Name : Rajendra Rakha Arya Prabaswara

Kelas : 3H

NIM : 1941720080

QUIZ BIG DATA

1. **Titles :** DeepEye : An automatic big data visualization frameword
2. **Authors :** Xuedi Qin, Yuyu Luo, Nan Tang, Guoliang Li
3. **Publisher :** TUP
4. **Year :** 25 Januari 2018
5. **URL :** https://ieeexplore.ieee.org/document/8268737
6. **Problems :**

   Data visualization converts data into images to help understand data an invaluable tool for explaining the importance of data to visually inclined people.

   In this paper,

   they presented the DEEPEYE system to answer this challenge. This system solves the first challenge by training binary classifiers to decide whether a particular visualization is good for a given data set, and by using supervised learning to rank models to rank the above good visualizations. Their proposed system tackles the third challenge by incorporating a database of optimization techniques for sharing calculations and pruning.

7. **Methods**

   DeepEye System

   In this section, they goal is to present the techniques we envision, which can solve the problem under study, which consists of automatically generate compelling stories of a given dataset.they propose a method to solve this problem by seamlessly integrating techniques from two perspectives namely perception by using machine learning based approaches, and the efficiently guessing of human perception from all possible data transformations by using database optimization techniques.

   Method

   1. **Visualization verification**

      From Training Dataset. they first collect tables that can be used for meaningful visualizations. For each table, we ask users to manually select two columns to generate different visualizations. Then, the users will rank them manually.

2. **Visualization search space**

For simplicity, they discuss the following five operators for one relation:

- Sort: sort the numerical values in a column
- Bin: partition, the values into different buckets
- Group-by: group the categorical values in a column
- Aggregate: use the aggregate functions, such as sum, min, and count, to compute a value for each group;
- None: do nothing to the original data.Naturally, for the easy understanding of some trend, they want some scale domain, e.g., x-scale, to be sorted.

Interactive Data Visualization Systems. There has also been work on building Data Visualization Management systems (DVMSs). Ermac is a system that compiles an LVP to a Physical Visualization Plan (PVP), where PVP can leverage optimizations of DBMSs.DeVIL extends Ermac by an SQLlike language to support interactive visualization.

DeVIL captures user-generated low-level events (e.g., mouse down, move, and up) as event streams, and extract compound events (e.g., mouse drag) as patterns, which are then treated as database views used for interactive visualizations.

Lyra is an interactive environment that enables custom visualization design without writing any code.They can be easily integrated into DEEPEYE after automatically generating visualizations, for user tuninand interactively using our system . SeeDB searches for the possible visualizations that have large deviations from some reference. This is then abstracted as a utility metric to quantify the interestingness.

8. **Conclusion**

In the proposed study, they have presented DEEPEYE,which is a novel data visualization system for automatically visualizing a given dataset in a compelling manner, after transforming the dataset using a combination of operators. thry obtained promising preliminary results using real-world data and real users. This indicates the potential of DEEPEYE in automating data visualization. Finally, they forecasted research challenges in the implementation of the proposed system.

9. **DOI**

10.26599/BDMA.2018.9020007

Electronic ISSN

209-0654

References

- VizCertify: A Framework for Secure Visual Data Exploration
- Multi-model Investigative Exploration of Social Media Data with BOUTIQUE: A Case Study in Public Health
- CrowdRL: An End-to-End Reinforcement Learning Framework for Data Labelling
- Human in the Loop: Industry 4.0 Technologies and Scenarios for Worker Mediation of Automated Manufacturing
- CrowdChart: Crowdsourced Data Extraction From Visualization Charts