Modern Recommendation Techniques in the Real World

Olivier Koch, Flavian Vasile

Criteo Research

June 28, 2018

criteo

About the speakers

- Olivier Koch, Machine Learning Lead, o.koch@criteo.com
- Flavian Vasile, Lead Research Scientist, f.vasile@criteo.com

http://cail.criteo.com

Course overview

- Introduction to the Problem of Recommendation
- Part 1. Classical approaches
- Part 2. Deep Learning approaches
- Part 3. Causality in Recommendation

Workshop agenda

- Intro - 30'

- Classical ML Approaches - 30'
- Hands-on application - 1h

- Neural networks for Recommendation - 1h
- *Lunch break*
- Hands-on application - 1h

- Causal Recommendation - 45'
- Hands-on application - 45'

Introduction to the Problem of Recommendation

The task

Given user-item interactions, find the top-k best items for each user

This is different from minimizing the RMSE of a predictive model trained on a user $\times$ item matrix.

Examples

- Movies: Spotify
- Clothes: Zalando
- Retail: Amazon
- Restaurants: Google Maps
- Courses: Coursera

Implicit vs explicit feedback

- Explicit feedback provides positive and negative examples (e.g. user ratings)
- Implicit feedback provides positive examples only (e.g. clicks, views, $\cdots$)
- Implicit feedback is much more abundant but incomplete
- You may have the choice (e.g. Netflix)

The Exploration / exploitation trade-off

- Exploitation = maximize expected reward
- Exploration = reduce uncertainty in the model
- Exploration is hard, but necessary
- Exploration is expensive (most explored items do not work)
- In practice: trade off exploration and exploitation (e.g. random)

Explore-exploit in top-N recommender systems via Gaussian processes, H. Vanchinathan et al, RecSys 2014

The Long Tail

- Long tail on users (most users rate very few items)
- Long tail on products (most items are rarely seen)
- Pay attention to data distribution!

The long tail of recommender systems and how to leverage it, YJ Park, A. Thuzhilin, RecSys 2008

Reco as an ML problem

- Early recommenders used heuristics (best-ofs)
- New recommenders use machine learning and optimize a function

Lessons from the Nextflix Prize Challenge

- Launched in 2006 by Netflix
- Open competition to beat the baseline (CineMatch)
- Training set: 100,480,507 ratings (480,189 users, 17,770 movies)
- Qualifying set: 2,817,131 ratings (50% test set, 50% quiz set)
- Improve by 10% to win $1,000,000

Lessons from the Netflix Prize Challenge, R. Bell and Y. Koren, SigKDD 2007

Lessons from the Nextflix Prize Challenge

- A trivial solution is almost as good as the baseline (average all ratings for a movie)
- The best method could not make it to production (ensemble)
- Optimizing RMSE $\neq$ optimizing user experience
- The contest shaked the industrial/academic community and paved the way for new recommendation algorithms

Lessons from the Netflix Prize Challenge, R. Bell and Y. Koren, SigKDD 2007

Evaluation of recommender systems

- Offline: evaluate new model on offline metrics
- Online: test new model with real users

Offline evaluation

- Typical metrics: RMSE, LLH, ranking metrics (NDCG)

- Convenient
- Hard to align with online behavior
- Useful to fine-tune a model or gather intuitions about the model behavior

Offline evaluation

RMSE for implicit feedback is wrong: people who viewed an item only few times may dislike it more than items they have never seen

Offline AB testing

- Leverage counterfactual or off-policy estimators
- Propose two variants of counterfactual estimates with different modelling of the bias
- Benchmark the estimators against uplift in revenue

- Requires to log a lot of additional data (candidates that were not displayed)
- Variance is a major challenge (sweet spot between minor and major changes)

Offline A/B testing for recommender system, A. Gilotte et al, WSDM 2018

Online evaluation

- Test your model on real traffic and measure the difference with the baseline

- Expensive (most ideas do not work)
- Bias removal is a big challenge
- You need to define a clear set of metrics for roll out decisions
- Online AB testing is a science

Offline evaluation is hard

- REVEAL 2018 Workshop
- Held with RecSys 2018 (Vancouver, CA)
- Submission deadline: July 28, 2018
- https://sites.google.com/view/reveal2018