UNIVERSITI TEKNOLOGI MARA

EXCAVATING CLUSTERS OF STAR COORDINATE VISUALIZATION THROUGH DIMENSION MANIPULATION

IZYAN IZZATI BINTI KAMSANI

Thesis submitted in fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Information Technology and Quantitative Science)

Faculty of Computer and Mathematical Science

November 2019

ABSTRACT

Recent advancement in high-performance technologies has resulted in collections of high dimensional data. High dimensional data consist of large number of data records and attributes. It is challenging when it comes to extracting high dimensional data into meaningful information. Cluster analysis is used as a method to provide useful summary so that it can be understood more straightforwardly. The summary obtained, however is difficult to interpret from raw data. In order to understand the clusters results, it is visualized in a low dimensional plane. There are many available visualization techniques that can handle high dimensional data. One of it is geometric projection technique. Examples of geometric projection techniques are; Scatterplot (SP), Parallel Coordinate (PC) and Star Coordinate (SC) visualization techniques. However, for this study, only SC visualization technique was explored in further details. SC can display high dimensional data into a limited space. SC performs by plotting data dimensions in a circular arrangement. Through this method, it assists users to view the data pattern as a whole in a single space. Nevertheless, mapping high dimensional data could produce clutter and overlap. Consequently, SC provides interactive features through dimensions manipulation to search for hidden information. There are dimensions arrangement and scaling. Excavating clusters which are also known as digging clusters can help novice users to determine the correlations among data dimensions. Without prior knowledge of relationship between data dimensions, novice users will be blindly searching for clusters. This is due to clusters appearance that depends on data dimension position and length. Hence, users need to identify the arrangement of data dimensions and which data dimensions contribute to the cluster appearance. Therefore, the study into manipulating interactive features of SC technique draws attention to be investigated. This thesis proposes a method to identify the preferable dimensions position and scaling axis that can help novice users to understand the distribution of plotted data in SC environment. It supports discovering clusters, interpreting the relationship patterns between data dimension. This method covers three parts; dimension arrangement (equal angle), dimensions angle (based 360 degrees) and dimension scaling (equal angle and based 360 degrees). Initially, the arrangement of dimension was measured using Pearson Correlation (PR); meanwhile, the obtained correlation value was converted into based 360 degrees; lastly, the dimension scaling applied Manhattan Distance (MD) to determine the contribution of each data dimension towards cluster performance. The proposed method has successfully produced clusters. It is positively supporting novice users with limited knowledge in using SC technique to further manipulate interactive features during their analysis.

ACKNOWLEDGEMENT

In the name of Allah, the Most Gracious, Most Merciful.
Alhamdulillah, praises to Allah whom with His blessings allowing me to complete this thesis.

Firstly, I wish to thank Allah for giving me the opportunity to embark on my PhD and for completing this long and challenging journey successfully. My gratitude and thanks go to my supervisor Dr Noor Elaiza Abdul Khalid for her inspirational and patient guidance, general support, and unselfish help.

My appreciation goes to my husband (Faizzul Fadly Bin Jaafar) who is always there in a happy or difficult situation. My children, Firas Irsyad, Iman Farisya and the growing baby in my womb are my main pillars in ensuring that I can finish my studies.

This thesis is dedicated to my parents especially my father (Kamsani Bin Johan) for giving encouragement and motivation to continue my study until this stage. Not to forget to my mother who has always been supportive of me. Thank you to my family-in-law who supported me to finish this journey. This piece of victory is dedicated to all of you. Alhamdulillah.

Thank you to my siblings (Farah, Afiqah, Hazirah and Aqilah) for their full support. Thanks to my friends for helping me with this project.

Last but not least, a sincere thanks to Universiti Teknologi MARA (UiTM) and Ministry of Higher Education (KPT) for providing me with the MyBrain Scholarship which had helped me greatly during the period of my studies.

Thank you very much.

TABLE OF CONTENT