the statistical properties of a specific population or cohort. In this type of analysis, the analytical techniques focus on the descriptive statistics of a data set, such as the average, standard deviation, variance, and correlations. The single-subject analysis focuses on comparing the properties of a particular subject against a particular population or cohort. This type of analysis is done by looking at the values against the z-score or normalized score to be able to see how the subject compares.

Figure 1 shows VisXplore, our visualization system designed to analyze and study complex clinical datasets. The system has been implemented using D3[6] and Extjs to allow the user to have an interactive experience inside of their current web browser without the need of installing any additional application. At the start of the application, the user is presented with a login which supports multiple users so that different researchers can use and analyze their own individual datasets. Upon login the user is presented with the ability to upload their tabular, temporal or hierarchical datasets using common file formats including CSV, Excel, and JSON. Once the dataset is loaded into the system, it becomes a report, allowing the user to have many different datasets / reports. When a specific report is selected, it launches the visualization tool and enables the different modules so the user can dive into exploring their data.

The main window of VisXplore has a section in the left side of the screen with two tabs: *the patients* and *clinical variables* tabs. The user can choose to only see data for certain clinical variables or patients by selecting them in the list. The researcher may see a similarity in certain patients or may want to further investigate a subset of the variables and they are able to by forming a group. We have added the functionality for the researcher to be able to create unique labels for their groups, drag and drop the variables/patients into their corresponding groups, and to download the data specific to that group.

Figure 1: Illustration of the different components and features of VisXplore.

The right hand side of the screen contains the various modules that we have implemented for analyzing data using visualization techniques. To aid in the analysis of clinical data we present the user with many visualization techniques such as area chart, bar graph, chord, co-occurrence matrix, histogram, lines, parallel, and windrose as well as many analytical tools to estimate correlations, averages, standard deviation, t-test, p-values, and many others.

The parallel coordinates visualization (Figure 2(a)) displays all of the patients on a slightly modified line graph. There are M equally spaced vertical axes displayed, each representing a clinical variable. Along these axes N lines are drawn through the axes, where each line contains the data for a patient. Analyzing this visualization helps discover trends in the data such as outliers from the norm or a linear relationship between axes. In addition, the user can interactively define a range within a single or multiple variables (axes), explore the data further, and even create cohorts or subgroups.

The chord visualization (Figure 2(b)) is generated by creating a circular plot consisting of boundaries, where each boundary is a clinical variable. Connections are drawn between boundaries and the thickness of each connection represents the strength of the correlation. Hovering the mouse over a connection displays the corresponding R-value to help the researcher understand the chart. Extra controls are available for the chart, such as only drawing connections above a chosen R-value.

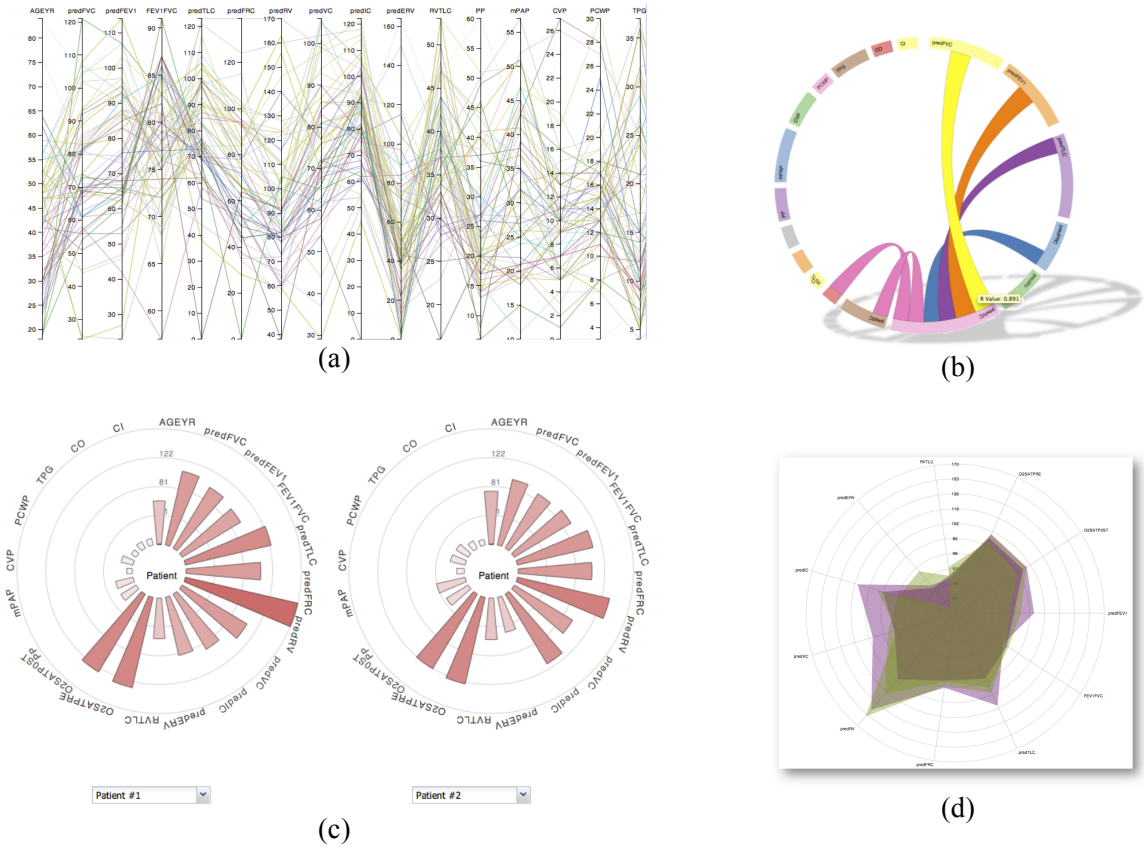
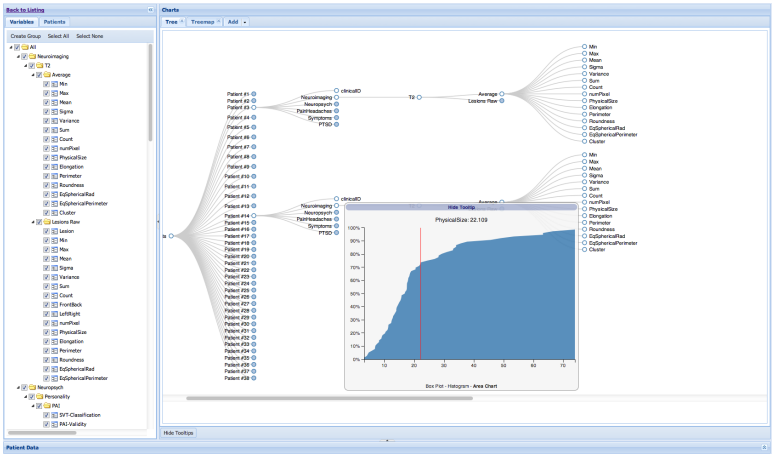


Figure 2: A subset of some of the visualization techniques implemented in VisXplore. (a) Parallel coordinates are used to quickly look at patterns and identify subgroups. (b) The chord visualization is used to study the associations between individual clinical variables. (c) A circular visualization (aka windrose) is used to compare patients. (d) A circular visualization is used to compare patients within a single window.

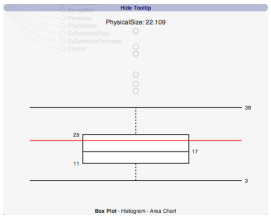
The windrose visualization (Figure 2(c)) is used to illustrate multiple patients side by side to make comparisons easier. A side-by-side chart is created where the two patients' data is drawn in a circular bar chart rather than a standard bar chart. This enables the researcher to compare the shapes of the patients, as similar patients will have similar shapes.

For navigating hierarchical data, we present the user with two useful visualizations: *hierarchical tree* and *treemaps*. The treemap visualization places all of the patients in a grid as individual rectangles. The area of the patient's rectangle corresponds to the amount of data available for that patient, which helps in looking at the overall data. Clicking on a patient's rectangle zooms the user in, displaying another grid of data sub-treemap for the patient and allowing the user to navigate through the hierarchy of data. Figure 3 (b) shows an example of the treemap visualization chart.

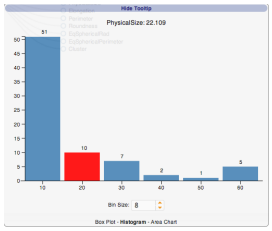


(a)

(b)



(c)



(d)

Figure 3: A subset of the hierarchical visualization techniques available within VisXplore. (a) A tree visualization is used to explore hierarchical data. When the user clicks in a leaf node a window appears with a line illustrating the value of subject under consideration against the aggregate results of the group. (b) A treemap visualization used to explore hierarchical data (c-d) When clicking in a leaf node the user can select different ways to compare the current subject against the group under consideration.

The second visualization, displays the data in a tree graph with the data connected by paths. When a node is clicked, an animation takes place to collapse or expand the data in the corresponding node causing the tree to change shape. Figure 3(a) shows an example of the how the user can interactively explore hierarchical data using the tree visualization.

The treemap and tree visualizations both give the user the functionality to navigate through their hierarchical data, but we needed a way for them to be able to analyze the leaf nodes on the graphs. We accomplished this by presenting the user with a boxplot, histogram, and area chart upon clicking on a leaf node. On each of these three graphs the user is shown where the node that they clicked on is in relation to the rest of the data, by means of a red line indicating the position. This enables single-subject analysis when navigating through the hierarchical data. Figures 3a-d show different visualization techniques.

IV. Results and Conclusion

Our visualization system has been tested with multiple clinical datasets, including blood samples, pulmonary datasets, TBI data, and has been shown to display useful visualizations to aid the user in analysis. The analysis of these datasets using our visualization system has provided effective results compared to the standard visualization tools used by clinicians and researchers. Therefore, we have designed and developed a flexible, easy-to-use visualization system capable of handling tabular, hierarchical, and temporal clinical datasets to assist clinicians and researchers in performing single-subject and group analysis on their patients. Future plans include incorporating machine learning techniques to be able to assist the clinician and researcher in looking for patterns in the data.

References

[1] Daniel A. Keim, “Information visualization and visual data mining”, IEEE Transactions on Visualization and Computer Graphics, v. 8 (1), Jan/Mar 2002

[2] Nan Cao; Gotz, D.; Jimeng Sun; Huamin Qu "DICON: Interactive Visual Analysis of Multidimensional Clusters", Visualization and Computer Graphics, IEEE Transactions on, On page(s): 2581 - 2590 Volume: 17, Issue: 12, Dec. 2011

[3] Michael Bostock, Vadim Ogievetsky, Jeffrey Heer, "D3: Data-Driven Documents", IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis), 2011

[4] Bostock, M., Heer, J., "Protovis: A Graphical Toolkit for Visualization", IEEE Transactions on Visualization and Computer Graphics, v. 15(6), 2009

[5] Heer, J., Mackinlay, J., Stolte, C., Agrawala, M., "Graphical Histories for Visualization: Supporting Analysis, Communication, and Evaluation", IEEE Transactions on Visualization and Computer Graphics, v. 16(6), 2008

[6] Michael Bostock, Vadim Ogievetsky, Jeffrey Heer, “D3: Data-Driven Documents”, IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis), 2011