Understanding and Exploring: Recommendations, Provenance, and Open Data

Rachel Pottinger
University of British Columbia
About this talk

• This talk is a mix of an overview of what my students and I (and then the group at large) are currently doing and where I’m hoping to collaborate with you all

• As such, if you see a spot where you have input, please let me know – I’d love to talk about it later
Exploring and understanding data (in 4 parts)

- **Exploration**: recommend items beyond the popular items in recommender systems
- **Exploration**: recommend regions of data to users of numerical data
- **Understand**: help non-DBA users understand data provenance information
- **Understand**: help users understand open data

Rachel Pottinger
http://www.cs.ubc.ca/~rap
**Exploration:** Recommend long tail items (joint work with Zainab Zolaktaf)

- Standard recommender systems algorithms tend to emphasize popular items
- This tends to cause recommendation consumers to only find things they already know
- But most items are “long tail”
Exploration: Recommend long tail items (joint work with Zainab Zolaktaf)

- Our work explores the trade offs between accuracy and coverage using a framework that models users’ long-tail novelty preferences
- We conduct thorough experiments on these issues, including looking at how density of data impacts the results
- See her poster!
Understand: help users understand data provenance (joint work with Omar AlOmeir)

- Database researchers have done a great job of exploring different provenance definitions and how to calculate it.
- However, this information is difficult to understand by non-DBA users, which makes it hard for users to trust their data.
- We created a desirable set of features for provenance exploration systems and implemented such a system.
- Our case study was on Global Legal Entity Identifiers.
- We’re looking for more data.
Understand: help users understand open data (joint work with Janik Andreas)

- Governments are increasingly creating open data sites
- However, these open data sites are hard to use – it’s hard to find the data that users are looking for
- We’re doing a case study on local data to look at some common open data issues:
  - Quality – granularity and details of available data
  - Metadata and data formatting
  - Availability and completeness
The broader group context

• In addition to myself, there are two other research faculty in our group
  • Laks Lakshmanan
  • Raymond Ng
Laks Lakshmanan

- Information Propagation in Social Networks and Media.
- Recommender Systems
- Data Cleaning and Data Quality Management → Emphasis on Big Data Streams
- Discovering and combating filter bubble
- Fake news detection and intervention
- Students and postdocs
  - PhD: Glenn Bevilacqua, Prithu Banerjee, Sharan Vaswani (joint with Mark Schmidt)
  - MSc: Alexandra Kim
  - Postdoc: Ezequiel Smucler (joint with Ruben Zamar, Statistics)
• Develop preventive, diagnostic or prognostic biomarkers to fight against heart, lung and kidney failures as half-time CEO of the PROOF Centre of Excellence for the Prevention of Organ Failures.

• Text mining with Giuseppe Carenini: create meta data, such as natural language summaries, to facilitate access e-mail, blogs, meeting minutes, etc.