“In a simple Netflix-style item recommender, we would simply apply some form of matrix factorization (i.e. NMF)”
From the Netflix Prize to today

2006

2013
Everything is Personalized
Everything is personalized

Over 75% of what people watch comes from a recommendation
Top 10

Diversity

Personalization awareness
But...

Netflix launches user profiles for individual recommendations

By Julianne Pepitone @juliannepe @CNNMoney August 1, 2013, 10:54 AM ET

Timmy’s love of watching cartoons on Netflix won’t ruin his parents’ Netflix recommendations anymore.

NEW YORK (CNNMoney)

Netflix is launching a tool on Thursday that will ensure your spouse’s ‘Kardashians’ binge will no longer wreck your list of recommendations for cerebral French films with a strong female lead.
Support for Recommendations

Social Support
EVERYTHING is a Recommendation
Consumer (Data) Science
Consumer (Data) Science

1. Start with a hypothesis:
   - Algorithm/feature/design X will increase member engagement with our service, and ultimately member retention

2. Design a test
   - Develop a solution or prototype
   - Think about dependent & independent variables, control, significance…

3. Execute the test

4. Let data speak for itself
Offline/Online testing process

Offline Experimentation

Initial Hypothesis

- Reformulate Hypothesis
  - no
  - Try different Model?
    - yes
    - Test Offline
    - Hypothesis Validated?
      - yes
      - Deploy Feature
      - no
    - no
  - yes
  - no

Choose Model

Train Model

Online Experimentation

Design AB Test

- Choose Control
- Deploy Prototype
- Observe Behavior
- Analyze Results
- Significant Improvements?
  - yes
  - Deploy Feature
  - no
Executing A/B tests

Measure differences in metrics across statistically identical populations that each experience a different algorithm.

- Decisions on the product always data-driven
- Overall Evaluation Criteria (OEC) = member retention
  - Use long-term metrics whenever possible
  - Short-term metrics can be informative and allow faster decisions
    - But, not always aligned with OEC
- Significance and hypothesis testing (1000s of members and 2-20 cells)
- A/B Tests allow testing many (radical) ideas at the same time (typically 100s of customer A/B tests running)
Offline testing

- Measure model performance, using (IR) metrics
- Offline performance used as an indication to make informed decisions on follow-up A/B tests
- A critical (and mostly unsolved) issue is how offline metrics can correlate with A/B test results.
- Extremely important to define offline evaluation framework that maps to online OEC
  - e.g. How to create training/testing datasets may not be trivial
Data & Models
Big Data @ Netflix

- > 40M subscribers
- Ratings: ~5M/day
- Searches: >3M/day
- Plays: > 50M/day
- Streamed hours:
  - 5B hours in Q3 2013

Member Behavior

- Time
- Geo-information
- Impressions
- Device Info
- Metadata
- Social
- Ratings

Demographics
Smart Models

- Regression models (Logistic, Linear, Elastic nets)
- SVD & other MF models
- Factorization Machines
- Restricted Boltzmann Machines
- Markov Chains & other graph models
- Clustering (from k-means to HDP)
- Deep ANN
- LDA
- Association Rules
- GBDT/RF
- …
SVD for Rating Prediction

- User factor vectors $p_u \in \mathbb{R}^f$ and item-factors vectors $q_v \in \mathbb{R}^f$
- Baseline (bias) $b_{uv} = \mu + b_u + b_v$ (user & item deviation from average)
- Predict rating as $r_{uv}' = b_{uv} + p_u^T q_v$
- SVD++ (Koren et. Al) asymmetric variation w. implicit feedback

\[
    r_{uv}' = b_{uv} + q_v^T \left( |R(u)|^{\frac{1}{2}} \sum_{j \in R(u)} (r_{uj} - b_{uj}) x_j + |N(u)|^{\frac{1}{2}} \sum_{j \in N(u)} y_j \right)
\]

- Where
  - $q_v, x_v, y_v \in \mathbb{R}^f$ are three item factor vectors
  - Users are not parametrized, but rather represented by:
    - $R(u)$: items rated by user $u$ & $N(u)$: items for which the user has given implicit preference (e.g. rated/not rated)
Restricted Boltzmann Machines

- Restrict the connectivity in ANN to make learning easier.
  - Only one layer of hidden units.
    - Although multiple layers are possible
  - No connections between hidden units.
- Hidden units are independent given the visible states.
- RBMs can be stacked to form Deep Belief Networks (DBN) – 4th generation of ANNs

![Diagram of Restricted Boltzmann Machines](image)

Figure 1: A restricted Boltzmann machine with binary hidden units and softmax visible units. For each user, the RBM only includes softmax units for the movies that user has rated. In addition to the symmetric weights between each hidden unit and each of the $K = 5$ values of a softmax unit, there are 5 biases for each softmax unit and one for each hidden unit. When modeling user ratings with an RBM that has Gaussian hidden units, the top layer is composed of linear units with Gaussian noise.
Ranking

- Ranking = **Scoring + Sorting + Filtering**
  bags of movies for presentation to a user
- Key algorithm, sorts titles in most contexts
- **Goal**: Find the best possible ordering of a set of videos for a user within a specific context in real-time
- **Objective**: maximize consumption & “enjoyment”

**Factors**
- Accuracy
- Novelty
- Diversity
- Freshness
- Scalability
- …
Example: Two features, linear model

Linear Model:

$$f_{\text{rank}}(u,v) = w_1 p(v) + w_2 r(u,v) + b$$
Example: Two features, linear model
Ranking improvement over baseline

0%  50%  100%  150%  200%  250%  300%

- Popularity
- + Ratings
- + More Features & Optimized Models

- NETFLIX
Learning to Rank Approaches

- ML problem: construct ranking model from training data

1. **Pointwise** (Ordinal regression, Logistic regression, SVM, GBDT, …)
   - Loss function defined on individual relevance judgment
2. **Pairwise** (RankSVM, RankBoost, RankNet, FRank…)
   - Loss function defined on pair-wise preferences
   - Goal: minimize number of inversions in ranking
3. **Listwise**
   - **Indirect Loss Function** (RankCosine, ListNet…)
   - **Directly optimize IR measures** (NDCG, MRR, FCP…)
     - Genetic Programming or Simulated Annealing
     - Use boosting to optimize NDCG (Adarank)
     - Gradient descent on smoothed version (CLiMF, TFMAP, GAPfm @cikm13)
     - Iterative Coordinate Ascent (Direct Rank @kdd13)
Other research questions we are working on

- Row selection
- Diversity
- Similarity
- Context-aware recommendations
- Explore/exploit
- Presentation bias correction
- Mood and session intent inference
- Unavailable Title Search
- ...

NETFLIX
More data or better models?
More data or better models?

I teach a class on Data Mining at Stanford. Students in my class are expected to do a project that does some non-trivial data mining. Many students opted to try their hand at the Netflix Challenge: to design a movie recommendations algorithm that does better than the one developed by Netflix.

Here's how the competition works. Netflix has provided a large data set that tells you how nearly half a million people have rated about 18,000 movies. Based on these ratings, you are asked to predict the ratings of these users for movies in the set that they have not rated. The first team to beat the accuracy of Netflix's proprietary algorithm by a certain margin wins a prize of $1 million!

Different student teams in my class adopted different approaches to the problem, using both published algorithms and novel ideas. Of these, the results from two of

Anand Rajaraman: Former Stanford Prof. & Senior VP at Walmart
More data or better models?

Sometimes, it's not about more data.
Norvig: “Google does not have better Algorithms, only more Data”

The Unreasonable Effectiveness of Data

Many features/low-bias models

Figure 1. Learning Curves for Confusion Set Disambiguation

[Banko and Brill, 2001]
More data or better models?

Sometimes, it’s not about more data
More data or better models?

THE PETABYTE AGE:

"All models are wrong. Some are useful."

Sensors everywhere. In smartphones, processors. Our ability to sense and understand the world...
“Data without a sound approach = noise”
More data +
Smarter models +
More accurate metrics +
Better approaches

Lots of room for improvement!
Thanks!

NETFLIX

We are hiring!