

User Interface for Biometric Data Analysis

ECE Capstone Design Project - Spring'13

Jaskirat Khangoora, Suhas Kumar, Arsalan Rahman

Advisor: Dr. D. Pompili and PhD student H. Viswanathan

Introduction: Big data is a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications. The three main properties of Big Data, known as the Big V's of Big Data, are Volume, Variety, and Velocity. Volume refers to the amount of data, variety refers to the types of data, and velocity refers to the speed at which data is generated and needs to be processed.

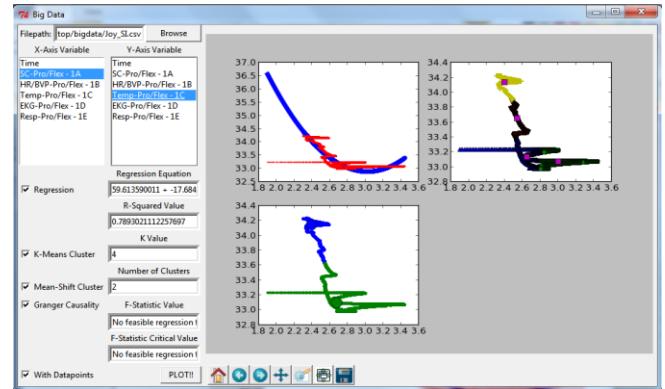
Motivation: Due to the technological improvements over the past decade, there has been a data explosion. According to IBM, we use about 2.5 quintillion bytes of data everyday and 90% of the data we have today has been created in the past two years alone. Almost every electronic device generates some kind of digital data. However, most of this data is not being used due to challenges such as storage, analysis, and visualization. It is difficult to process and analyze such large data sets using traditional data processing applications and on-hand database management tools.

Medical devices have also joined the big data revolution. These devices easily accumulate around 6.5 MB of data in a matter of 5 minutes. Someone in the intensive care unit may be hooked up to multiple devices for extended periods of time. The ability to understand relationships between different biometrics and alarm hospital personnel about possible emergencies in real time would be a life saving application of big data analysis.

Design: Due to Python's relative ease, quickness, and open-source ability, we used this language to implement our Graphical User Interface (GUI). Python's GUI library, Tkinter, was a fairly easy programming tool but gave us

trouble with the geometry of the widgets and the maximization of the plot windows. We then had to include the numpy, numerical python, scipy, science python, libraries to construct the equations necessary for each of the four analysis tools: regression, k-means cluster, mean-shift cluster, and granger causality.

After setting up the basic equation and interface programming aspect of this project, we concentrated on strategies used to control the large volumes and high velocities of the datasets available to us. In order to control the volume and the variety we used two methods. In the first method we would sample only 5-second portions of the data and run the requested analyses on that sample and then repeat the same method for the next 5-second portion. In the second method, we would sample the first 1500 data points instead of 5 seconds worth of data.



Conclusion: Using the user interface we developed, we ran analysis on sample biometric data. Two methods of Big Data Analysis were carried out. Both methods utilized the simple process of sampling Big Data and running the appropriate analysis on that sample. The results of the analysis on the samples were fairly consistent with the analysis run on the total data. However, there were several abnormalities from the sampling analysis compared to the analysis of the full dataset. This goes to show that real-time analysis is doable and is a fairly consistent manner of analyzing large volumes and high velocities of data. However, there has to be a human component to back up the program due to the abnormalities mentioned above. This is especially true in applications concerning life and death.