International Journal of Database Theory and Application
Vol.9, No.9 (2016), pp.49-58
http://dx.doi.org/10.14257/ijdta.2016.9.9.05

Comparative Exploration of Features for Data Mining Results by Legend Navigation Interactive Technique

*Muzammil Khan[1], Sarwar Shah Khan[2] and M. Daud Awan[3]

[1,2]*Department of Computer & Software Technology, University of Swat, Pakistan*
[2,3]*Department of Computer Science, Preston University Islamabad, Pakistan*
[1]*muzammilkhan86@uswat.edu.pk*, [2]*sskhan0092@yahoo.com*,
[3]*drdaudawan@hotmail.com*

Abstract

In this age of information, which is available in different forms and these information play imperative part to enhance our knowledge base and so our social life. Information presentation is as important as information itself because to interpret the inside knowledge and interaction with it makes it more effective. This study investigates legend navigation interactive technique for comparative exploration of descriptive data mining results in order to communicate the insight information quickly, easily and in well understandable form. The legend navigation interactive mechanism is applied to two visualization techniques (column charts and bar charts) by performing descriptive data mining task on published Amazon dataset. The experimentation is done with 41 volunteers, selected by simple random sampling technique. The interactive technique is comprehensively analyzed in both visualization techniques considering visualization features. Equate the results with drill down interactive mechanism and discussed the utility of mechanism based on visualization features. The legend navigation approach and drill down interactive mechanism showed better performance in column chart comparatively.

Keywords*: Legend navigation interaction, interactive visualization technique, interactive data mining results, interactivity*

1. Introduction

Enormous data are generated by different sources, these data are of diverse type, stored in various formats and considered to be useless until to convert it into user based required information. Information is the essential part of our lives and the focus is to interpret this information as easily as possible. This interest encourages thinking of new ways to represent information and evolve new techniques. Data mining analysis techniques are of great importance, in order to draw out valuable information from data repositories.

The *Data mining analysis techniques* are the rich innovative technology apply to big data (huge dataset stored in data warehouses *etc.*) for the purpose of extracting knowledge [1-2]. Data mining tasks use to identify patterns at different granularity, abstraction and aggregation. These tasks can be categorized into two groups *i.e.* descriptive and predictive data mining tasks, depends on the type of patterns to be extracted from the dataset [2-4]. The descriptive data mining tasks are employing to data characterization and discrimination; means to summarize the characteristics or features extraction; existing inside the dataset and compare these features. The data characterization and discrimination results can be presented in various forms that are bar charts, pie charts, curves, data cubes, tables, rule form *etc.* [2]. On the other hand, predictive data mining tasks use to infer future happenings on the bases of previous data and the dependency among them. Classification, regression analysis, deviation detection *etc.* are known techniques.

ISSN: 2005-4270 IJDTA

The best way to convey the information protract in the form of results through mining techniques is to visualize the information. These visualized results need to convey the inside easily, effectively and with minimal efforts.

Visualization is an effective tool to depict information and communicate the insight easily. Data or information visualization use for diverse purposes like analyzing the data, exploring with different aspects, discovering new facts, illustrating of these facts, to represent huge data compactly from different viewpoints *etc.* in understandable form [5-6]. To explore the information easily and effectively by different perspective, interaction mechanisms with visualization techniques are utmost important. The study [8] comprehensively discusses different data and information visualization techniques, challenges in visualization, common interactive mechanisms.

The ultimate goal of visualization interaction (user interaction with the visualization technique) is to provide information easily, manipulate the information effectively and effortlessly by the users [7]. To explore the information the users interact with the visualization by different means like mouse over, click, double click, even can add multiple interactive options by mouse right click *etc.* The selection of interactive mechanisms depends on the information to explore, the patterns derived and the characteristics to analyze.

There are several challenges regarding information visualization, discussed in [8]. To make data mining results more productive, the major challenges of information visualization must be considered *i.e.* selection, presentation, interaction and make sure that the visualization features exist in visualization technique.

The selection of visualization technique and interaction mechanism depends on the nature of the data to be visualized. This study is the continuation of the study [16], which discusses an interactive technique "Drill Down Approach" for descriptive data mining results. Drill down approach explores lower level details of the information. Here the study introduces another interactive mechanism "Legend Navigation Approach" for data discrimination (compares the characteristics or features), to make the data mining results more effective and informative.

1.1. Legend Navigation Interactive Approach

The legend navigation approach is best suited interactive technique for data discrimination. Data discrimination means to compare characteristics or features of a single targeted class or with general features of another class or with multiple classes in order to get better understanding. The Figure 1 and Figure 2 are the examples of legend navigation interactive technique.

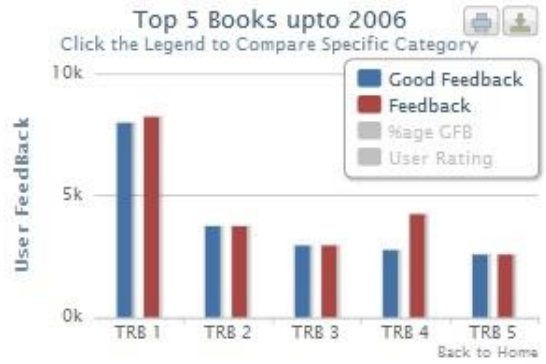

Figure 1. Top Rated Books Up to 2006 on the Bases of User's Feedback

Figure 1 shows the comparison of two features and Figure 2 compares one feature of a target class. With legend navigation interactive technique we can specify multiple features depends on the existance of potential features in the dataset, user requirement and interests.

Figure 2. Best Book Up to 2006 and its User's Feedback

The information provided by the given interactive visualization technique can be represented using four different charts with almost eight backward and forward navigations. On the other hand the stated interactive technique reduces these navigations to two and provides the comparison of all four features of the targeted class by legend navigation interactive mechanism.

2. Experimental Setup

2.1. Visualization Evaluation Technique

The worth, reliability and significance of a research study can be analyzed by choosing a suitable evaluation procedure based on some criteria. Evaluation is one of the substantial steps of research process [9]. The selection of evaluation technique depends on the nature of the research study. A number of techniques are discussed in [11] and categorized in various types based on their utilization in different domain [10]. The study evaluates the effectiveness of interactive technique in information visualization by considering visualization features, as briefly described in Table 1.

Table 1. Visualization Features and its Description

Feature	Visualization Feature Description
Functionality	Functionality means, up to what extent the visualization presents the functionalities compelled by the end user?
Effectiveness	Visualization techniques should be advisable (valuable, meaningful) and convey the insight effectively.
Efficiency	Up to what extent the visualization techniques provide help to the users and lead to achieve the intended or expected result with better performance?
Usability	The visualization techniques provide easy user interaction and should be clear and understandable.
Usefulness	Are the visualization techniques is beneficial, what is the overall user perception?

This study is evaluated by using questionnaire based control experiment. In questionnaire based control experiments the participants are asked to execute a number of task based experiments on functional prototype and fill the feature based questionnaire or conduct individual interviews.

2.2. Participants

The simple random sampling technique is used for evaluating legend navigation interactive technique. In simple random sampling technique each individual has equal chance to be selected for experimentation from the population. The population of the study is based on the dataset and its attributes. 41 volunteers are selected from the population by applying simple random sampling technique and the prototype is presented to them for evaluation. The participant's demographics are summarized on the bases of age and education, as follows;

Table 2. Participant's Demographics on the Bases of Age and Education

Age		
Age Range	Male (73%)	Female (27%)
21-25	19%	14%
26-30	41%	10%
31-35	13%	03%
Education		
Education Degree	Male (73%)	Female (27%)
BS	12%	05%
MS	41%	15%
PhD	20%	07%

2.3. Dataset

In the study the data mining techniques are applied to the Amazon Books dataset published in 2008 [11]. The dataset has 7 years of data from year 2000 to year 2006, size of the dataset is 8.3 GB, contains user's feedback or reviews, helpful feedback, book ratings, user's ranking *etc*. The task is to find out the best book on the bases of some specified criterion.

2.4. Survey Design

To evaluate the study we choose the control experiment evaluation technique based on survey or questionnaire. Quantitative research methods are useful to identify relationships among variables, measure and analyze using statistical techniques [12]. After studying numerous systems [13] conclude that 78% of evaluation is based on surveys. In surveys, questionnaire is used to collect the data using functional prototype over sample of real users selected from population and generalize the results to the whole population [14]. Questionnaire used for the evaluation contained five features, visualization techniques and interactive mechanism, using five-point Likert scale which is commonly used in questionnaires, and is the most widely used scale in survey research [15].

Five-point Likert Scale

5. Strongly Agree **4.** Agree **3.** Neutral **2.** Disagree **1.** Strongly Disagree

The questionnaire contains questions related to different tasks. The main goal of the study is to provide interactive usable visualization techniques for data mining results to ensure the best way to convey information inside the huge data by minimal efforts and help to identify the best book on the bases of different criteria using tasks.

Legend navigation approach in column chart is implemented for task “Best book on the bases of high percent of good feedback with highest feedback” and bar chart is implemented for task “Best book on the bases of high percent good feedback with highest feedback with highest rating”.

3. Result Analysis & Discussion

According to the experimental setup, the task based control experimentation is performed on the functional prototype using questionnaire. In this section we analyze different features of interactive mechanism in visualization techniques on the bases of user’s responses and discussed the results. The results are compared with the result of drill down interactive mechanism [16] with legend navigation interactive mechanism. Interactive mechanism in combination with visualization techniques formulates two combinations.

3.1. User’s Response

The user’s responses for each combination have related questions based on directly and indirectly effective parameters or features of visualization techniques. Statistical measures mean, median and standard deviation are applied to the user’s responses as;

Sample mean

$$\bar{x} = \frac{1}{\mathrm{n}} \sum_{i=1}^{n} a_i$$

Standard deviation

$$S = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

All measures are discussed in detail below for the interactive technique on the bases of user's feedback.

3.2. Legend Navigation Approach in Column Charts

The users are interviewed for legend navigation interactive technique in column chart for all features which are under observation. The user response indicates relatively weak results for effectiveness and usability while rest of the features represents encouraging results. About 84% of the participants agree that the technique compelled the functionality demanded by them (Q 4.3), 48% for effectiveness (Q 8.3), 79% for efficiency (Q 13.3), 43% for usability (Q24.3), and 87% for usefulness (Q 26.3).

Table 3. Statistical Measures for Legend Navigation Approach in Column Charts

	Q 4.3	Q 8.3	Q 13.3	Q 24.3	Q 26.3
Mean	4.03	3.48	3.93	3.41	4.25
Median	4	3	4	3	4
Standard Deviation	0.54	0.49	0.57	0.5	0.63

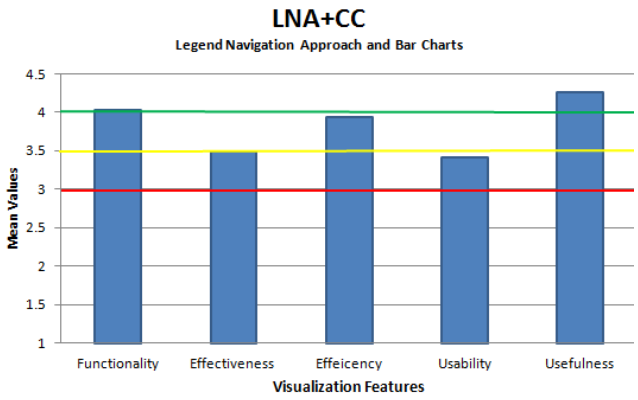

Figure 3. Legend Navigation Approach in Column Charts

Figure 3 illustrates that the legend navigation interactive approach in column charts shows good results for all visualization features. The bar length in the figure indicates the average response of the users while performing tasks. The standard deviation varies (0.49 to 0.63) which shows the stability of the user's responses. The resulted values sit near the red line shows discouraging results, near to yellow line presents fairly good results and near to green line indicates the appreciable responses of users for the visualization features.

3.3. Legend Navigation Approach in Bar Charts

Similarly to investigate the effectiveness of the legend navigation interactive technique in bar charts, the users are interviewed. The user response indicates relatively weak results for effectiveness and poor results for usability while rest of the features represents encouraging results. About 76% of the participants agreed that the functionality feature (Q 4.4) compelled the user requirement, 39% for effectiveness (Q 8.4), 77% for efficiency (Q 13.4), 32% for usability (Q24.4), and 78% for usefulness (Q 26.4).

Table 4. Statistical Measures for Legend Navigation Approach in Bar Charts

	Q 4.4	Q 8.4	Q 13.4	Q 24.4	Q 26.4
Mean	3.83	2.74	3.9	2.45	4.06
Median	4	3	4	2	4
Standard Deviation	0.52	0.55	0.53	0.56	0.67

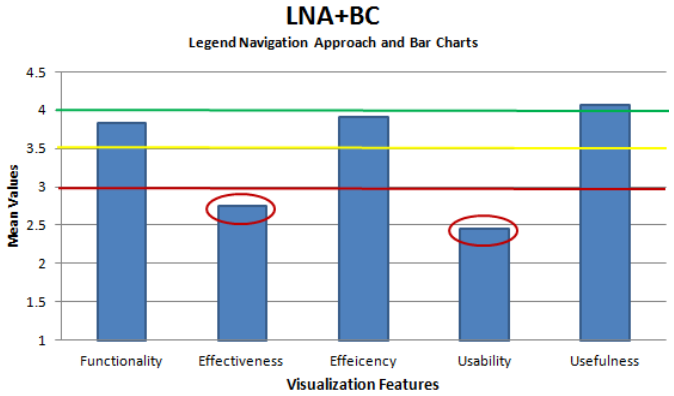

Figure 4. Legend Navigation Approach in Bar Charts

The Figure 4 illustrates the effect of visualization features in legend navigation interactive mechanism in bar charts. After interviewing the participants it has been concluded that features effectiveness and usability represent poor results, it is difficult to efficiently manipulate the interaction in visualization.

3.4. Comparison

Figure 5 clearly illustrates the visualization features in column and bar charts, applying legend navigation interactive technique. The interactive approach in column chart shows better performance for all visualization features while in bar chart the features "effectiveness" and "usability" present poor results.

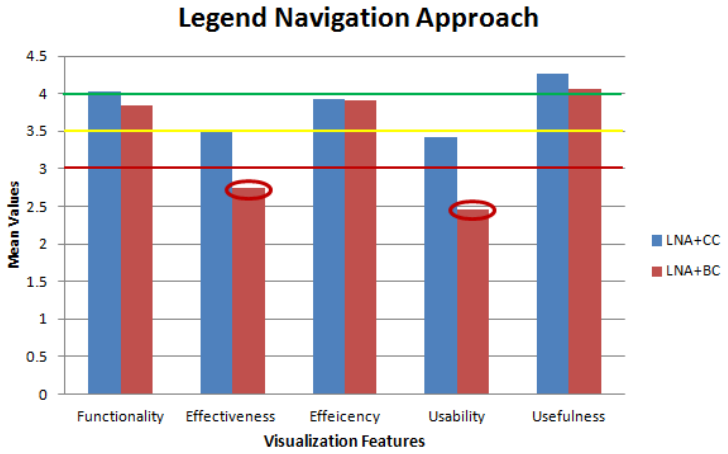

Figure 5. Comparison of Legend Navigation Approach in Column Charts and Bar Charts

Figure 6 is comprehensive comparison of drill down interactive approach and legend navigation interactive approach in both column chart and bar chart information visualization techniques. Both interactive techniques shows promising performance in column charts for all five visualization features and in bar chart the features "effectiveness" and "usability" present poor results (encircled) for both interactive mechanisms.

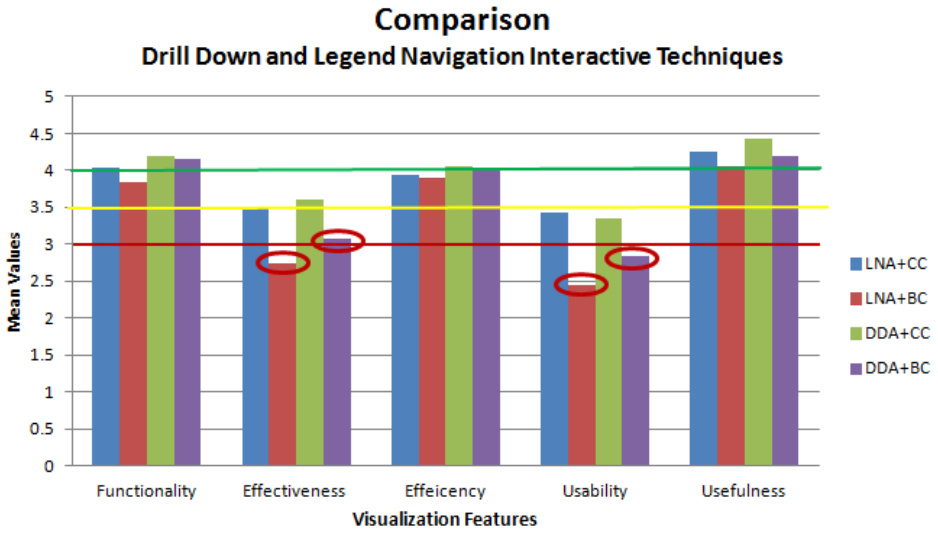

Figure 6. Comparison of Drill Down Approach and Legend Navigation Approach in both Column Charts and Bar Charts

5. Conclusion

This article introduced a new visualization interaction technique for descriptive data mining results in column charts and bar charts. The experimental setup section contains the evaluation technique applied, participant demographics, questionnaire and data source information. The last section discussed the results of interactive technique, include abstract user's response.

The study found that the legend navigation interactive approach in column charts showed the promising results for all information visualization features up to user satisfaction level as compare to bar charts and in bar charts presents weak results for usability and effectiveness because of rare use of the technique. At the end both interactive techniques are compared and presented the utility of each technique with respect to visualization features.

5.1. Future Work

The study can be extended in different dimensions as briefly discussed below;

- To introduce new visualization techniques and related interactivity considering users aspects as well as system aspects.
- To experiment and identify the number of characteristics and target classes that is optimum for comparison in a single chart.
- To extend the study for other purposes or different domains or potential datasets.
- To experiment the interactive techniques on mobile devices in order to enhance the user range and identify the utility of these technique in limited resources.

Acknowledgment

I owe special thanks to my university colleague Mr. Azhar Uddin, Department of English and Foreign Languages; for proof-reading the article and thereby significantly improve the English.

References

[1] U. Fayyad, G. P. Shapiro and P. Smyth, "From Data Mining to Knowledge Discovery in Databases", AI Magazine, vol. 17, **(1996)**, pp. 37–54.
[2] K. Han and Jiawei, "Data Mining: Concepts and Techniques", Morgan Kaufmann Publishers Inc. San Francisco, CA, USA, **(2005)**.
[3] S. Velickov and D. Solomatine, "Predictive Data Mining: Practical Examples", 2nd Joint Workshop, Artificial Intelligence in Civil Engineering, Cottbus, Germany, **(2000)**.
[4] U. Fayyad, G. P. Shapiro, P. Smyth and R. Uthurusamy, "Advances in Knowledge Discovery and Data Mining", MIT Press, **(1996)**.
[5] S. K. Card, J. D. Mackinlay and B. Shneiderman, "Readings in information visualization: using vision to think", Morgan Kaufmann, **(1999)**.
[6] G. J. Kowalski and M. T. Maybury, "Information storage and retrieval systems: theory and implementation", Springer, vol. 8, **(2000)**.
[7] R. Kosara, H. Hauser and D. L. Gresh. "An interaction view on information visualization", State-of-the-Art Report. Proceedings of EUROGRAPHICS, **(2003)**.
[8] M. Khan and S. S. Khan, "Data and information visualization methods, and interactive mechanisms: A survey", International Journal of Computer Applications, Citeseer, vol. 34, **(2011)**, pp. 1–14.
[9] M. K. Trochim, William and J. P. Donnelly, "Research methods: The concise knowledge base", Atomic Dog Publisher, **(2005)**.
[10] R. Mazza, and A. Berre, "Focus group methodology for evaluating information visualization techniques and tools", Information Visualization, IV'07. 11th International Conference, IEEE, **(2007)**, pp. 74-80.
[11] M. Khan, "Interactive Data Mining Results Visualization on Mobile Devices", ISBN 978-3-659-46354-9, LAP Publisher USA, **(2013)**.
[12] N. Jindal, and B. Liu, "Opinion spam and analysis", Proceedings of the international conference on Web search and web data mining, **(2008)**, pp. 219-230.

[13] K. J. Smith, "Quantitative versus qualitative research: An attempt to clarify the issue", Educational researcher, JSTOR, vol. 12, **(1983)**, pp. 6-13.
[14] L. V. Velsen, T. V. Der Geest, R. Klaassen and M. E. L Steehouder, "User-centered evaluation of adaptive and adaptable systems: a literature review", Knowledge Engineering Review, Cambridge Univ Press, vol. 23, **(2008)**, pp. 261.
[15] A. Bryman, "The debate about quantitative and qualitative research: a question of method or epistemology", British Journal of Sociology, JSTOR, **(1984)**, pp. 75-92.
[16] J. A. Gliem, and R. R. Gliem, "Calculating, interpreting, and reporting Cronbachs alpha reliability coefficient for Likert-type scales", Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education, The Ohio State University, Columbus, OH, **(2003)**.
[17] M. Khan, F. Hussain, and I. Khan, "Single Level Drill Down Interactive Visualization Technique for Descriptive Data Mining Results", International Journal of Grid & Distributed Computing. vol. 7, no. 4, **(2014)**.

