Optimization of Parallel Coordinates for Visual Analytics

By

Liang Fu Lu

Supervisor：Mao Lin Huang

A Thesis submitted in Fulfillment for the Degree of Doctor of Philosophy

in

Faculty of Engineering and IT

University of Technology, Sydney Australia

March 2016

CERTIFICATE OF AUTHORSHIP/ORIGINALITY

UNIVERSITY OF TECHNOLOGY SYDNEY

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

SIGNATURE OF STUDENT

_____

ACKNOWLEDGEMENTS

How time flies. It has been nearly five years since for my research study I first toke my feet on the grand and attractive university-University of Technology, Sydney, where I had been dreaming for years to go. Fortunately, I acquired a variety of knowledge in the short but precious semesters like a tiny boat loading to full capacity of great achievements, expectations and dreams was sailing backward in the sun shine through the ordeal from the sea storm. At the tranquil and solemn campus, everywhere is surrounded by the strong and profound academic atmosphere, resounded with vigorous study aspirations and left the track of our familiar figure shadows and good memories, which arouses a strong sense of gratitude in my heart.

Firstly, I would like to gratefully acknowledge Mao Lin Huang, an A/Prof in the field of information visualization, who has who has earned a great reputation at home and abroad for his numerous contributions to scientific researches and practical achievements, for the enthusiastic supervision in the process of the research. This thesis grew out of a series of dialogues with him. Each step of completing the thesis including selection, proposal, research and writing is dissolved in Prof. Huang's hard work and guidance. What was deeply impressed me is not the toil and sweat he scattered in numerous readings and modifications, but the wise academic vision, noble life faith and rigorous scholarship attitude he had, which will benefit me a lot in my rest whole life. He brought me closer to the reality I had initially perceived, and eventually enabled me to grasp its rich complexity.

Furthermore, I owe sincere thankfulness to the members of Visualization Team and fellow researchers and the professors of iNext Research centre. Their scholarly attainments and instructions in humor, serious and patient way gave me great help and

inspirations, and that besides gaining knowledge also benefitted me many like friendship, concern, trust and so on.

This research also benefited tremendously from many researchers and staffs in the University of Technology, Sydney. In addition, thank you to the participants in the usability study for the cooperation and valuable feedbacks which are taken as a key factor in providing a necessary guarantee on the datum I collected in accuracy and precision and the evidences in objectivity and scientificity for the thesis.

Last but not least, I owe my deepest gratitude to Huan Xu, my wife, who bravely took the burden of supporting the family alone without any complaints and totally poured her love onto my daughter and me and my whole family. All of her toil and sweat, tolerance and understanding, selfless support and devotion set up the strong will and belief for me in my study. I really appreciated sincerely my extended family for their continuous encouragement and support to make this PhD thesis possible. Thanks my daughter Wenxuan Lu for constantly delivering me angel smiles which accompanied me all the way. Her sweet smiles once encouraged me stand up again after fall a million times and then bravely and firmly faced the difficulties I had encountered. You are all my love most worthy of spending my lifetime loving.

Contents

Figure List

Table List

Equation List

Algorithm List

Abstract

The visualization and interaction of multidimensional data always requires optimized solutions for integrating the display, exploration and analytical reasoning of data into a kind of visual pipeline for human-centered data analysis and interpretation. However, parallel coordinate plot, as one of the most popular multidimensional data visualization techniques, suffers from a visual clutter problem. Although this problem has been addressed in many related studies, computational cost and information loss still hamper the application of these techniques, which leads to large high dimensional data sets. Therefore, the main goal of this thesis is to optimize the visual representation of parallel coordinates based on their geometrical properties.

At the first stage, we set out to find optimization methods for permuting data values displayed in parallel coordinate plot to reduce the visual clutter. We divide the dataset into two classifications according to the values and the geometric theory of the parallel coordinate plot: numerical data and non-numerical data, and missing data may exist between them occasionally. We apply Sugiyama's layered directed graph drawing algorithm into parallel coordinate plot to minimize the number of edge crossing among polygonal lines. The methods are proved to be valuable as it can optimize the order of missing or non-numerical value to tackle clutter reduction.

In addition, it is true that optimizing the order is a NP-complete problem, though changing the order of the axis is a straightforward way to address the visual clutter problem. Therefore, we try to propose in the research a new axes re-ordering method in parallel coordinate plot: a similarity-based method, which is based on the combination

of Nonlinear Correlation Coefficient (NCC) and Singular Value Decomposition (SVD) algorithms. By using this approach, the first remarkable axis can be selected based on mathematical theory and all axes can be re-ordered in line with the degree of similarities among them. We also propose a measurement of contribution rate of each dimension to reveal the property hidden in the dataset.

In the third stage, we put forward a new projection method which is able to visualize more data items in the same display space than the existing parallel coordinate methods. Moreover, it is demonstrated clearly in the research that the new method enjoys some elegant duality properties with parallel coordinate plot and Cartesian orthogonal coordinate representation. Meanwhile, the mean crossing angles and the amount of edge crossing between the neighboring axes are utilized in this research to demonstrate the rationale and effectiveness of our approaches.