Selecting, Quantifying, Optimizing, and Understanding Visualization Techniques: A Computational Intelligence-Based Approach

By

Tufail Muhammad

Supervised by

Dr. Zahid Halim

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Computer System Engineering

Faculty of Computer Science and Engineering
Ghulam Ishaq Khan Institute of Engineering Sciences and Technology
Topi, Khyber Pakhtunkhwa, Pakistan
Fall, 2016
Dissertation examination committee:

Dr. Zahid Halim  
*Advisor*, Faculty of Computer Science and Engineering, Ghulam Ishaq Khan Institute of Engineering Sciences and Technology, Topi, PAKISTAN. 🇵🇰

Prof. Dr. Keith C.C. Chan  
*Foreign Evaluator*, Big Data Lab, Department of Computing, The Hong Kong Polytechnic University, Hung Hom, Kowloon, HONG KONG. QS World University Rank=116 🇭🇰

Prof. Dr. Ivan Viola  
*Foreign Evaluator*, The Institute of Computer Graphics and Algorithms, Vienna University of Technology, AUSTRIA. QS World University Rank=197 🇦🇹

Prof. Dr. Michael John Watts  
*Foreign Evaluator*, Academic Head of Programme, Information Technology, Auckland Institute of Studies, Auckland, NEW ZEALAND. 🇳🇿

Dr. Muhammad Tanvir Afzal  
*External Examiner*, Department of Computer Science, Capital University of Science & Technology Islamabad Expressway, Kahuta Road, Zone-V, Islamabad, PAKISTAN. 🇵🇰

Dr. Muhammad Zohaib Zafar Iqbal  
*External Examiner*, Department of Computer Science, National University of Computer and Emerging Sciences, A.K Brohi Road, Sector H-11/4, Islamabad, PAKISTAN. 🇵🇰

Dr. Ahmar Rashid  
*Internal Examiner*, Faculty of Computer Science and Engineering, Ghulam Ishaq Khan Institute of Engineering Sciences and Technology, Topi, PAKISTAN. 🇵🇰
Dr. Ghulam Abbas

*Internal Evaluator (Pre-screening)*, Faculty of Computer Science and Engineering, Ghulam Ishaq Khan Institute of Engineering Sciences and Technology, Topi, PAKISTAN.

Dr. Rashad M Jillani

*Internal Evaluator (Pre-screening)*, Faculty of Computer Science and Engineering, Ghulam Ishaq Khan Institute of Engineering Sciences and Technology, Topi, PAKISTAN.

The work in this dissertation has been carried out at the Faculty of Computer Science and Engineering, Ghulam Ishaq Khan Institute of Engineering Sciences and Technology (GIKI), Topi, Pakistan. The research was supported by the Higher Education Commission of Pakistan (HEC) under the Indigenous Ph.D. Fellowship Program.

Scholar PIN-: PS3-254

Copyright©2015 by Tufail Muhammad
Dedication

To, my father (late)
Abstract

The modern age is contingent on the ever increasing data volumes, its ubiquitous usage, and pervasive nature. This perspective is potentially inevitable for adequately powerful and efficient data analysis and exploration to extract information. Information visualization is a prominent technique to visually explore and analyze large pertinent volumes of data effectively. The plethora of visualization techniques brings new challenges for naïve users in the selection of most appropriate visualization technique for their data. Furthermore, visualization must also be aesthetically appealing and perceptually pleasing to the human cognition. This situation necessitates a framework for predicting visualization technique based on two aspects: the underlying dataset and the task to be performed on the data. Additionally, the resultant visualization must be optimal in the context of aesthetics and human perception.

To investigate these problems, this dissertation presents contributions in three perspectives that subsume information visualization aspects: automatic visualization technique selection, quantifying and optimizing visualization layout, and visualizing software traces. The study provides computational intelligence (CI)-based model to predict a visualization technique based upon the metadata of original dataset and relevant tasks. Similarly, visualization metrics are formulated to objectively measure visualization quality. Based on these metrics, an evolutionary algorithm optimizes the visualization layout. Finally, the hierarchical visualization technique is used to study the application programming interface (API) objects usage for the program trace. The trace is collected using the bytecode instrumentation technique. This dissertation has three parts.

The first part of the dissertation aims to provide a solution to an appropriate visualization technique prediction for a specific dataset. The study shows that the selection of an appropriate visualization technique is dependent upon the metadata of the dataset that is to be visualized. Additionally, the relevant tasks to be performed on that particular dataset are also considered. A customized dataset has been built using the knowledge that exists in the contemporary literature on various visualization techniques. The dataset comprises of four
Abstract

metadata attributes, relevant task, and the visualization techniques. Moreover, the study develops an artificial neural network (ANN)-based model for an appropriate visualization technique prediction with five input and eight output neurons. The ANN-based model is trained on the dataset with eight visualization techniques: histogram, line chart, pie chart, scatter plot, parallel coordinates, map, treemap, and linked graph. Optimal neural network architecture is obtained by evaluating various structures with different network configuration. Several well-known performance metrics, i.e., confusion matrix, accuracy, precision, and sensitivity of the classification are used to compare various neural network architectures. Additionally, the best ANN model is compared with five other well-known classifiers; k-nearest neighbor (k-NN), naïve Bayes (NB), decision tree (DT), random forest (RF), and support vector machine (SVM). The best ANN model is also tested on twenty benchmark datasets to predict the most appropriate visualization technique. A qualitative comparison with the state-of-the-art approach is also presented. The results show that the proposed technique is useful in selecting an appropriate visualization technique for a given dataset with high accuracy.

The second part of this dissertation provides solution for the design of an optimal visualization using visualization quality metrics. Initially, the study focuses on the design parameters that contribute towards the quality of a particular visualization technique. Furthermore, visualization metrics are proposed to measure visualization aesthetic and perceptual characteristics. They include: effectiveness, expressiveness, readability, and interactivity. An evolutionary algorithm (EA)-based framework for optimizing the layout of a visualization technique is also proposed using the design parameters and the visualization metrics. The EA uses a fixed sized chromosome and the fitness function, which is based on the linear combination of the aforementioned four visualization metrics. Treemap visualization technique is used as a case study for layout optimization using the EA. Additionally, the results are evaluated using control experiments and benchmark tasks. A comparison of the results is also presented with the state-of-the-art approaches using internal and external criteria. In general, the findings show that the proposed approach is effective in designing aesthetically appealing and perceptually pleasing visualizations.
The last part uses treemap-based visualization for the analysis of API objects used in the software. Particularly, in understanding which collection API’s objects are used during runtime of a Java program. The work consists of two aspects: the extraction of collection APIs information from a Java program using bytecode instrumentation, and development of a visualization tool to analyse the collected traces using treemaps. Initially, a bytecode instrumentation tool is developed to probe and collect runtime information from the Java programs. The extracted information is logged into an extensible markup language (XML)-based file. Furthermore, synthesis of the log file is performed using treemap-based visualization tool. The visualization shows the complete log file on a single screen, where API objects are shown in a hierarchy. The visualization provides a global view to the usage of collection API objects at different locations of a program. The instrumentation part is evaluated using twenty benchmark and ten real world applications. The results show that the instrumentation tool causes minimal runtime overheads. The visualization tool is evaluated using controlled experiments with twenty four participants divided into two distinct groups. The results show that the proposed tool enables the viewer to comprehend more information with less effort and time. The proposed approach will help the developer in understanding the Java collection API’s object usage, and it will assist in program comprehension and maintenance.