Recommending for Impact: Intentions, Algorithms, and Metrics

Joseph A. Konstan
konstan@umn.edu
www.grouplens.org
A Bit of History

» Researchers
• MAE, MSE, RMSE
• Correlation (rate/pred)
• Error Rates
• Retention Rates
• Top-k Metrics
• Survey Preferences

• Most of all: statistically-significant improvements!

» Businesses
• Click-through rate
• Conversion rate
• Lift

• Customer return and retention rates
• Time-on-site

• Most of all: customer engagement, retention, and revenue
This Isn’t the Talk I had Prepared

» I became upset during the conference
  - Research papers showing results that were only better under non-existent circumstances – and worse otherwise!
  - Research papers making assumptions that had no grounding in reality, with no recognition or path towards generalizing.
  - Papers claiming important and statistically significant results so tiny it is hard to see how they may to actual impact.
Three Views on Impact

» Business impact of a recommender system (or innovation)

» Impact of a research innovation

» Impacts on users (individually or collectively) beyond business
Common Ideas

» Algorithms are just one part of a complex system
  • Domain
  • User base
  • Available data
  • User interface
  • Broad usage context …. 
Common Ideas

» Impact is always compared with some reasonable alternative or baseline
  • We’re almost entirely past the age of being impressed that we can beat random
  • The relevant alternative is not always obvious
  • Recommendations come with a cost: computation, space/time/attention
    • E.g., if we took “comparative products” off the page, would more people look deeply at and buy this product? Would they be happier?
My Goal Here

» Not to:
  • Share research results
  • Offer brilliant and tidy solutions

» Instead to offer:
  • Some provocative thoughts
  • An articulation of what we might already agree upon
  • A bit of critique of where we are, and hope for where we might go
Business Impact

» This is an area where we’re actually somewhat good, but often have too much tunnel vision

• What’s good – reduction of value to terms of art in business.
• How do we move from immediate impact towards lifetime value?
• Incorporating a broader view of possible “dividends” of effective recommendation.
• Beyond the single firm? Results in shareable and aggregable form.
Research Impact

» Getting beyond inward-looking metrics
» Greater access to experimentation, user impact
» Studying recommender impact in context of full system
  • What are characteristics of use?
  • What are realistic alternatives?
Broader Impact

» How do we consider other benefits, such as:
  • Diversifying somebody’s perspective
  • Connecting people
  • Improving a whole industry

» And where do we measure/consider broader negative impact?
  • Disproportionate/biased recommendation
A Few Other Thoughts

» I believe in metrics … somewhat
  • Never singular – need multiple perspectives
  • Need to be careful to capture meaningful differences, meaningful baselines

» And we clearly need new ones
  • User engagement, user experience, user task success, site/corporate brand and reputation, …
A Few Other Thoughts

» Industry believes there is tremendous impact; how do we quantify this across the field, and trace it back to the research innovations that spawned it?
  • Critical for justifying continued investment in research, both from industry and from public sources
  • “Tire tracks” examples that trace specific wins back to funding sources that contributed
A Few Other Thoughts

» The role of knowledge, theory, expertise in an era of data …

• How do we ensure that knowledgeable alternatives force our algorithms to consider incorporating knowledge
  • Understanding humans, domain, features, heuristics, etc.

• Examples: coins, decision-making, …
Impact is Hard Word

» Data makes statistical significance free

» How to encourage, reward research and presentations that address impact

» Not hard for its own sake, but valuing impact