A System for eCommerce Recommender Research with Context and Feedback

Sean Pfister

RichRelevance

http://code.richrelevance.com

sean@richrelevance.com

Agenda

- Introduction to {rr} and RecLab
- Review of context and data challenges in recommender research
- Overview of the RecLab approach
- RecLab models of behavior and context
- RecLab computational model
- RecLab in the real world

Introduction to {rr} and RecLab

- RichRelevance, a.k.a. {rr}, is the leading provider of product recommendations for eCommerce merchants.

- {rr}'s top five merchant partners are in aggregate 60% larger than Amazon.com [Comscore, Nov. 2010].

- RecLab is {rr}'s new approach to enabling recommender researchers to participate in algorithm development and optimization.

reclab

Evaluating Recommenders

- Offline metrics – widely used
  - Precision/Recall
  - Diversity
  - RMS error

- User-centric metrics – emerging
  - Session length
  - Click-through rate
  - Conversion
  - Overall sales lift

reclab

The Critical Role of Context

- Relevance is fundamentally contextual
  - What is relevant changes radically from click to click
- Alignment of recommendation algorithms and context is critical to performance
  - Even more so when merchandising objectives are considered.

reclab

Context I: TV Item Page

- Focus is on discovery and comparison.

Context II: TV in the Cart

- Focus is on accessories
- Competing TVs are gone

LG BD550 Network Blu-ray Disc Player

Today: $149.99
Compare at: $159.97
You Save: $9.98 (6%)
★★★★★

Mount-It! Dual Arm Articulating Flat Panel Monitor Mount for 13 to 30-inch Screens

Today: $24.99
★★★★☆

Sylvania SYL1542EMXC Full Motion 15 inch - 42 inch Plasma/LCD TV Corner Mount

Today: $58.49
★★★★☆

LG BD570 Blu-ray Player

Today: $171.99
Compare at: $197.00
You Save: $25.01 (13%)
★★★★☆

reclab

Context III/IV: Product Type

- You can convince me to buy a second relevant DVD.

- I will only ever buy one washer every 5-10 years.

reclab

The Role of *In-Situ* Feedback

- A key motivating factor in user studies of recommenders.

- Observation of critical intermediate metrics (e.g. click-through) provide important insights.

- Downside is user studies are complex, expensive, and necessarily small scale.

reclab

Data Access Issues

- Production-scale data sets are not widely available.

- Merchants are reluctant to share
  - Competitive concerns
  - Privacy issues can be untenable, e.g.:
    - AOL
    - Netflix Prize 2

Performance Issues

- Model builds must be scalable to billions of raw data points.

- Runtime calls to services without strict performance SLAs are problematic
  - Lack of responsiveness can offset the benefit of recommenders

richrelevance

reclab

The RecLab Approach

If you can't bring the data to the code, bring the code to the data.

Specifically

- Provide and API to access data and build models;

- Provide an environment with billions of real data points to build models;

- Provide runtime APIs to serve recommendations to live shoppers;

- Provide a synthetic environment and data for development and debugging.

richrelevance

reclab

RecLab Context Data

- Click-stream of pages
  - Home, category, search, item, add to cart, purchase, rating, ….

- Annotated with instrumentation data
  - Date/time, user id, session id, search terms, search results, items added to cart, cart contents, ….

Data Access and Performance

- Synthetic tests run efficiently on a single machine.

- Production model builds happen on large-scale {rr} fleet.

- Production runtime served by {rr}'s distributed low-latency architecture.

- Private data remains private and secure.

- Aggregate results are provided to researchers.

Developer Support

- Familiar map reduce programming model
- Detailed tutorial and examples
- Quick start on a single machine with zero configuration
- Synthetic data sets for testing
- YAML (human-readable) data format
- Utilities to convert from other data formats to reproduce existing results
- All code is open-source Apache 2 licensed.

reclab

Summary

A call to academic and research teams: do what you love doing with RecLab and you could win a significant prize!

The Challenge: Generate measurable lift over RichRelevance's existing algorithms on an IR Top 50 online retailer's site.

Contest rules, participating retailer, percentage lift increase, terms and conditions available online.

- ***Interested parties may sign up to receive information at http://code.richrelevance.com/reclab/follow.html***

Thank You

Downloads, documentation, examples, and tutorials at:

http://code.richrelevance.com
Twitter: @RecLabs