Designing a Study for the Parallel Coordinates Diversity Visualizations Technique

Maduka Ogba
DREU Summer Research intern
1 Trinity Place, 1582
San Antonio, Texas, U.S.A
1-210-954-3462
maduka.ogba@gmail.com

1. OVERVIEW

Many times and in many areas of study, we are very interested in representing the strength of particular objects in a system. We want to know how strong one object is for a certain attribute. However, in those same areas of study and more, we are also highly interested in representing the diversity of strengths among all the objects in our system. For example, in a fantasy football game, although we would like all our players (objects) to be strong in their health (attribute), we want each player to have different skills (diverse) that will enhance the chances of winning the fantasy football game (system)[1].

One other example and the one we will be focusing on for this particular experiment, is that of the college admissions process. Most, if not all, admissions offices in colleges and university require some form of minimum academic standards from all the applicants that they admit. This means that each applicant has to be able to be represented as above the minimum standards for admittance. However, admissions offices are a lot more interested in admitting the pool of applicants that are highly diverse in the range of economic status, race/ethnicity, high school location, academic interests, and perhaps extra curricular activities. The problem is that it is difficult to see this bigger picture without some tool that could help to put all the information about all the incoming students together

Diversity visualization techniques are graphic tools used to help visualize the level of diversity of a set of objects among different attributes. In other words, taking in a set of data, the visualization technique should be able to produce a complete visual description of the distributed strengths of all the objects in a system.

The goal of Dr. Metoyer's diversity research team is to help admissions officers identify and select a diverse group of applicants from the entire applicant pool. The research team has proposed a modified technique called the parallel coordinates diversity visualization technique (henceforth called pc technique). It is believed that this technique emphasizes diversity in a high dimensional data set unlike previously proposed techniques [2, 3, 4].

2. RESEARCH PROJECT

The goal of the research project was to design and create a web-based application that will be used in conducting an experimental study to analyze the pc diversity visualization techniques. As an overview, the program is intended to:

1. Take in data (questions, files, images, and answers) effectively, and efficiently
2. Display the data taken in dynamically to the study taker
3. Allow the study taker to make choices by clicking buttons, navigating through the application, and submitting information.
4. Store all required results and data into the database

The timeline for this project was ten weeks, and the expectation was that some test runs will be ran before the end of the time here.

3. DESIGN OF THE STUDY

The study was designed in tandem with the design of the web program. The study will involve a study taker sitting in

front of a monitor and will go through the whole program with the supervision of a certified test giver.

The study will consist of fifty six questions total. It will involve comparing two different visualization techniques, so each will have twenty eight of the same questions attributed to it. For each technique, six questions will involve comparing two images, three with 100 attributes, and the other three with 1000 attributes, asking which image represents a more diverse set of objects. This will help convey how efficient each method can represent different levels of diversity. Next, there will be a section with one image asking how diverse each of these images is. The participant will also be asked about specific attributes and their degree of diversity. All these will ensure that the participant is able to make good judgments on diversity using the two techniques.

There is a set time limit of 60 seconds on each question. The time limit for each question is displayed on the page however the actual countdown will not show. This is to eliminate any form of pressure from the test and give the participant a fair ground for answering these questions

To avoid participant and experimenter bias, these questions will be reordered in different ways. Each ordering is called a session template and each user is assigned a session from the session templates. For example, the order at which the participant is given one technique before another will be interchanged. In addition, the order at which the participant is shown the 100 attribute question before the 1000 attribute image question will be interchanged. The point of this reordering is to remove any instant where it may be possible for the participant to have some form of bias on a certain question. This is also to remove anything that will skew the overall results of the study.

4. DESIGN OF THE WEB PROGRAM
Before the program was built, we came up with a design for the entire structure. There are three areas of concentration for the web program. There are the forms that support backend usage, those that support frontend inputs, and the database structure.

4.1 Backend Forms
The backend forms are those forms that make it easy for the experimenters and study givers to manipulate data. These involve inputting question descriptions, possible answers, image files and time limits assigned to each questions. There is another form section for managing sessions, which are a predefined ordering of questions for the study, in the database. There is also the managing user section where the administrator could assign user study codes so that each user is uniquely assigned and identified in our database.

4.2 Frontend Forms
The frontend forms are those forms that enable the user to interact with the questions and submit choices and answers for the study. These results are then stored into the database. These forms include the user login page where the user enters her or his personal user code obtained from the administrator, and the actual question display form.

4.3 Database Structure
The MySQL database used is made up of eleven tables that are used to store and retrieve information about the user, the questions, the answers, image URLs, and results. These tables are connected to the forms in different sections and are manipulated from the pup forms to ensure all the data are placed in, and retrieved from, the correct tables.

5. TECHNICAL SKILLS REQUIREMENT
For this project to be done effectively, it was required to master the PHP scripting language. PHP was needed for this project because it is particularly great and efficient at creating dynamic web pages. This was the basis of the web program that was created. Most of the pages needed to be dynamic to fit the criteria that were set in the initial stages.

Also, the knowledge of how to manipulate a MySQL database was absolutely necessary in storing, arranging, and retrieving data effectively. All records of every data were stored in tables in a created database. MySQL was used because it was the most available.

6. OVERALL CONTRIBUTION TO THE PROJECT
The design and creation of the web program was split between another DREU participant and me.

I created the backend forms for entering the questions (and all the other necessary information) into the database. The person entering the forms will be given the option to insert the name of the question, the URL to the images and legend

that they wish to use for the question, the time limit set, and specify what kind of question it is (that is, multiple choice, scale, Yes/No, True/False, 1 line text field, Likert style, or multiple line question). There is also a special option for if they wish to insert the URL of a page that the study will be redirected to. The example of this happening is when the study taker is redirected to a tutorial for a particular technique.

I also worked on the dynamic display of the questions in a form. These question forms are placed in the format that there will be one or two images displayed above the question and the answer. There will be one question per page .The exception to this rule is the Likert questions, which will be grouped together and displayed on one page.

7.   RESULTS

We were able to get a working program done. This program at this stage can allow the experimenter or administrator to enter a unique user ID for a particular user. The user could then log in with the specific user code that the administrator gives, and the user cannot use any other code. This ensures accuracy and security on the whole program. The user can go through the sequence of questions that the experimenter puts together. All the questions have time limits, and once the time limit is reached, a pop up box alerts the test taker that this has occurred and redirects the test taker to another page with a different question. The user also goes through subjective questions which involve the user's perception of the technique.

8.   ACKNOWLEDGMENTS

9.   REFERENCES

1   Pham, T., Hess, R., Ju, C., Metoyer R., 2009. What Does Diversity Look like? Submitted to IEEE InfoVis Poster Session, 2009

2   J. Pearlman, P. Rheingans, and M. des Jardins. Visualizing diversity and depth over a set of objects. IEEE CG&A, 27(5), 2007.

3   E. Kandogan. Visualizing multi-dimensional clusters, trends, and outliers using star coordinates. In Proc. KDD, 2001.

4   R. Kosara, F. Bendix, and H. Hauser. Parallel sets: Interactive exploration and visual analysis of categorical data. IEEE TVCG, 12(4), 2006.