

Overview of Big Data Visualization

Sagargouda S.Patil

Department of Computer Science and Engineering, PES University, Bangalore.

[Email-id:sagar.cs.kle@gmail.com](mailto:sagar.cs.kle@gmail.com)

ABSTRACT- Day by day, data is growing at a faster rate in every field. Data which are very large in size and cannot be processed using traditional database systems are known as big data. Before the days of data visualization and big data, businesses struggled to grasp concepts such as customer behavior and fraud. Big data helped these insights come to light. Because of the way the human brain processes information, using charts or graphs to visualize large amounts of complex data is easier than poring over spreadsheets or reports. Data visualization is a quick, easy way to convey concepts in a universal manner and can be experimented with different scenarios by making slight adjustments. This paper discusses about big data Visualization, its challenges and tools.

Keywords— big data, big data visualization, data visualization, interactive visualization, visualization tools.

I INTRODUCTION

Big data [1] Visualization is a representation of data in graphical format or pictorial format. It enables us to evaluate things visually so that we can identify new patterns. There are some key points to be considered for big data visualization such as: (1) Meta data (2) Visualization tools must be interactive. There are some myths related to visualization are as follows:

- Every data must be visualized
- Good data should be visualized
- Visualization will be always evident right decision
- Visualization leads to the inevitability

Big data visualization is not easy as compared to traditional data sets. In larger data set visualization, many researchers use feature extraction and modeling techniques to reduce data size. The Following table shows the benefits of big data visualization according to a survey report [1].

Benefits	Percentages (%)
Improved decision-making	77
Better ad-hoc data analysis	43
Improved collaboration/information sharing	41
Provide self-service capabilities to end users	36
Increased return on investment (ROI)	34
Time savings	20
Reduced burden on IT	15

Table 1.Benefits of visualization tools

II LITERATURE SURVEY

This section focuses on the conventional visualization methods and how big data visualization came into the picture.

Conventional visualization method [2] is a traditional method which uses tables, bar chart, pie chart, histogram, Venn

diagram, flow chart, entity relationship diagram for data visualization. These methods do not work well for huge datasets containing structured and unstructured data respectively. Given below are the 2 additional methods which are used often in conventional visualization.

- (1) **Parallel Co-ordinates:** It is used to plot data elements across multi-dimensions.

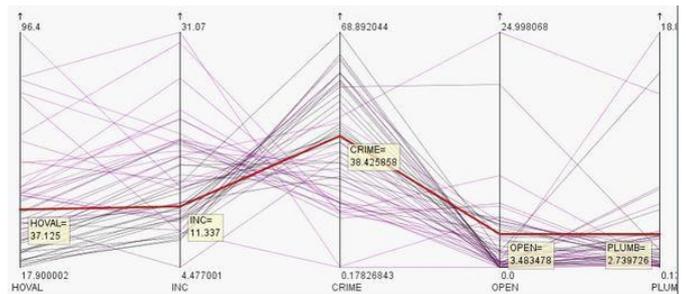


Fig 1. Parallel Coordinates

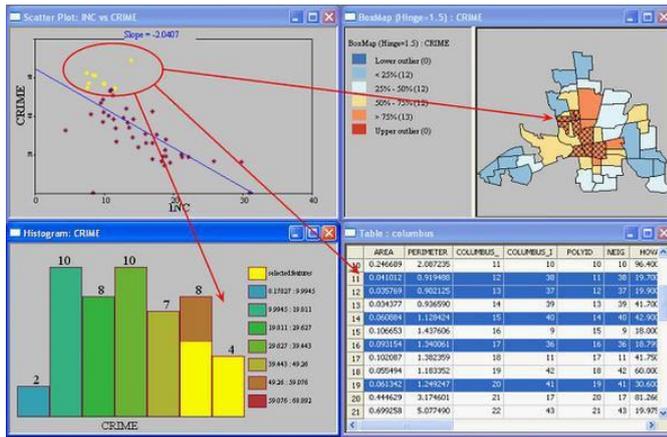
- (2) **Semantic network:** It is a logical relationship between different concepts.

Conventional data visualization conveys that these visualizations could use in business analytics and other data systems, but cannot be used for the scientific process.

Visualizations are not only static, but also they can be interactive. Interactive visualization [4] help in zooming (zoom in or zoom out), focus, overview and detail.

Below given are the Steps for interactive visualization.

- 1] Selecting
- 2] Linking
- 3] Filtering
- 4] Remapping



A. Fig 2. Interactive brushing and linking

Big data visualization is important because of the following reasons.

- It provides clear knowledge about patterns of data
- Detects hidden structures in data
- Identify areas that need to be improved
- Help us to understand which products to place where
- Clarify factors which influence human behaviour

III. CHALLENGES OF BIG DATA VISUALIZATION

This section describes the challenges of big data visualization [3][10]. Visualization based methods of big data challenges are presented by **four Vs** as follows:

- **Volume**= the methods are developed to work with a large number of datasets.
- **Variety**= the methods developed to combine as many data sources (structured +unstructured data).
- **Velocity**= with the methods, business can replace batch processing with real time processing.

Visualization of heterogeneous big data (structured, unstructured, semi-structured) is a biggest problem. Because of the big data size, parallelization is a challenge in visualization and Speed is a desired factor for big data analytics. So, effective data visualization is a key point of the discovery process in the field of big data [3].

Following are some of the problems of big data visualization.

- **Visual noise:** Objects which are present in the data set are related to each other. So

users cannot divide them into separate objects on screen.

- **Information loss:** There is information loss while reducing data sets.
- **Large image perception:** Visualization methods are not only related to aspect ratio and resolution but also it takes physical perception into consideration.

Potential solutions to some challenges and problems of data visualization are [6][7]:

- **Meeting the need for speed:** One solution is to deal with **Hardware** by increasing memory and massive parallel processing can be used.
- **Understanding the data:** Proper domain Expertise is solution.
- **Displaying meaningful results:** solution is to cluster data into smaller groups that are visible effectively.
- **Dealing with outliers:** solution is to remove outliers from data or create separate chart for outliers.

IV. BIG DATA VISUALIZATION TOOLS

This section discusses about different big data visualization tools most of which run on the Hadoop platform [3][10] having following modules:

- Hadoop Distributed File System(HDFS)
- Hadoop Map Reduce
- Hadoop YARN(Yet Another Resource Negotiator).

Some of the big data visualization tools available are as follows:

- **Statistical Analysis System(SAS) Visual analytics:** It uses intelligent auto charting to create best possible visual data we select.
- **Polymaps:** It is a free java script library for making interactive maps in modern web browsers.
- **Flot:** It is a java script plotting for JQuery. It supports visualization for interactive charts, data points, zooming and stacked charts.
- **Google Maps:** This tool enables developers to build visual mapping programs for any website or application.
- **Microsoft Excel:** It is also platform for data analysis and creates visualization and represents data in the form of charts and graphs.
- **R-Project:** **R** is a free software environment for statistical and graphics. It helps in data manipulation, calculation and analysis.

- **Tableau Public:** It is a simple and user-friendly tool to create an iterative data visualization quickly and puts them on the website [5][9].

CONCLUSION

The strong and growing trend toward visualization-based data discovery tools is expected to continue in the years to come, as more businesses seek better, more cost-effective ways to derive meaning from their big data. Big data Visualization provides an immense opportunity not only to manage the growing volume, variety, and velocity of new and existing data but also to turn that data into value. It allows access to challenging data sets, it allows exploration, can be fun, and provides useful information in an efficient way. This paper discusses big data visualization basics, tools, challenges which will benefit future researchers of big data visualization to continue their work based on the concepts discussed here.

ACKNOWLEDGEMENT

I am grateful to the valuable guidance provided by Prof. Nitin V Pujari, HOD, Department of Computer Science & Engg, PES University, Bangalore and Dr. Usha Sakthivel, HOD, Department of Computer Science & Engg, RRCE, Bangalore and special thanks to Prof. Rashmi G, Department of Computer Science & Engg, Nandi Institute of Technology and Management Sciences, Bangalore, Prof. V.R. Badri Prasad, Dept of Computer Science & Engg, PESIT, Bangalore and Prof. Dinesh Singh, Dept of Computer Science & Engg, PESIT, Bangalore for providing me a valuable help and support.

REFERENCES

- [1] M. Khan, S.S. Khan, Data and Information Visualization Methods and Interactive Mechanisms: A Survey, *International Journal of Computer Applications*, 34(1), 2011, pp. 1-14.
- [2] Intel IT Center, Big Data Visualization: Turning Big Data Into Big Insights, White Paper, March 2013, pp.1-14.
- [3] V. Sucharitha, S.R. Subash and P. Prakash, Visualization of Big Data: Its Tools and Challenges, *International Journal of Applied Engineering Research*, 9(18), 2014, pp. 5277-5290.
- [4] P. Simon, The Visual Organization: Data Visualization, Big Data, and the Quest for Better Decisions, *Harvard Business Review*, June 13, 2014, pp. 1-8.
- [5] C.L. P. Chen, C.-Y. Zhang, Data-intensive applications, challenges, techniques and technologies: A survey on Big Data, *Information Sciences*, 275 (10), August 2014, pp. 314-347.
- [6] B. Porter, Visualizing Big Data in Drupal: Using Data Visualizations to Drive Knowledge Discovery, Report, University of Washington, October 2012, pp. 1-38.
- [7] T. A. Keahey, Using visualization to understand big data, Technical Report, IBM Corporation, 2013, pp. 1-16.
- [8] P. Fox and J. Hendler, Changing the Equation on Scientific Data Visualization, *Science*, 331(11), February 2011, pp. 705-708.
- [9] I. B. Otjacques, UniGR Workshop: Big Data- The challenge of visualizing big data, Report, Gabriel Lippmann, 2013, pp. 1-24.
- [10] H. Childs, B. Geveci, J. Meredith, K. Moreland, C. Sewell, E.W. Bethel, T. Kuhlen, W. Schroeder, Research Challenges for Visualization Software, Joint Research Report of Lawrence Berkeley National Laboratory, Oak Ridge National Laboratory, Sandia National Laboratories, Los Alamos National Laboratory, RWTH Aachen University (Germany), May 2013, pp. 1-11.