



# **A System for eCommerce Recommender Research with Context and Feedback**

Sean Pfister

RichRelevance

<http://code.richrelevance.com>

[sean@richrelevance.com](mailto: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.

# Evaluating Recommenders

- Offline metrics – widely used
  - Precision/Recall
  - Diversity
  - RMS error
- User-centric metrics – emerging
  - Session length
  - Click-through rate
  - Conversion
  - Overall sales lift


# 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.


# Context I: TV Item Page

- Focus is on discovery and comparison.

Photos



View larger



## LG 42LD450 42-inch 1080p LCD TV

Rating 4.8 ★★★★★ 5 reviews  
[Read Reviews](#) [Write a review](#)

**Today: \$639.99**

Receive 5% Back and Free Shipping With Club O

0% Intro APR\* [View Credit Card Details](#)

Item#: 13095280

A great way to get into high-definition entertainment, the affordable 42-inch LG 42LD450 LCD HDTV offers 1080p Full HD resolution for crystal-clear detail, multiple AV Modes to optimize the picture depending on what you're watching.

**High Sell Out Risk**

Quantity:

Warranty:  
[Select your Platinum Protection Plan](#)

[Add to Cart](#)

[Like](#) [Be the first of your friends to like this](#)





[Details](#) [Shipping & Returns](#) [Payment Methods](#)

A great way to get into high-definition entertainment, the affordable 42-inch LG 42LD450 LCD HDTV offers 1080p Full HD resolution for crystal-clear detail, multiple AV Modes to optimize the picture depending on what you're watching, and an amazing 100,000:1 dynamic contrast ratio for more stunning colors and deeper blacks. Other features include a 4ms (millisecond) response time, capability to play MP3 audio files and JPEG photo slideshows from a USB drive, stereo speakers with 20 watts of power and Dolby Digital decoding, and three HDMI inputs.

### Compare to these Retailers [View this list](#)


J&F **\$899.00** [Compare](#)



### People Who Viewed This Also Viewed

 <p>LG 42LD450C 42-inch 1080p LCD TV <b>Today: \$671.94</b> Compare at: \$876.32 You Save: \$204.38 (23%) ★★★★★</p>	 <p>LG 55LE5400 55-inch 1080p 120Hz LED TV <b>See price in cart</b></p>
 <p>Panasonic TCP42C2 42-inch 720p 600Hz Plasma TV with HDMI Cable <b>Today: \$527.99</b> Compare at: \$599.00 You Save: \$71.01 (12%) ★★★★★</p>	 <p>LG 37LD450 37-inch 1080p LCD TV <b>Today: \$533.99</b> Compare at: \$699.97 You Save: \$165.98 (24%) ★★★★★</p>

# Context II: TV in the Cart

- Focus is on accessories
- Competing TVs are gone



			
<b>LG BD550 Network Blu-ray Disc Player</b> <b>Today: \$149.99</b> Compare at: \$159.97 You Save: \$9.98 (6%) ★★★★★	<b>Mount-It! Dual Arm Articulating Flat Panel Monitor Mount for 13 to 30-inch Screens</b> <b>Today: \$24.99</b> ★★★★★	<b>Sylvania SYL1542EMXC Full Motion 15 inch - 42 inch Plasma/LCD TV Corner Mount</b> <b>Today: \$58.49</b> ★★★★★	<b>LG BD570 Blu-ray Player</b> <b>Today: \$171.99</b> Compare at: \$197.00 You Save: \$25.01 (13%) ★★★★★

# 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.





# 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.

# 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

# The RecLab Approach

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

# Specifically

- Provide an 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.

# 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.



# 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**

