Parallel Coordinates

Addanki Adithya
Department of Computer Science, The University of Akron,
Akron, Ohio-44304. aa207@zips.uakron.edu

Abstract

The project's objective is to implement a system in WPF for Parallel Coordinates. The report starts by explaining about parallel coordinates, their existence, various application areas, and limitations. The paper also discusses the technical challenges and possible extensions to the project.

1. Introduction

Visual Communication is the fundamental process for exploring and dissemination of information through visual aids such as diagrams, sketches, photographs, charts, video, animations etc., thus improving the quality of comprehension, memory and inference. The strength of visualizations lies in the design principles which connect the visualizations with viewer's perception and cognition of underlying information that is conveyed. The key to good visualizations use the best practices to either emphasize the relevant information or de-emphasize irrelevant information [5].

2. Background

2.2 Parallel Coordinates

Visualizing information that has more than 3 dimensions is not understandable to the human eye. Parallel coordinates are introduced by Philbert Maurice D'Ocagne in 1885[4].

Parallel coordinates are one of the most famous visualization techniques, while initially confusing, they are a very powerful tool for understanding multi-dimensional numerical datasets [1]. Represents multiple dimensions with the help of parallel axes. Samples in data set are represented as poly lines spanning the axes.

3. Design

3.1 UML Diagrams

We build models to [6]

- To communicate the desired structure and behavior of our system.
- To visualize and control the systems architecture.
- For better understanding of the system we are building; often exposing opportunities for simplification and reuse.
- To manage risk.

3.1.1 Use case diagram

A Usecase diagram is used to illustrate the static design view of the system. Given below is the use case model for our system. Representations and relationships used inside of the system could be checked through references [6].

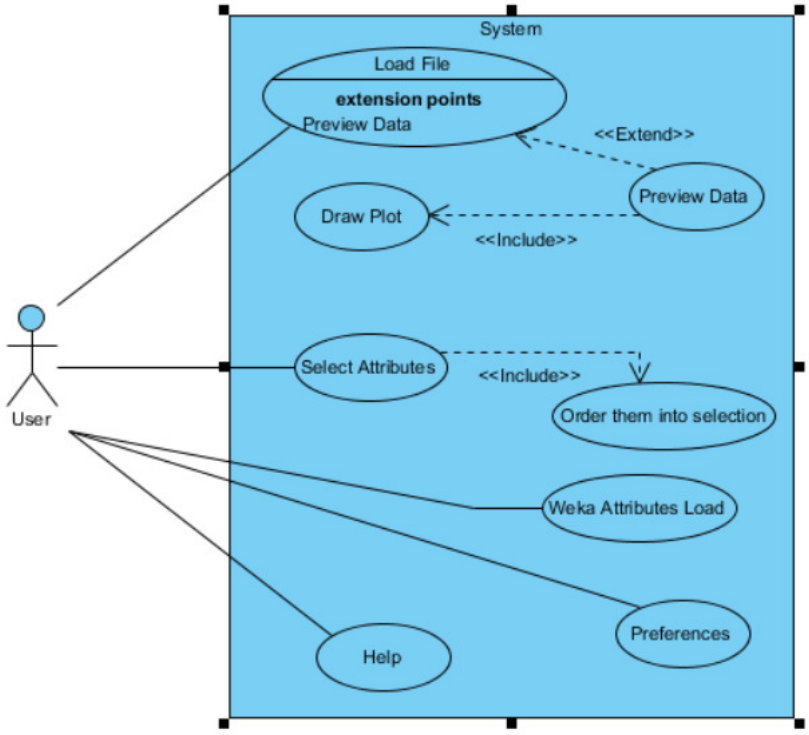

Figure 3.1.1.1: Use Case Diagram

The use case diagram illustrates the requirements on a given system. It gives an understanding of how the model is implemented and the actions the user can perform over the system through his interaction. In other words it represents the behavioral view of the system.

3.1.2 Class Diagram

A class diagram depicts a static view of the system, the way different classes interact with each other, their relationships and dependencies between the classes.

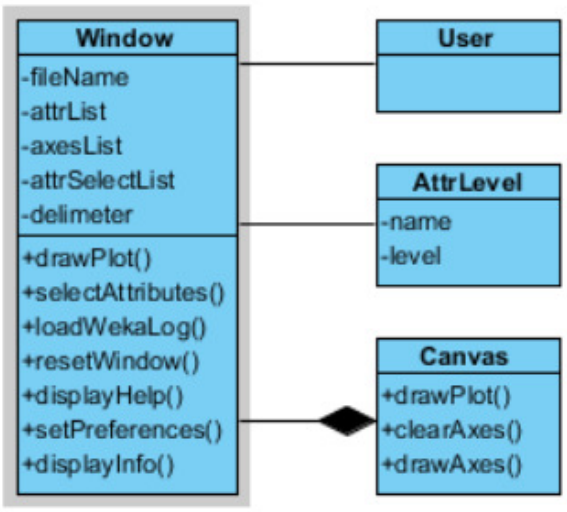

Figure 3.1.2.1: Class Diagram

As shown in the [Figure 3.1.2.1] we have the windows, canvas and other event listeners that allow the user to interact with the visualized model and render again when required based on user selection.

3.2 Implementation

3.2.1 Dataset

Three gorges project dataset that has recorded responses to a questionnaire before and after the relocation of 1.3 million Chinese people. Prospective panel study involving two phases (late 2002 to early 2003 and 2006) of F2F interviews with 1530 people (975

designated migrants and 555 non-migrants). Though started with 1530 subjects, the second wave was able to capture a success rate of 70%, thus resulting in the final sample size of 1056; 350 Non-Migrants, 286 designated Migrants who have not yet moved, 420 remaining subjects who are designated and moved.

3.2.2 Algorithm Implementation

The implementation is done in WPF, with code behind programmed in C#.NET. The application has been implemented to handle three kinds of files for parallel coordinate plots; CSV, XLS and XLSX. Also, once the plot is drawn the attributes could be filtered through user selection or by loading a Weka Log file; TXT and LOG files.

4. Results

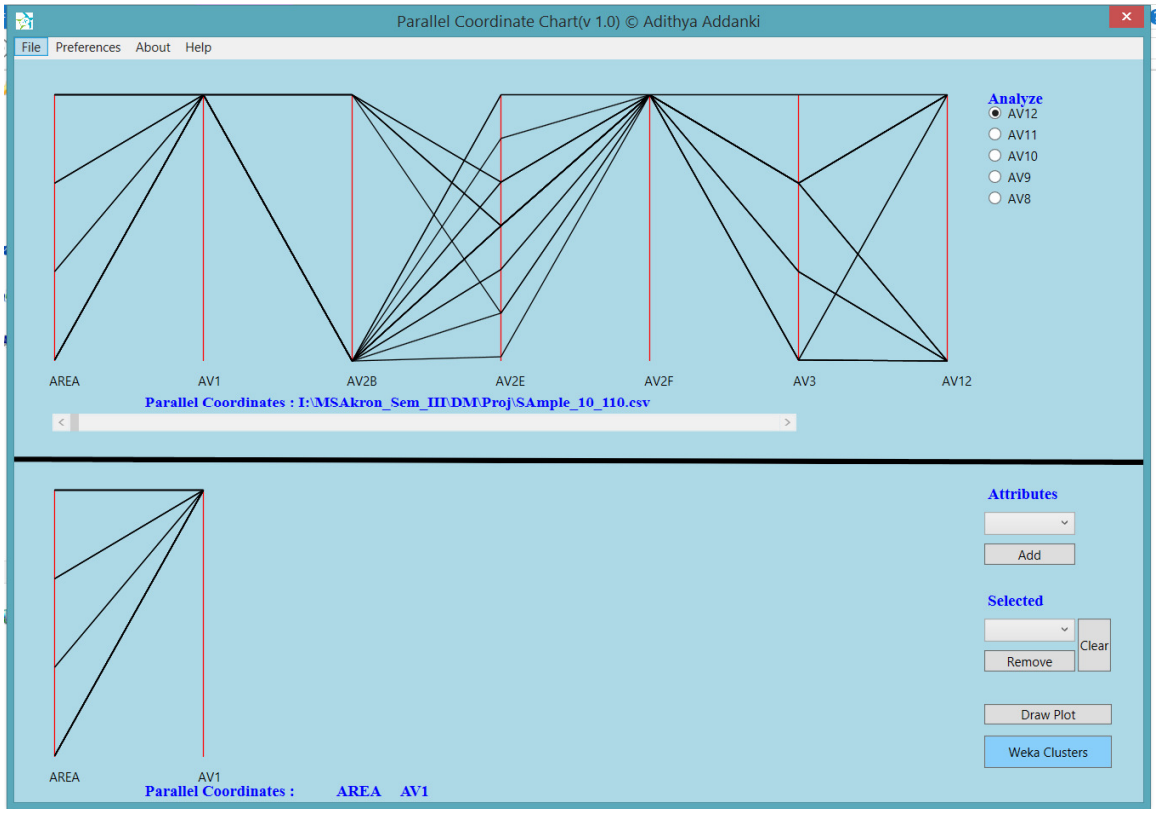

Figure 4.1: Sample Dataset Loaded into Application

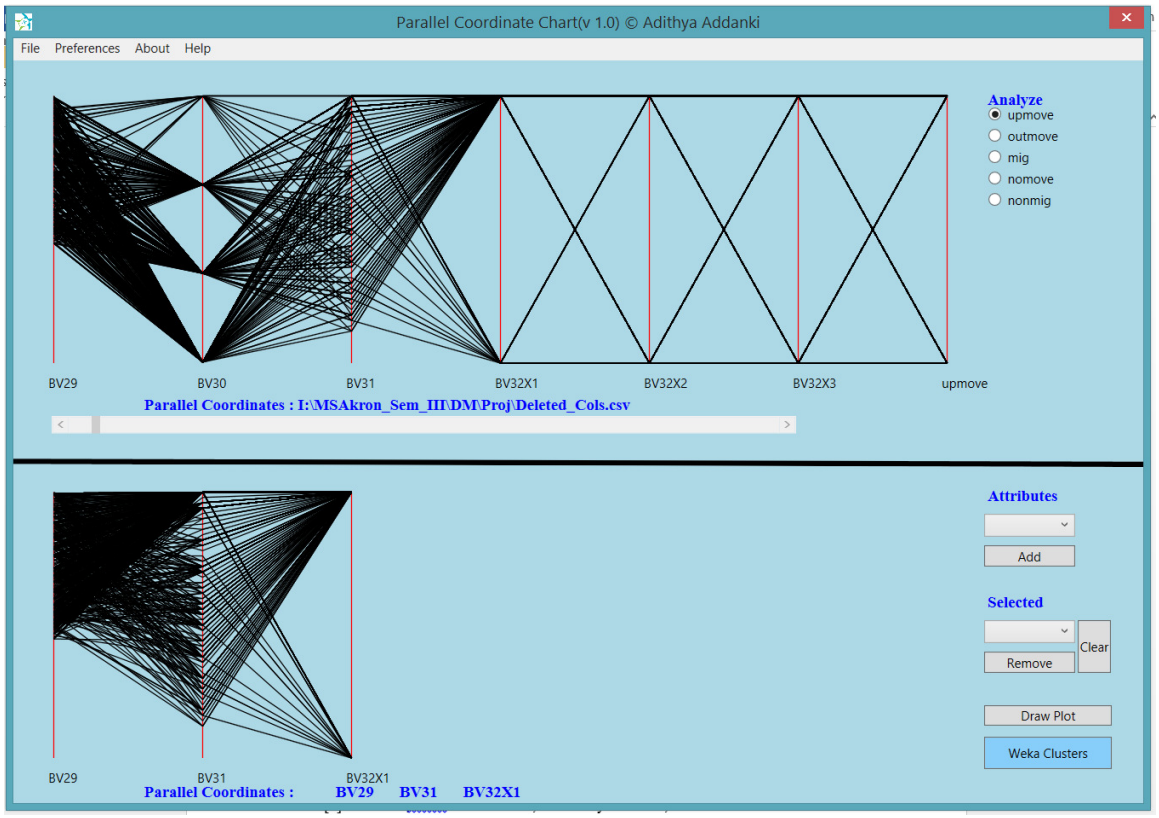

Figure 4.2: TGP Dataset Loaded into Application

5. Observations

Representing heterogeneous data types and scaling the range of values to fit to fixed axes was a technical glitch. A few of limitations are listed below:

- Each axis can only be compared with two other axes at once;
  - One to the left another to the right
- More range of values in each attribute might cause confusion to the user. Use with caution with regards to categorical data.
- ASCII based text characters only[character sets and digits]
- Rearrangement of attributes | axes order might be required to view patterns
  - Good arrangement of axes could be dealt using heuristics

6. Conclusion

Parallel coordinates are a very versatile and useful technique for finding structures in moderately-sized datasets. With a bit of experience, it is possible to very quickly recognize patterns and even estimate the strength of correlations, etc. Parallel Coordinate plot especially helps identify the patterns, clusters, characteristics of interest in a dataset.

Possible extensions to the application include:

- Brushing- A provision to analyze the parallel coordinate plot where clusters are formed.
- Rearranging axes- drag and drop axes to suit the ordering the user is interested in.

Parallel coordinates find its applications in Data Mining, Traffic Control, Process Control, and Computer Vision. Thus it has become one of the most important tools in Data Visualization.

7. References

[1] . https://eagereyes.org/techniques/parallel-coordinates

[2] . http://www.xdat.org/

[3] . Dr. Xiao, Data Visualization; University of Akron, Lecture Notes.

[4] . http://en.wikipedia.org/wiki/Parallel_coordinates

[5] . Agrawala, M., Wilmot, L. I., & Berthouzoz, F. (2011). Design principles for visual communication. Communications of the ACM, 54(4), 60-69. doi:10.1145/1924421.1924439

[6] . Grady Booch, James Rumbaugh, Ivar Jacobson. The Unified Modeling Language User Guide (2005).ISBN-10: 0-321-26797-4. ISBN-13: 978-0-321-26797-9.