Flexible Views in Visualizing Multiple Response Survey using Murvis

Siti Z.Z. Abidin, M. Bakri. C. Haron and Zamalia Mahmud

Abstract—In a survey investigation, subjects will normally provide only one answer for each question. However, when they are allowed to provide more than one answer, a specific technique is needed to represent the observed data in a visual form that could be easily seen and viewed in terms of its similarity and dissimilarity among the survey attributes. One of the common techniques used in visualizing the results is by using multidimensional scaling (MDS). MDS is often used to provide a visual representation of the pattern of proximities (i.e., similarities or distances) among a set of objects that allow results to be interpreted according to the survey subjects and attributes. However, too many subjects and attributes will produce massive output points (coordinates). In order to enhance the visualization, a tool called Murvis (Multiple Response Visualization) has been developed using Java programming language to provide users with the flexibility in visualizing the MDS output coordinates in 2D and 3D space. Murvis allows users to add colors to the graphic visual and present the output in many different and flexible views. With the latter, analysts of multiple response survey are able to illustrate a more informative research findings. A small scale survey involving three data sets are used to test the usability and effectiveness of the tool which has some impending significances.

Keywords— Attribute, Clustering, Distance ratio algorithm, Java program, MDS output, Visualization.

I. INTRODUCTION

In a survey investigation, it is important to see the clustering pattern of subjects who provide multiple responses to a survey question. The clustering pattern should indicate the similarity of answers among the subjects. Other than survey, there are other disciplines that require clustering of data or information for classification and analysis. Such disciplines include bioinformatics[1], chemical processing [2], education[3] and many more.

In order to produce the pattern, proximity multidimensional scaling (MDS) is one of the common techniques used in visualizing the results that reduces the dimension of data for presenting information in clusters. The results can be interpreted according to the survey subjects and attributes. However, too many subjects and attributes will produce massive output points (coordinates) in the presentation and hence difficulty in the visualizing output in two dimensional (2D) space. An attempt has been made to ease the analysis of results by transforming the output into three dimensional (3D) view [4], however, there is no flexibility in visualizing the results. There are also several tools available in the market for data analysis but most of it focus on the raw data processing rather than enhancing the visualization of the output results.

In order to enhance the visualization, we propose a tool called Murvis (Multiple Response Visualization) [5] to provide users whom the main target group are researchers, to visualize the MDS output coordinates in 2D and 3D colored space with flexible views. The MDS results are reclassified based on attributes given in the datasets. Murvis is developed using Java programming language to read all the MDS output coordinates. The distances between all the coordinates are recalculated to cluster the results into several classes according to the given attributes. The calculation is also used for visualizing the results in 3D colored space, whereby another dimension is added to the 2D output coordinates in terms of height. The heights are determined using the distance ratio algorithm [4].

In a small scale study, a dataset that consists of 50 respondents is used. The functionality testing is performed to verify that the visualization of the result is produced in colors and both 2D and 3D views. Three datasets are used to do the testing on the tool. The study aims to assist researchers or statisticians in analyzing their multiple response data more effectively.

This paper is organized as follows; Section 2 discusses on visualization and other related works while Section 3 gives the details of this tool. Section 4 presents the results before the concluding remarks in Section 5.

II. DATA VISUALIZATION

Data visualization is a way to present and display information that encourages appropriate interpretation, selection and association of data. Other than data, it is also possible to visualize the relation between several entities for the purpose of understanding the information that they cover such as the relation between lots of documents [6].
requirements engineering (RE), information is visualized by transforming the written and verbal requirements into graphics or visual representations [7].

In this research, the study is focused on shifting the load from numerical reasoning to visual reasoning and exploiting the capabilities of human eye to detect information from pictures or illustrations. Finding trends and relations in the data from a visual representation is much easier and far more time-saving compared to looking through text and numbers [8]. Interactive manipulation and control of visualization are important tools which allow users to focus on the region of interest. He also stated that the important aspects of interactive visualization can be broken into three categories; computation (the ability to speedily compute visualization), display (the ability to quickly display the computed visualization), and querying (the ability to interactively probe a displayed visualization for the purpose of further understanding) [9].

Moreover, color is important and frequently used in data visualization. A reasonable number of colors must be used to avoid difficulties for the viewer to get any useful information [8]. Furthermore, the visualization can be presented in 2D or 3D space according to the suitable data and the needs of the evaluators.

A. 2D and 3D Visualization

There are many opinions regarding the comparison of 2D and 3D visualization when choosing the best display of result. Choosing the best dimension to use depend on the field itself. The views can be applied differently, depending on the domain, tasks specification and mental registration [10]. Two dimensional views are often used to create precise relationships, and 3D views are used to gain a qualitative understanding and presenting ideas to others [11].

In general, 2D views are good for viewing details of a specific part and navigating or measuring distance precisely. Whereas, three-dimensional displays are good for gaining an overview of a 3D space and understanding 3D shape. Since 3D and 2D views have their own advantages, having both visible presentations may benefit certain tasks such as orienting and positioning objects relative to one another [12]. 3D representation is found to have harder interaction with objects when compare to 2D [13]. Interacting with 3D can be quite challenging since the interpretation in 3D space is not straightforward and depend on user’s experience [14]. Obvious difference between 2D and 3D display techniques is how easy user can relate and combine information from different views [10]. A good 3D design will result in giving an effective visualization since it includes few added elements to the visualization such as colors, objects, spaces and sizes [8].

Many techniques have been developed to map multidimensional data to 3D scenes because data from different fields often require different visualization techniques [15]. One of the advantages of 3D is its ability to show the relationships between variables [16].

B. Tools for 2D and 3D Visualization

When There are several data visualization tools available to assist user in data analysis process. Most of it include color enhancement and raw data clustering in order to see the similarity pattern. Examples of 2D visualization tools are Projection Explorer [17] and FlowVisMenu [18], and examples of 3D visualization tools are OriginPro [19] and Scatterplot3D [20].

Projection Explorer is a multidimensional visualization tool used for visualizing multidimensional data through high speed precision projections. It includes multiple projection techniques and has various functionalities in visualizing the data and interacting with them in order to find interesting data relations. Fig 1 shows the sample output produced by the Projection Explorer.

Another 2D tool, FlowVisMenu is a radial menu containing a scatterplot that can be popped out transiently and manipulated with rapid, fluid gestures to select and modify axes of the scatterplot. The program’s user interface makes exploring along multiple metrics easy. The smooth transition and integration of views are useful during data analysis process. Fig 2 illustrates the sample output.
For 3D, OriginPro is a specialized program for data analysis providing Fast Fourier Transformation analysis, profile plots and 3D color maps surface of computerized tomography (CT) images. One of the functions in this software is Contour Plot that is useful for delineating organ boundaries in images as shown in Fig 3.

![Fig 3. 3D color surface map of an x-ray CT brain scan][19]

Scatterplot3D is an R package for the visualization of multivariate data in a three dimensional space where R is a programming language for data analysis and graphics. Fig 4 depicts the example output.

![Fig 4. Scatterplot3D][20]

The purpose of 3D visualization tools is to give better insight of the data. In each of the tools, the data is treated as a solid mass that can be navigated, analyzed, zoomed and rotated. Interactivity and flexibility in the visualization tools can help users during the analysis process. However, in all these tools, user cannot manipulate the results of the visualization.

III. Murvis

Murvis is designed to help users to analyze the results of a MDS proximity scaling more efficiently. It takes all the output coordinates from the MDS results as the input data before calculating the distance between one coordinate in relative to all other coordinates. The tool provides the concept of wizard to guide users whom usually are not familiar with the technical issues of computers. Therefore, there are menus and buttons for the user interface.

The tool is not only used for 2D colored representation, but also recalculating the 2D coordinates to transform the visualization into 3D colored space while providing the flexibilities in viewing the output through zooming, rotation and region of interest.

For a case study, a small scale survey instrument comprising three questions/items are developed. The aim of the survey is to find out about users’ preferences and reasons in choosing five different colors, i.e., red, blue, green, purple, and yellow. Although this is a simple survey, it has significances in evaluating people’s choices toward determining their favorite colors. The fabric industry may be interested in knowing the outcome as this survey can provide useful information for fabric manufacturers to identify which color should be given priority and produced most.

There are 50 subjects in the survey, representing 50 cases and a total of 10 choices of answers that signify 10 attributes. The survey questions and some of its examples of X and Y coordinate values are illustrated in Fig 5.

![Fig 5. Sample survey questions and MDS coordinates][21]

Refering to the first survey questions, there are five (5) choice of answers, which is also referred to as attributes. The second question has three attributes while the final question has two attributes. Subjects are allowed to choose more than one answer with a total of maximum ten answers for each survey. This leads to ten attributes that represent multiple choices of answers.

Based on the survey questions, there are three possible classifications that can be made if the clustering is to be performed based on attributes, preferred choice of colors, reasons for the choice, and gender.
Each of the attribute value has its X and Y coordinates to present its location in the 2D space map, whereby the region of output has four sections; upper left, lower left, upper right, and lower right.

A. Framework

Based on ADDIE model, a framework of Murvis is designed to structure the phases involved during the implementation process. The framework is shown in Fig 6.

In the preliminary phase, related journals and articles are read for defining clearer objectives and scope, while in prototype design phase, the 3D representation and required calculation are identified. For the development phase, several rules and algorithms are applied and appropriate Java graphics packages are selected, modified and written. This phase can be considered as the most difficult part of the work as it requires skills and knowledge in graphics programming. Section 3.2 will describe the details of the workflow at this stage.

After the tool is developed completely, the functionality and usability tests are carried out and several test cases are used. For result analysis phase, several MDS visualization views are compared especially in 2D and 3D spaces. Details steps of the methodology is illustrated in Fig 7.

The prototype interface design is also one of the important steps in the methodology because the focus of the end user is to do the analysis without having difficulty in using the tool.

B. Main Work Flow

The work flow is the core part of the tool. The process is depicted in Fig 8. There are three phases, namely, input, process and output. For the input data acquisition, the tool relies on the output coordinates produced by MDS. The coordinate values are the position of X and Y points. The values are prepared in a text file. Each input file must follow a specific format, such as respondents’ input is represented at each line in the file to carry the information of one respondent. Details on the attributes and their associated names are also provided properly to the tool as an input file.

There are three activities in the process phase that include adding another dimension to the original 2D so that the 3D view will be produced, classifying all the attributes in the dataset so that colors can be applied for better information analysis, and determining the suitable objects for data representation.
The last phase, the output, is the aim of the tool to provide the 3D visualization that can be viewed either in 2D or 3D space with flexible control and manipulation.

C. Algorithm

There are several algorithms applied to ensure the accuracy of the visualized information. Among all, the calculation of distance between one coordinate in relative to other coordinates in space is very important. The tool will recalculate the existing MDS output coordinates to determine the required values as shown in Fig 9.

![Fig 9. Determining distance between one coordinate to all other coordinates](image)

To ensure that every data has its own distance to every other data coordinate, linear search algorithm is applied to find and calculate the distance between all the coordinates. In Fig 9, $D1$, $D2$, $D3$, $D4$ and $D5$ are the distance of the center point to all other points $ATT1$, $ATT2$, $ATT3$, $ATT4$ and $ATT5$. The calculated distances are also used to calculate the third dimension using distance ratio formula [4] with density to indicate the similarity factors. Some of the formulas used to calculate the distance ratio are as follows:

\[
\text{distance} = \sum_{j=1}^{n} \sum_{i=0}^{n-1} \text{dist}
\]  
\[
\text{ratio}_i = \frac{\text{distance}_i}{\text{maxVal}}
\]  
\[
Z_i = 1 - \text{ratio}_i
\]

The $\text{distance}$ is the calculation distance between one coordinate to all other coordinates in the system, while $\text{ratio}$ is the percentage of total distance of one coordinate in relative to the maximum distance found among all the coordinates. Lastly, the value $z$ indicates the density (for height) for each of the coordinate.

The larger value in heights (in 3D space) indicates that the particular point is located near to many other points, meaning that many respondents choose similar answers. When the points are closed to the measuring point, the shorter the distance will give the larger value in $z$.

Data for this tool is in the form of classes and attributes. Classes represent questions in the survey and within each of the class, there are attributes that represent answers for that particular question. All the coordinates will be classified by each class and its result will be used during the process of color assignment based on the chosen classification. To add color enhancement to the output, the nearest distance between the attribute’s coordinate needs to be identified in each class.

D. Structure

When implementing the tool using Java programming language, it can be guaranteed that the flexible views can be achieved as programming can provide users with possible interactive control. There are thirty three Java Classes written to make up this tool, as illustrated in Fig 10.

![Fig 10. List of Java Classes](image)
All the Java Classes can be summarized in a package diagram that consists of six main packages as shown in Fig 11.

![Fig 11. Package diagram](image)

In the diagram, the relation of each entity is drawn to indicate that the input data, graph manipulation and graphical user interface (GUI) are among the important components that require proper design and implementation.

The algorithm that classifies all the points must support flexibilities so that it will work with every data set of the unknown number of respondents, survey questions and attributes. The pseudo codes for classification and color enhancement algorithm is as follows:

```plaintext
for i<-0 to number of respondents + attributes
    attInClass = number of attributes in 1st class;
    Initialize min value;
    minIndex = 0;
    for j<- 0 to number of attributes -attInClass;
        dist = distance between this point to attribute j;
        if min > dist
            min = dist;
            minIndex = current index;
        if no more attribute in current class
            store minIndex for this class;
            if has more class
                attInClass = get number of attributes in next class;
                reset min and minIndex value
    store classification result for this point
```

With the consideration of any number of respondents and attributes, the tool can accept any given MDS coordinates.

**IV. RESULTS**

In visualizing the new MDS output, the result is determined by comparing the visualization of normal MDS output and the enhanced output with colour, thus giving better visualization and understanding.

There are two main windows in the interface; the control and visualization windows. Users can select and manipulate any commands in the control window, while the output results are displayed in the visualization window. Fig 12 shows the example output of the tool.

![Fig 12. Wizard windows](image)

Within the windows, several buttons are provided to ease the manipulation of input, process and output of the visualization. For the Graph Scene Window, the canvas of an instance of the Graph object will be embedded to the Java SWT panel. This Graph object will contains information about current visualization state.

```java
JFrame frame = new JFrame();
frame.add((javax.swing.JComponent)graph.getCanvas());
```

In the Control GUI window, there are three main drop-box menus for users to choose. The data will be classified by on the fly, and color information for each attribute. Fig 13 shows the example window.

![Fig 13. Control GUI](image)
function according to the subjects’ actual choices and preferences for the answers. A custom event handler must be added to each of the function to make this prototype interactive, where the result of each action will have effects in the visualization at real-time.

A. Controlling Presentation

Since color is among the important features that need to be provided to the users, color selections must be provided so that the visualization can be viewed according to the users’ needs. Fig 14 illustrates the color window provided for users. The color window is popped out when the selection is performed.

Fig 14. Color window

Besides color enhancement, users can view all the input data and any information regarding the coordinate values. The information is presented in a table forms if any necessary values need to be checked.

The most challenging part in implementing the tool is to provide flexibilities in the visualization process. All menus and functions in the tool must be set to be dynamic that can accept any input dataset. Users can choose their own data and the data can be set visible or invisible. This function is important during data analysis since user can only focus on the data that they are concern about. Users are also given a menu called visualization that will list the possible options in the visualization as depicted in Fig 15. By using menu and buttons, selection to the appropriate choice can be made as quickly as possible.

Fig 15. Visualization choices

With the wizard approach, it is easier for users even without the computer science background to operate and use the tool to set the visualization of the new MDS output.

B. New Enhanced Visualization

For viewing the result in 3D space, the tool will automatically read the input data and recalculate the distance between all the coordinates in the dataset. Since the original MDS output is in 2D space or only top view, the tool will apply colors to the original view before converting the presentation into 3D space. Fig 16 depicts the visualization of the results in 3D space and its side view to show the similarity factors in 3D side view.
The 3D visualization allows results to be viewed in many different angles. For example, the 3D side view helps users to see the height of the coordinate value and the clustering pattern of the similarity.

There are two other datasets used to test the tool. For the first dataset, the aim is to compare the presentation of the normal MDS with the enhanced output. The number of participants is increased up to 200 people. The results have shown that colors can differentiate the analysis of the survey. Fig 17 shows the comparison of the visualization.

Another survey is conducted on 98 subjects in order to find out the preference media in enquiring information. Sixteen types of media are listed and the pattern of clustering is determined according to gender. This dataset is used to test the flexibility in using different number of cases and attributes. This tool has successfully processed the dataset and provide the output as shown in Fig 18.

The usability test is also performed to get users’ response on the ease of the visualization. Thirty people are involved and 93.3% agreed that color can enhance the visualization, while 90% have positive feedback towards 3D visualization.
V. CONCLUSION

In this paper, we have presented our work on enhancing the visualization of multiple response survey using a tool called, Murvis. The tool is capable of providing users with flexible views of 2D and 3D results for MDS proximity scaling. Details on its framework, work flow, algorithms and programming structures are also discussed. The attributes used in a survey are identified and classification of results can be performed based on users’ preferences. The tool it able to analyze the output not only by providing colors to the output coordinates, but also flexible clustering of attribute classes in 2D and 3D rotating and zooming. Thus, statisticians and researchers can illustrate a more informative research findings and produce more versatility in their analysis of multiple response survey data.

ACKNOWLEDGMENT

The authors would like to thank the Ministry of Higher Education (MOHE) Malaysia for the financial support through its Fundamental Research Grant Scheme and Research Management Institute (RMI), Universiti Teknologi MARA Malaysia for its continuous support in research.

REFERENCES


