Towards a holistic workflow of visual data exploration

Doris Jung-Lin Lee, Aditya Parameswaran

Abstract

test abstract

1 Introduction

2 Precise Visual Querying

Intention driven task-based querying (Precise search)

2.1 Challenges

Theres a large space of possibilities, manual search is tedious. Either using one-size-fits-all statistics, templates, heuristics as a solution or problem only applicable to a subset of analytic tasks[**?**, **?**]. Propose VQS as a solution[**?**]. The goal here is to help novice submit precise queries without SQL background, easy to use interface. Our study found that VQS does more than just this, but still not enough.

2.2 Effortless Data Exploration with *zenvisage*

- Precise Search Fail to understand intricacies, need more expressivity/flexibility.
- No perfect training workload, real-world data + task is noisy and complex.

2.3 Challenges Ahead

- Precise Search Fail to understand intricacies of user need/intent, need more expressivity/flexibility for querying.
- No perfect training workload, real-world data + task is noisy and complex.
- towards more holistic model for insight discovery

Bulletin of the IEEE Computer Society Technical Committee on Data Engineering

3 Hypothesis Formation

3.1 Supporting cycle of visual analysis

- Essential ingredient in facillitating intelligent vague querying and exploration.
- This is a human process (cite [**?**],[**?**])
- Iterative Hypothesis Exploration/Refinement : argue that the following properties is important to sustain this cycle of visual analysis

3.2 Towards 3Is of rapid hypothesis generation support

Towards 3Is Interactive, Iterative, Informative Interactive (how natural is it to move between analysis steps, facilitate fluid analysis and not get stuck) : interactivity, feedback (latter is quite unexplored), and recommendation, expressivity (how easy is it to express what to do via interactions) and diversity of actions that could be performed. Iterative: query refinement, dialogue (not a one-shot query) Informative: not just task-based interestingness but more explanation based (causality, introduce distribution awareness notion in viz-sum)

4 Intelligent Search

Accounting for user interaction, mental models. More global objective taking into account user with the goal of dataset understanding rather than task completion.

4.1 Challenges

- Problem of cold-start recommendation
- Inferring user intent in querying and context is important (both in terms of user input and what is recommended)
- tools can not assume user has querying intention. exploration without intention, user dont know what they are searching for –¿ Recommendation.
- The important thing here is identifying what should be done by the system v.s. requested from user. Inappropriate choice of these will result in lack of expressibility and user feeling lack of control of analysis, limiting exploration.
- Need for a unified framework of inference to take all of these into account (e.g. natural language, etc)

4.2 STORYBOARD: Navigating Through Data Slices with Hierarchical Summary of Visualizations

5 Towards Dataset Understanding

5.1 Challenges

- Problem of cold-start recommendation (as discussed earlier use may not always know what to query for)
- Within a dataset, structure and provenance is essential to help users navigate and provide users with sense of coverage and completion. This is an important but underexplored area. (viz-sum, Sarvghad et al 2017)
- schema and attribute understanding (coverage, etc)

5.2 Examples

- understanding distributions (distribution awareness)
- providing overview recommendations (representative trends and outliers)

6 Concluding Remarks