

Single Level Drill Down Interactive Visualization Technique for Descriptive Data Mining Results

Muzammil Khan, Fida Hussain and Imran Khan

*Department of Computer Science & Software Technology
University of Swat, KP Pakistan*

muzammilkhan86@uswat.edu.pk, fida@uswat.edu.pk, imran@uswat.edu.pk

Abstract

Information technology plays vital role to enhance our knowledge and improve social life. Information presentation is as important as information itself, and interaction with these information enable one to understand these information quickly and easily. In this article, the information is explore up to one level granularity by introducing single level drill down interactive technique for descriptive data mining tasks results in order to convey inside of the data quickly, easily and effectively. The experimentation being done on Amazon dataset and two information visualization techniques i.e. column charts and bar charts. The interactive technique is comprehensively analyzing in both visualization techniques with respect to the visualization features. The drill down approach in column chart shows better performance comparatively.

Keywords: *Drill down interaction, interactive visualization technique, one level interactive mechanism, interactive data mining results*

1. Introduction

Information is now the essential part of our lives, and the focus is to get that information as easily as possible, which encourage the new ways to think and evolve new advancement. Huge amount of data are generated from different sources, these data are of diverse type, stored in various formats and useless until to convert it into some required information. In order to draw out valuable information from these data repositories, data mining techniques are of great importance.

Data mining is an emerging innovative rich technology encompassing a great deal of potential to draw out new and useful information in massive datasets. In simple word data mining is the process to extract useful patterns or knowledge from large amount of data stored in different information repositories i.e. databases, data warehouses *etc.*, [1, 2]. Data mining system should be able to identify patterns at various granularity *i.e.*, different level of abstraction or aggregation. Data mining tasks can be categorized into two classes depending on the type of patterns that need to be identified, *Descriptive data mining tasks and Predictive data mining tasks* [2-4].

Descriptive data mining tasks are used to characterize the features existing inside the data and compare them. Data characterization (to summarize the general characteristics or features of target class of data) and data discrimination (comparing characteristics or features of one target class to the general features of another target class or set of contrasting classes) are the two sub classes of descriptive data mining. 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].

Predictive mining tasks are those tasks that infer future happenings on the bases of previous data and the dependency among them. Classification, regression analysis, deviation detection *etc.*, are predictive data mining techniques.

The best way to convey information draw out using data mining techniques are to visualize these results in a way that communicate the information easily and effectively.

The main objective of *visualization* is, to find out insight of the data and used to analyze, explore, discover, illustrate, and communicate information in well understandable form. Visualization is used to present huge amount of information coherently, compactly, from different viewpoints, and provides several levels of details [5, 6]. The visualization techniques need some interaction techniques that help to explore this information up to different abstraction. The exploration, analysis, and for the representation of data or information visualization, interactive techniques are exceptionally momentous.

The user should get information easily and effectively understand the information for which the user is looking for, for this purpose provide interaction mechanism that make it possible to manipulate visualization effectively and effortlessly as probable [7]. Users can interact with interfaces or visualization in different ways by means of mouse over, single click, double click, or can add multiple interactive options by mouse right button click. The selection of interactive technique is mostly depends on the data set and the attribute inside that data set. A comprehensive study by [8] has a details discussion about data and information visualization techniques, categories and known interactive mechanisms.

1.1. Challenges in Information Visualization

Beside other challenges discuss in [8]. In the production of productive data mining results visualization the major challenges of information visualization are information selection, presentation, interface interaction and the insurance of visualization features. Briefly discuss below;

Information selection: To select those information which are utmost important to the users.

Information Presentation: To display useful information with the most suitable way by selecting feasible visualization technique.

Interface Interactivity: To manage information by the efficient and effective interactive mechanisms for the users.

Visualization features: To produce effective visualization, techniques need to have the visualization features like functionality, usability, effectiveness, efficiency and usefulness. The questions these features address is depicted in Table 1 and comprehensively discuss in [10].

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?

The visualization technique depends on the nature of the data to visualize. To make the data mining results more effective and informative, this study focus on drill down approach to explore data mining results in different dimensions and granularity with column charts and bar charts which is best suited to descriptive data mining results [2].

1.2. Drill Down Approach

Drill down approach is use to access lower level details of the data from a higher level or drill down is to move from abstract or aggregated information to detailed information. In this study the prototype provides one level granularity. The following Figures 1 and 2 are the example of one level drill down approach.



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



Figure 2. Best Book up to 2006 and its User'S Feedback

2. Experimental Setup

2.1. Visualization Evaluation Technique

Evaluation is another crucial and utmost important step of any research process. Evaluation is the systematic way to analyze and assess the worth, reliability, significance of research study on the bases of some criterion [9]. The visualization techniques can be evaluated by observing the features need to have effective visualization techniques. Evaluation techniques can be categorized into several types based on their utilization in different domains [10]. In [11] contain comprehensive detail about evaluation techniques.

The best way to evaluate this study is to use questionnaire based control experiment. In questionnaire based control experiments the participants are asked to execute a task based

experiment on functional prototype and fill the feature based questionnaire or conduct individual interviews.

2.2. Participant’s Demographics

The study considers the simple random sampling techniques, where each individual of the population have equal chance of being selected as a sample. For the evaluation of the visualization techniques the function prototype is applied to the real users. The study includes 41 volunteer participants. 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. Data Source Domain

In this study the data mining tasks are applied to the Amazon Books data set published in 2008 [11]. The data set have 7 years of data from year 2000 to year 2006, size of the data set is 8.3 GB, contains user’s feedback or reviews, helpful feedback, book ratings, user’s ranking *etc.* The aim 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 identified relationships among variables, measure and analyze using statistical techniques [12]. According to the [13] 78%of evaluation is based on surveys after studying various systems. In surveys, questionnaire is used to collect the data using functional prototype over sample of real users selected from population and generalize the results for the whole population [14]. Questionnaire for this study is based on the five features, visualization techniques and interactive mechanisms, using five-point Likert scale (A Likert scale is a psychometric scale concerned with the theory and technique of educational and psychological measurement) 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 on mobile device to enable one to identify the best book on the bases of different criterion which

exposed by using tasks. Tasks are about to “Find out the best book on the bases of some criteria”, as explained.

Drill down approach in column chart is implemented for mining task “Best book on the bases of getting maximum number of feedback or user response” and Drill down approach in bar chart is implemented for mining task “Best book on the bases of getting maximum number of positive feedback or user response”.

3. Result Analysis & Discussion

According to the proposed framework and experimental set up, the functional prototype is used to conduct questionnaire based control experiment. This section analyzes and discusses the results of the experiments being done for the interactive techniques in column charts and bar charts visualization techniques.

3.1. User’s Response

The user’s responses for both the combinations of interactive mechanism and visualization techniques. Each combination has related questions based on directly effective parameters and indirect parameters or features of visualization techniques. Related measures mean, median and standard deviation are discussed in detail below.

3.2. Drill Down Approach in Column Charts

The usage of drill down approach in column chart shows stable results for all five features. Approximately 85% of the participants agreed with the functionality of this approach (Q 4.1), 65% of the participants agreed for effectiveness (Q 8.1), 85% for efficiency (Q 13.1), 49% for usability (Q 24.1), and 90% show their agreement for usefulness (Q 26.1). Mean, median and standard deviation are given in Table 3.

Table 3. Statistical Measures for Drill Down Approach in Column Charts

	Q 4.1	Q 8.1	Q 13.1	Q 24.1	Q 26.1
Mean	4.19	3.61	4.06	3.35	4.41
Median	4	4	4	3	4
Standard Deviation	0.65	0.6	0.57	0.55	0.56

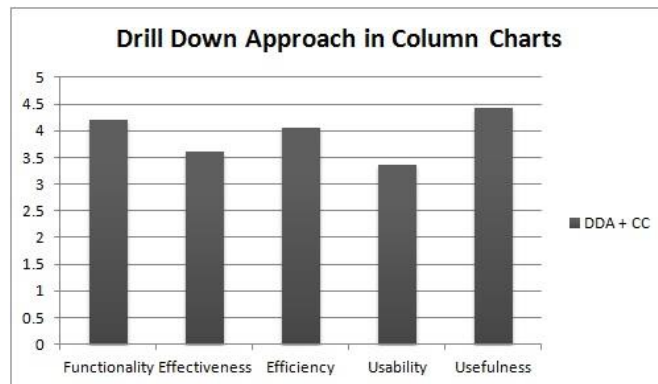


Figure 3. Drill Down Approach in Column Charts

The Figure 3 shows the effectiveness of visualization features for drill down approach in column charts. The length of bars corresponds to the mean values of the user’s responses. Standard deviation ranges (0.55 to 0.65), which shows the consistency of user’s responses. All the features present fair results.

3.3. Drill Down Approach in Column Charts

The drill down approach in bar chart shows weak results for effectiveness and usability while other features show good results. Almost 80% of the participants agree for the functionality feature (Q 4.2), 49% for effectiveness (Q 8.2), 83% for efficiency (Q 13.2), 41% for usability (Q 24.2), and 79% for usefulness (Q 26.2). Mean, median and standard deviation are given in Table 4.

Table 3. Statistical Measures for Drill Down Approach in Bar Charts

	Q 4.2	Q 8.2	Q 13.2	Q 24.2	Q 26.2
Mean	4.16	3.06	4	2.83	4.19
Median	4	3	4	3	4
Standard Deviation	0.68	0.55	0.51	0.52	0.74

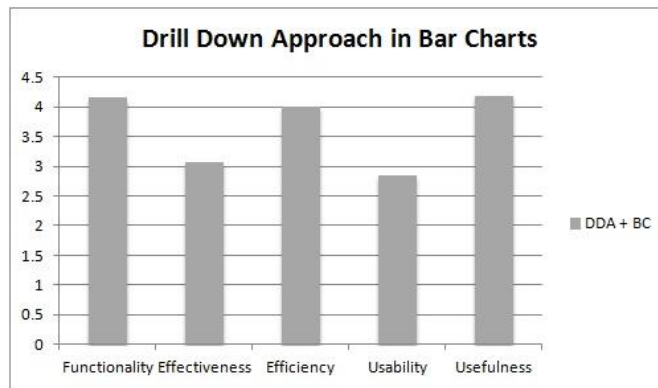


Figure 4. Drill Down Approach in Bar Charts

The Figure 4 shows the effectiveness of visualization features for drill down approach in bar charts. The length of bars corresponds to the mean values of the user’s responses. Standard deviation ranges (0.51 to 0.74), which shows the consistency of user’s responses. The functionality, efficiency and usefulness features show promising results.

3.4. Comparison

The following figure compare drill down approach for column charts and bar charts information visualization techniques. The visualization features show better results for column charts as compare to bar charts. It has been observed that the column chart shows better results because it is common to use comparatively as depicted by Figure 5.

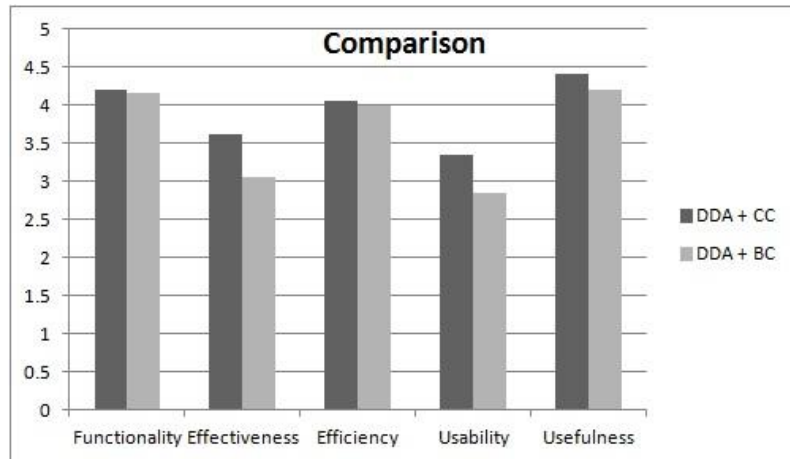


Figure 5. Comparison of Drill Down Approach in Column Charts and Bar Charts

4. Conclusion and Future Work

This article introduced a new visualization interaction technique for descriptive data mining results in column charts and bar charts. The challenges face by attractive information visualization, discussed drill down interactive technique for descriptive data mining result with pictorial example. 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 drill down approach in column charts shows the promising results for all information visualization features up to user satisfaction level as compare to bar charts. Drill down approach in bar charts presents weak results for usability and effectiveness because of rare use of the technique.

The study can be extended as;

- To introduce new visualization techniques and related interactivity for mobile devices considering users aspects as well as system aspects.
- To introduce different levels of abstraction, the study use one level details of the related data, it can be enhanced to two and three level details, depending on the data and user requirements.
- To extend the study for other purposes or different domains or data sets.
- To introduce another alternative interactive technique *i.e.*, legend navigation interactive technique to explore results of data discrimination mining tasks.

References

- [1] U. Fayyad, G. Piatetsky-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. M. Fayyad, G. Piatetsky-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. Helwig 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] Trochim, M. K. William and J. P. Donnelly, "Research methods: The concise knowledge base", Atomic Dog Pub., (2005).
- [10] R. Mazza and A. Berre, "Focus group methodology for evaluating information visualization techniques and tools", Information Visualization, 2007. 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] J. K. Smith "Quantitative versus qualitative research: An attempt to clarify the issue", Educational researcher, JSTOR, vol. 12, (1983), pp. 6-13.
- [14] L. Van Velsen and T. Van 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).

Authors



Muzammil Khan, Lecturer, Department of Computer Science & Software Technology University of Swat. I Received BS Computer Science Degree (Gold Medalist) Session 2005-09 from University of Malakand, MPhil Computer Science Degree Session 2010-12 from Quaid-i-Azam University Islamabad and Currently doing PhD Computer Science from Preston University Islamabad, Pakistan. Research interests are Human Computer Interaction, Usability, Data Mining on Mobile devices, Visualization, Mobile Visualization and Wireless Networks.



Fida Hussain, Assistant Professor, Department of Computer Science & Software Technology University of Swat. I am working on this topic from last 6 months. This work is done in group, in collaboration with my colleagues.



Imran Khan, Lecturer, Department of Computer Science & Software Technology University of Swat. I am working on this topic from last 6 months. This work is done in group, in collaboration with my colleagues.