Beyond Data

From User Information to Business Value

October, 2013

Xavier Amatriain
Director - Algorithms Engineering - Netflix

@xamat

“In a simple Netlfix-style item recommender, we would simply apply some form of matrix factorization (i.e NMF)”

Machine Learning

notes, thoughts, and practice of applied machine learning

Music Recommendations and the Logistic Metric Embedding

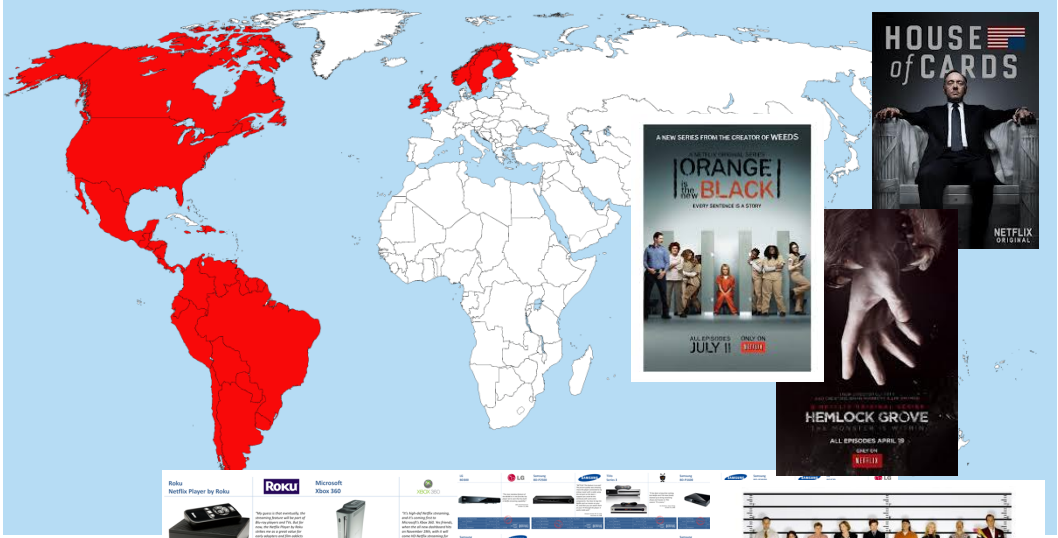
Posted on October 28, 2013

0

From the Netflix Prize to today

NETFLIX

2006



amazon web services™

2013

Everything is

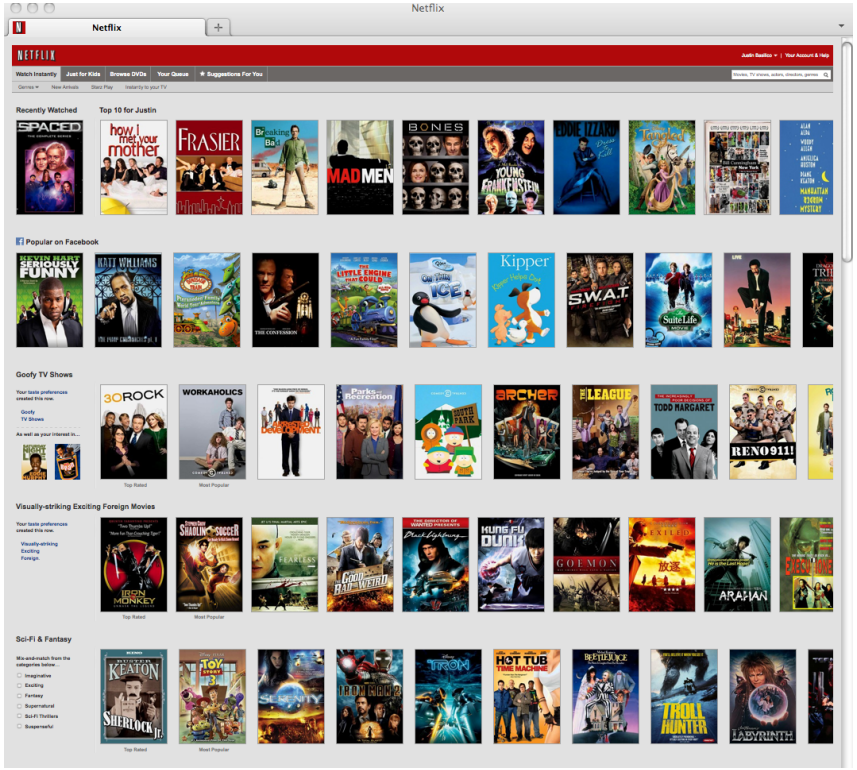
NETFLIX

Personalized

Everything is personalized

Ranking

Rows



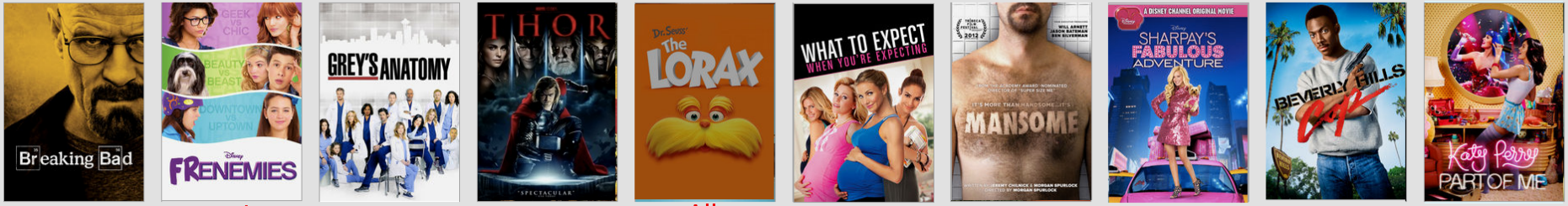
Over 75% of what people watch comes from a recommendation

NETFLIX

Top 10

Personalization awareness

Top 10 for Xavier



Dad | Daughter | Mom | Son | All | Mom | Dad | Daughter | Son | Daughter

Diversity

NETFLIX

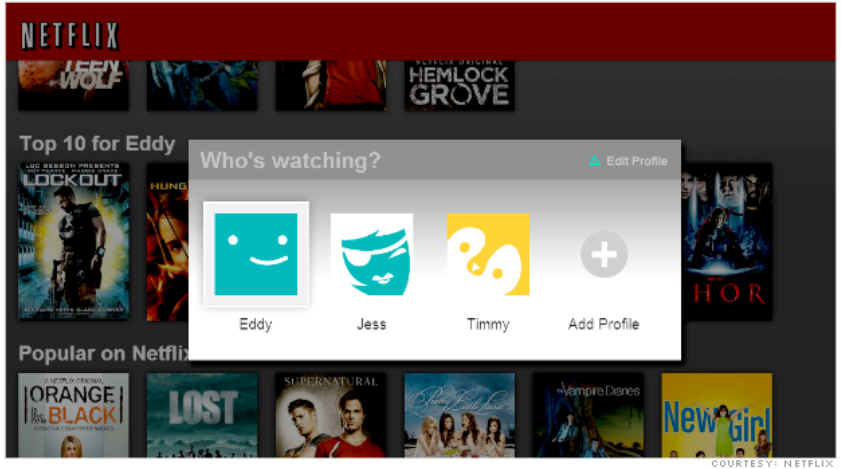
But…

Netflix launches user profiles for individual recommendations

CNNMoney

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

Recommend 1.2k | Tweet 67 | Share | +1 44 | Email | Print



COURTESY: NETFLIX

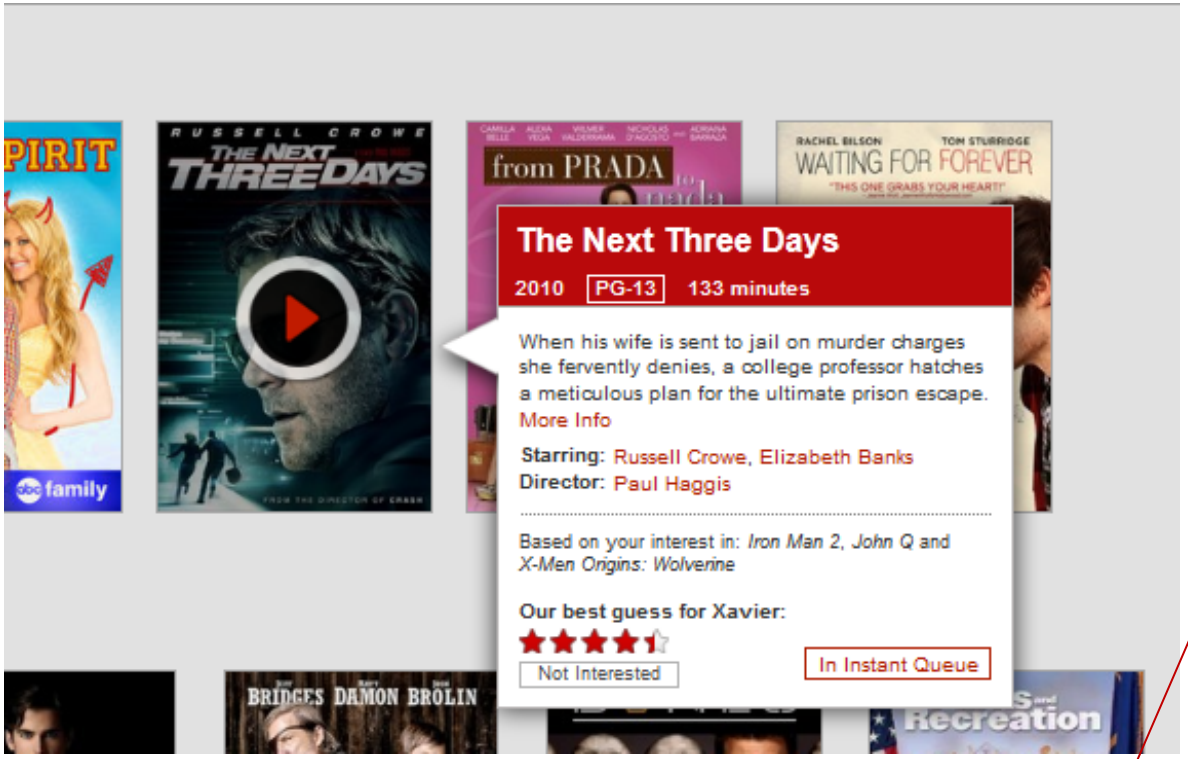
Timmy's love of watching cartoons on Netflic won't ruin his parents' Netflix recommendations anymore.

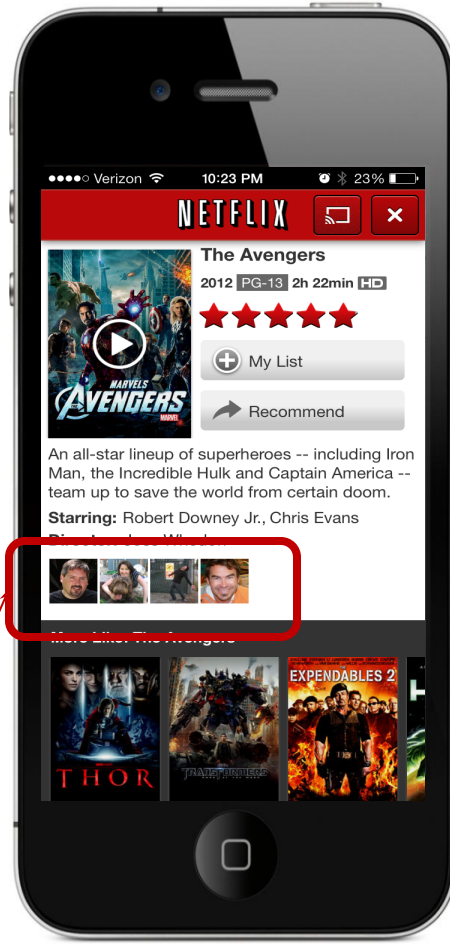
NEW YORK (CNNMoney)

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.

NETFLIX

Support for Recommendations





Social Support

NETFLIX

Genre Rows

NETFLIX

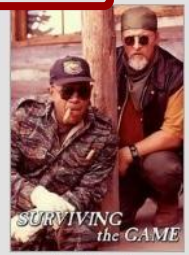
Watch Instantly | Just for Kids | Browse DVDs | Your Queue | ★ Suggestions For You

Genres ▾ | New Arrivals | Starz Play | Instantly to your TV

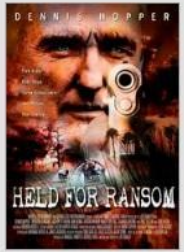
Suspenseful Wilderness-survival Action & Adventure

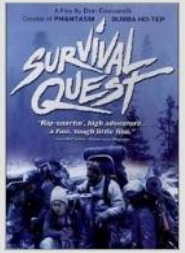
Based on your interest in...

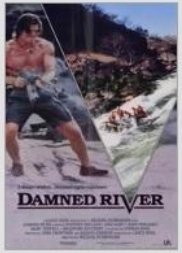
Top Rated



Most Popular





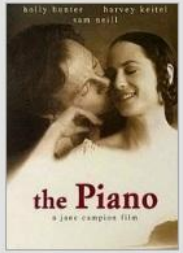


Independent Dramas Featuring a Strong Female Lead

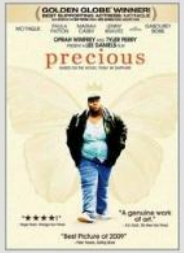
Your taste preferences created this row.

Independent

As well as your interest in...



Top Rated

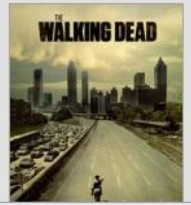


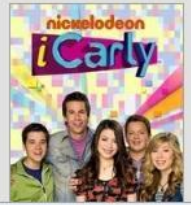
Most Popular

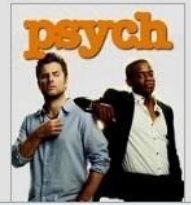
TV Shows

Mix-and-match from the categories below...

- ☐ Family-friendly
- ☐ TV Comedies
- ☐ Cartoons
- ☐ Kids' TV Shows
- ☐ TV Docs

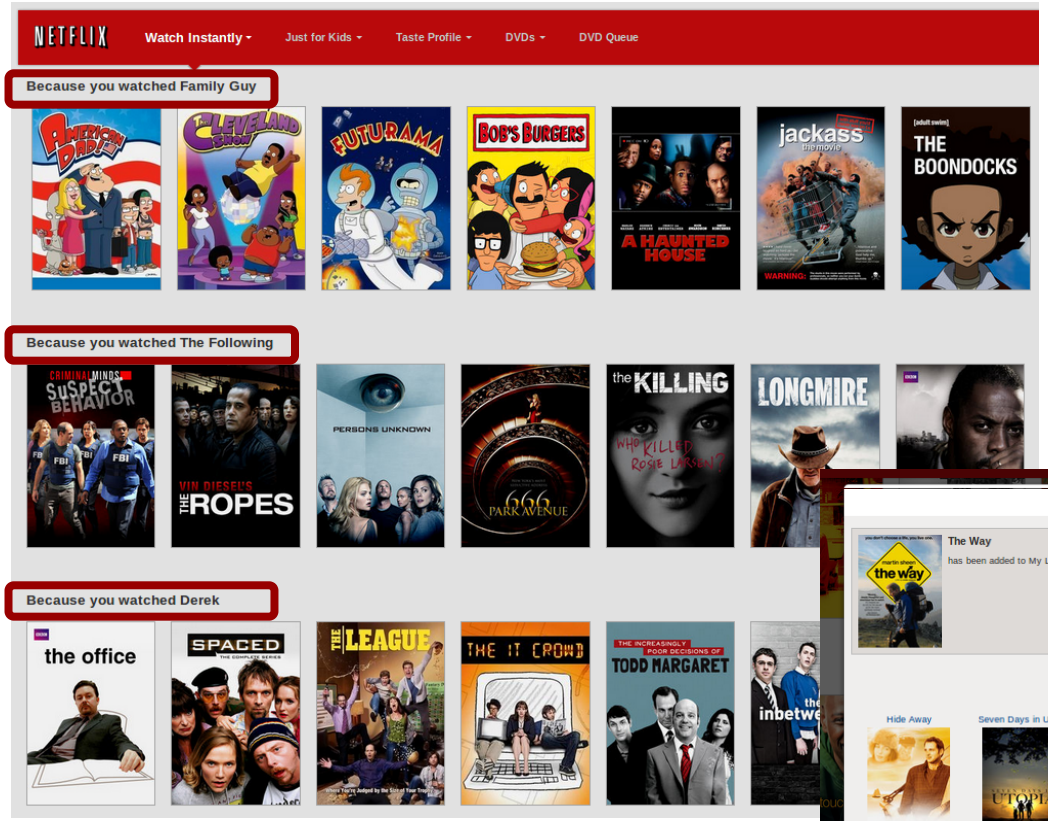


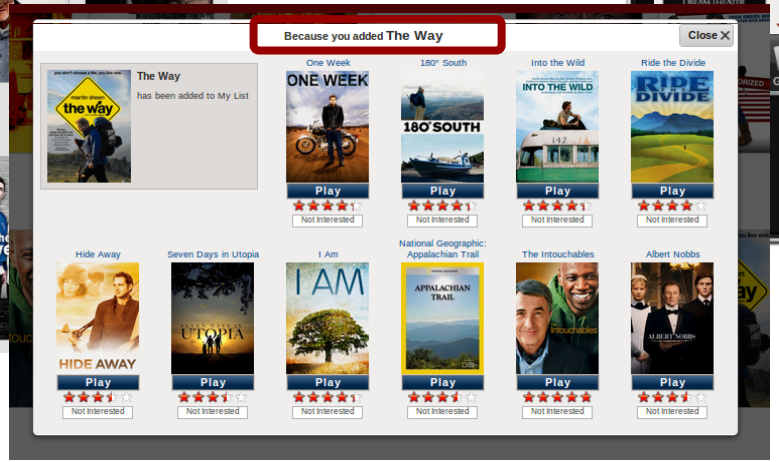


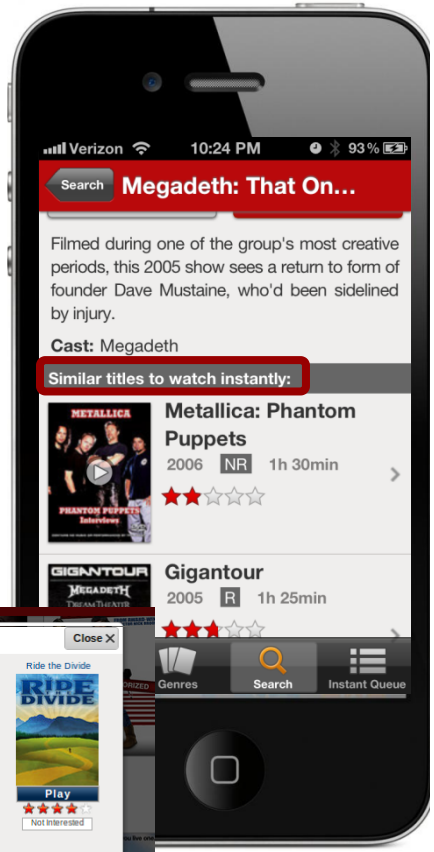


NETFLIX

Similars


NETFLIX
Watch Instantly
Just for Kids
Taste Profile
DVDs
DVD Queue
Because you watched Family Guy
Because you watched The Following
Because you watched Derek


Because you added The Way
Close
The Way
has been added to My List
One Week
180° South
Into the Wild
Ride the Divide
Hide Away
Seven Days in Utopia
I Am
National Geographic: Appalachian Trail
The Intouchables
Albert Nobbs
Play
Not Interested


Verizon
10:24 PM
93 %
Search
Megadeth: That On...
Filmed during one of the group's most creative periods, this 2005 show sees a return to form of founder Dave Mustaine, who'd been sidelined by injury.
Cast: Megadeth
Similar titles to watch instantly:
Metallica: Phantom Puppets
2006 NR 1h 30min
Gigantour
2005 R 1h 25min
Genres
Search
Instant Queue

EVERYTHING is a Recommendation

NETFLIX Watch Instantly Just for Kids Taste Profile DVDs Movies, TV shows, actors, directors, genres Michael

Recently Watched

My List See All

Top 10 for Michael

Popular on Netflix

Alphas
2011-2012 TV-14 2 Seasons
A team of individuals gifted with extraordinary neurological abilities is tasked with solving a series of high-profile crimes. Among them are an autistic man whose mind works as fast as a computer and an FBI agent with super strength. More Info
Starring: David Strathairn, Ryan Cartwright
Creators: Zak Penn, Michael Karnow
+ My List

FRONT PAGE BUSINESS SMALL BUSINESS MEDIA SCIENCE GREEN COMEDY ARTS NE

Tech TEDWeekends • CES 2013 • Social Media • Women In Tech • Tech Videos • Influencers And Innovation

Could Iron Man's Lab Soon Be A Reality?

Facebook To Introduce New Photo Feature

Netflix's New 'My List' Feature Knows You Better Than You Know Yourself (Because Algorithms)

The Huffington Post | By Dino Grandoni
Posted: 08/21/2013 1:44 pm EDT | Updated: 08/22/2013 8:31 am EDT

Like 55 people like this. Be the first of your friends.

30 12 2 7 107

GET TECHNOLOGY NEWSLETTERS:
Enter email SUBSCRIBE

NETFLIX

Consumer (Data) Science

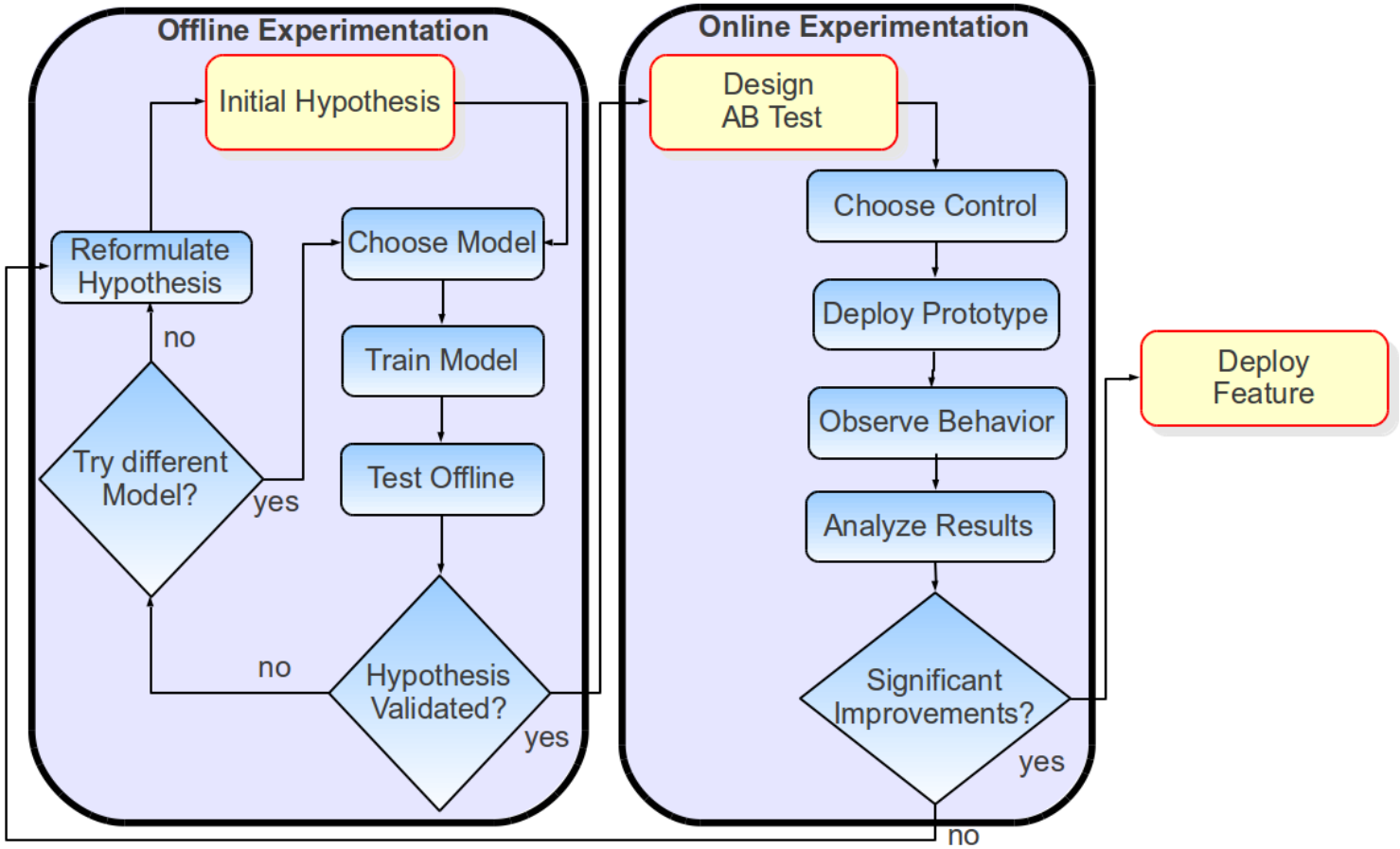
NETFLIX

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

NETFLIX™

Offline/Online testing process



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)

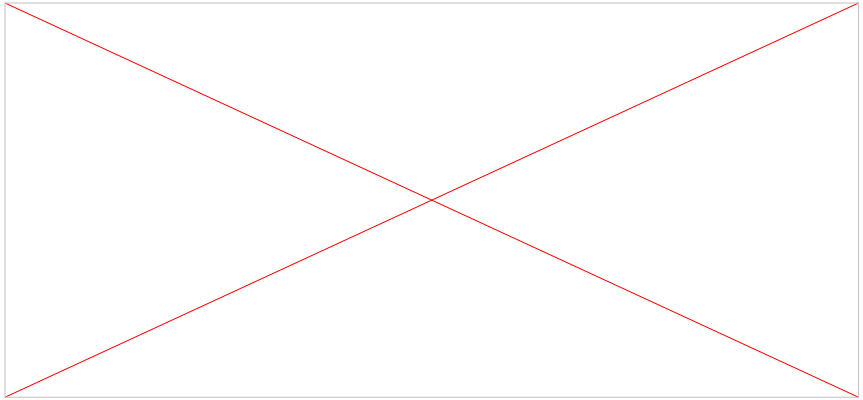
NETFLIX

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

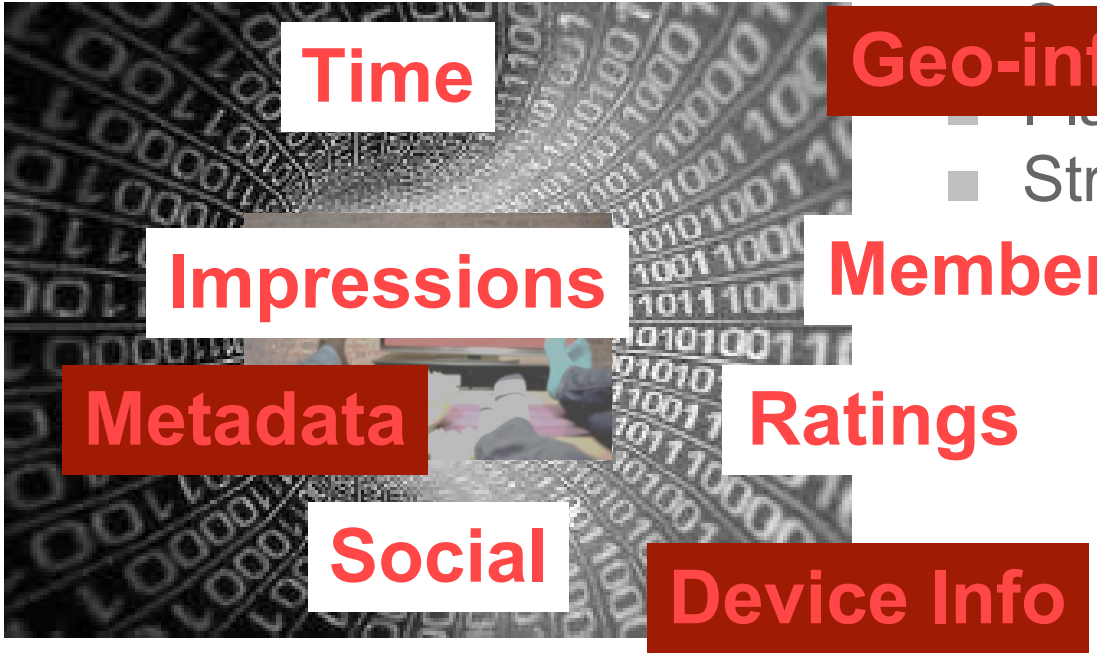
NETFLIX

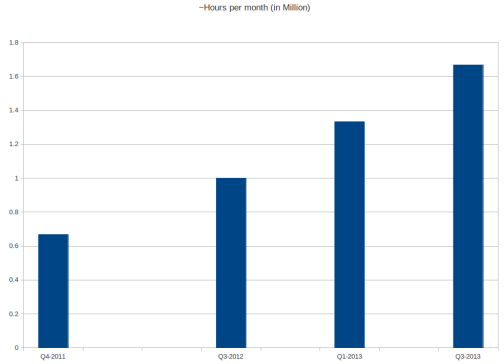
Data & Models



Big Data @Netflix

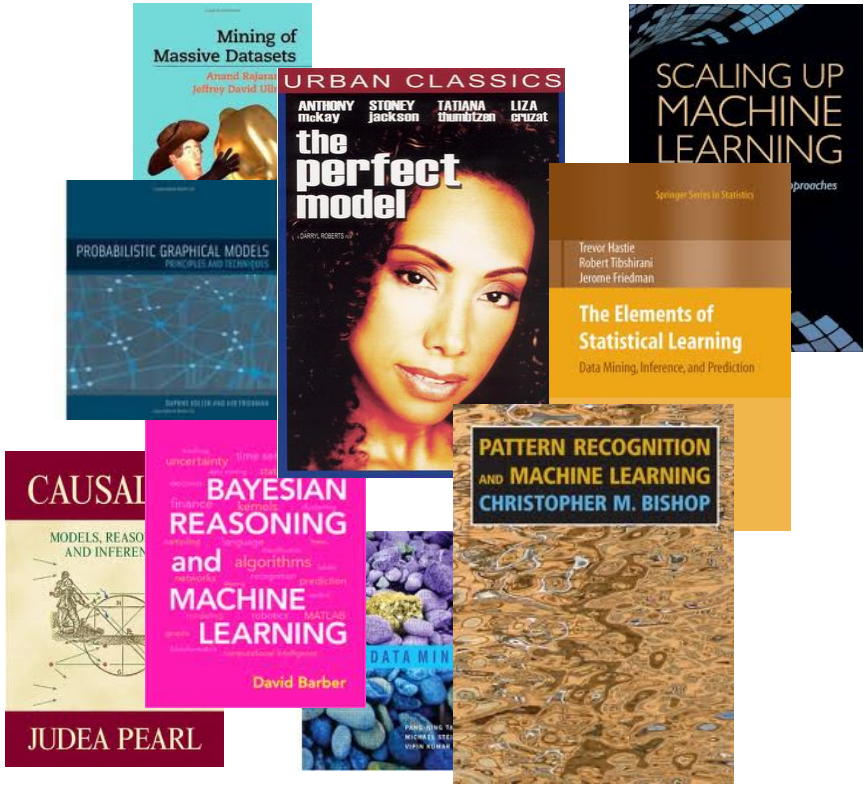
- > 40M subscribers
- Ratings: ~5M/day
- [illegible]/day
- Plays: [illegible]M/day
- Streamed hours: Q3 2013





NETFLIX™

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
- …

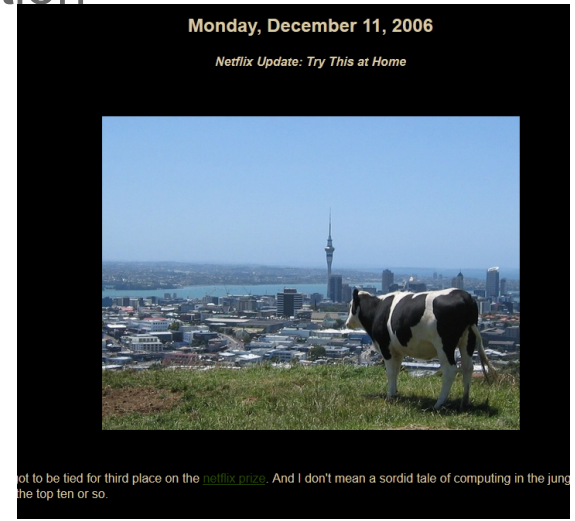
NETFLIX

SVD for Rating Prediction

- User factor vectors $p_u \in \Re^f$ and item-factors vectors $q_v \in \Re^f$
- 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
  - $q_v, x_v, y_v \in \Re^f$ 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

Restricted Boltzmann Machines for Collaborative Filtering

Ruslan Salakhutdinov RSALAKHU@CS.TORONTO.EDU
Andriy Mnih AMNIH@CS.TORONTO.EDU
Geoffrey Hinton HINTON@CS.TORONTO.EDU
University of Toronto, 6 King's College Rd., Toronto, Ontario M5S 3G4, Canada

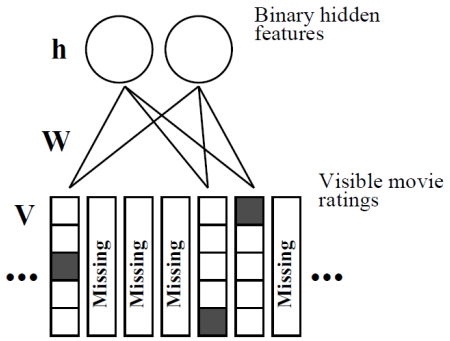
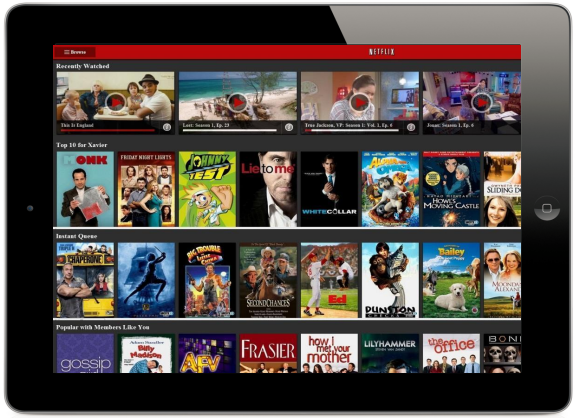


Figure 1. A restricted Boltzmann machine with binary hidden units and softmax visible units. For each user, the RBM only includes softmax units for the movies that user has rated. In addition to the symmetric weights between each hidden unit and each of the $K = 5$ values of a softmax unit, there are 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

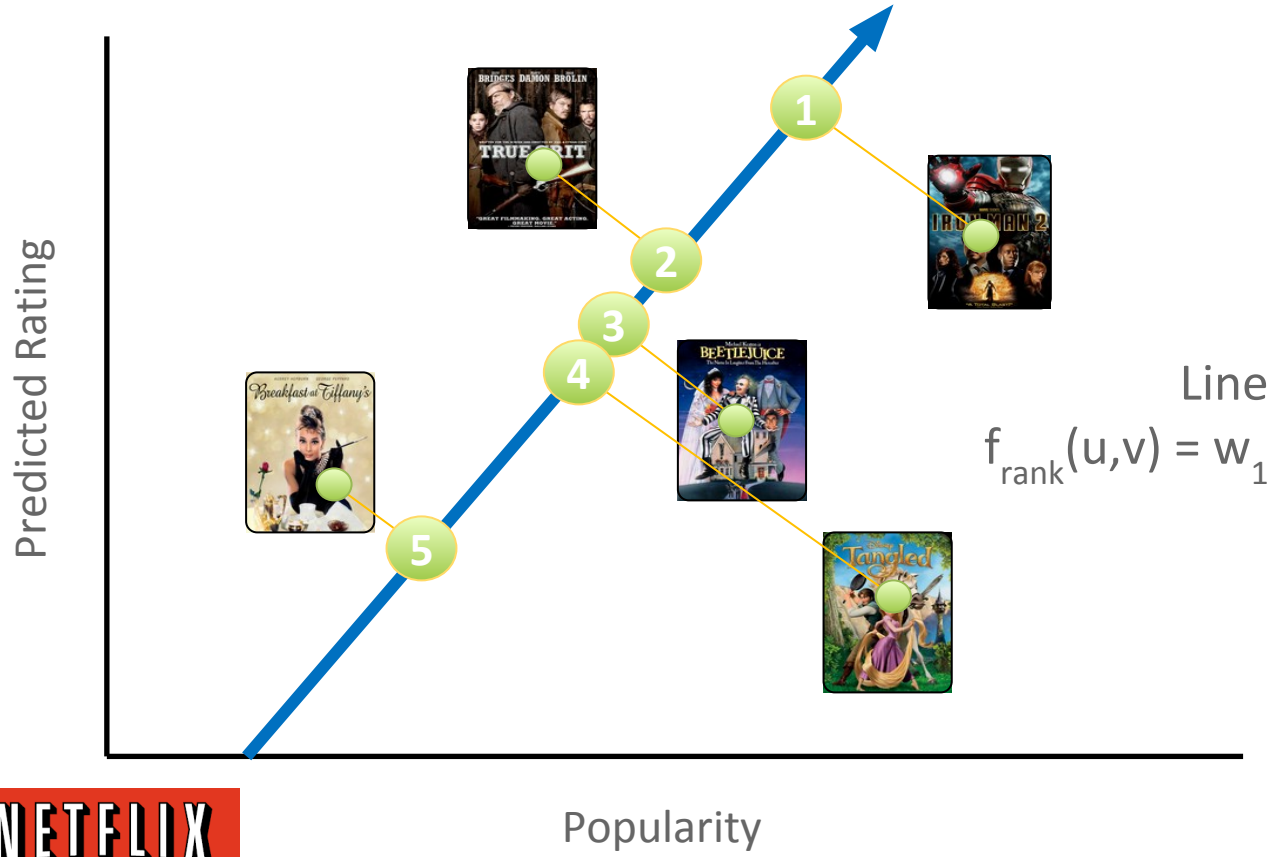
- Ranking = **Scoring** + **Sorting** + **Filtering** bags of movies for presentation to a user
- Key algorithm, sorts titles in most contexts
- Goal: Find the best possible ordering of a set of *videos* for a *user* within a specific *context* in real-time
- Objective: maximize consumption & "enjoyment"

- Factors
  - Accuracy
  - Novelty
  - Diversity
  - Freshness
  - Scalability
  - …



NETFLIX

Example: Two features, linear model



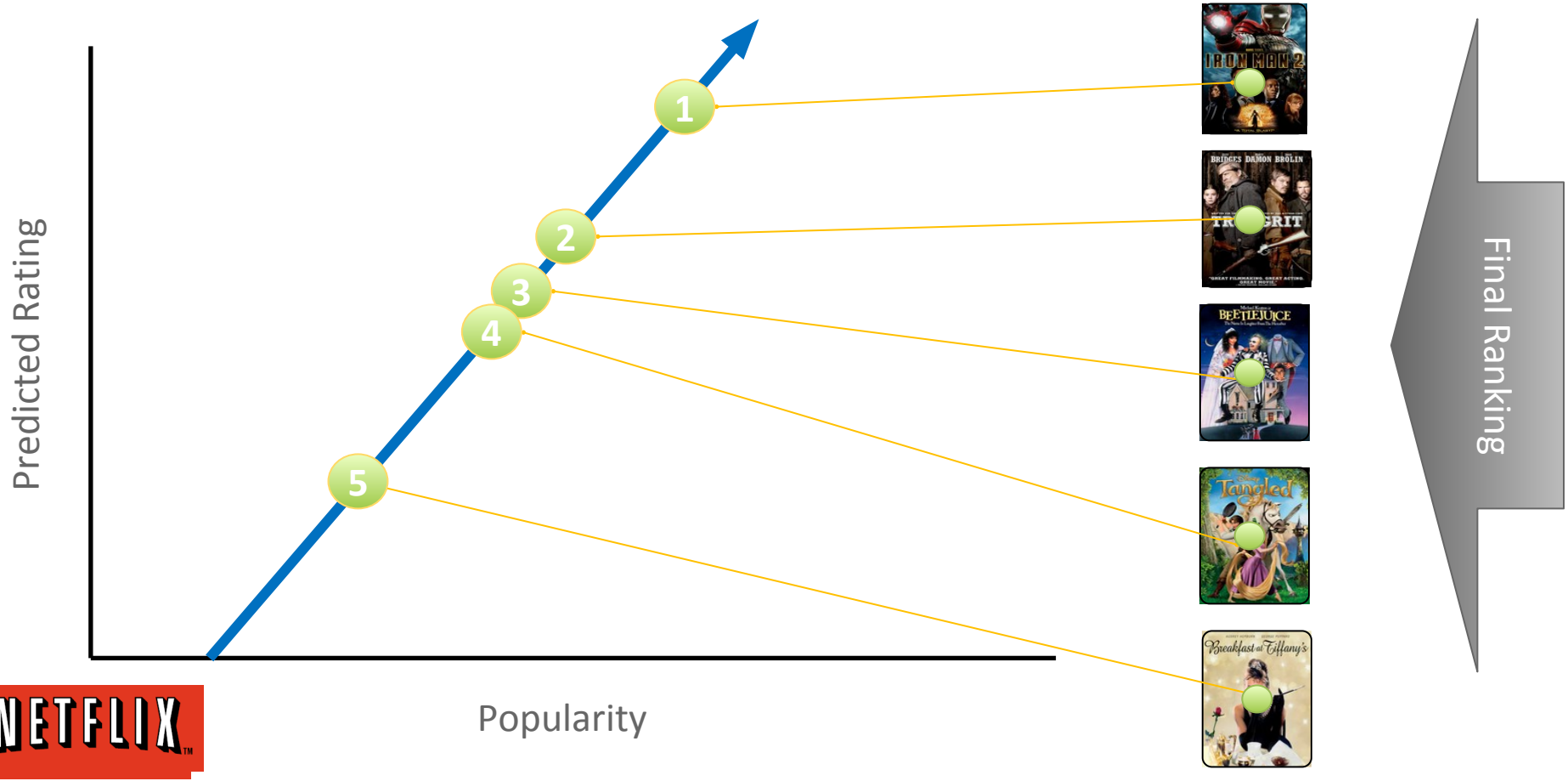
Linear Model:

$$f_{rank}(u,v) = w_1\, p(v) + w_2\, r(u,v) + \ldots$$

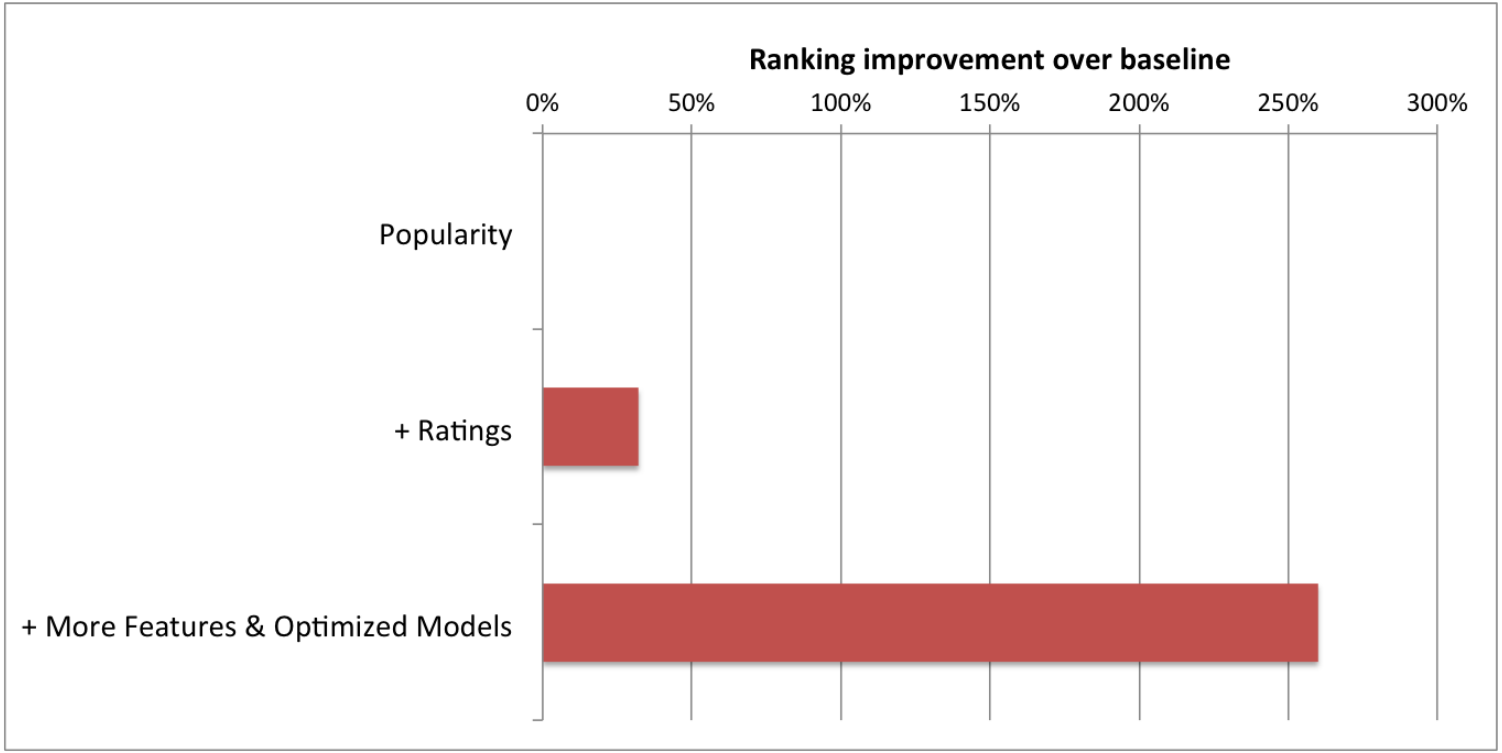
Final Ranking

NETFLIX

Example: Two features, linear model



Ranking


Ranking improvement over baseline
0%
50%
100%
150%
200%
250%
300%
Popularity
+ Ratings
+ More Features & Optimized Models

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)

NETFLIX

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
- ...

More data or better models?

NETFLIX

More data or better models?

Datawocky

On Teasing Patterns from Data, with Applications to Search, Social Media, and Advertising

« Enumerating User Data Collection Points | Main | Traveling: In India this week »

More data usually beats better algorithms

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

ABOUT

Anand Rajaraman

Datawocky

RECENT POS

Goodbye, Kosmix. H @WalmartLabs

Retail + Social + Mob @WalmartLabs

Creating a Culture of Innovation: Why 20% not Enough

Reboot: How to Rein Technology Startup

Really?

Anand Rajaraman: Former Stanford Prof. & Senior VP at Walmart

NETFLIX

More data or better models?

Sometimes, it's not about more data

Recommending New Movies: Even a Few Ratings Are More Valuable Than Metadata

István Pilászy *
Dept. of Measurement and Information Systems
Budapest University of Technology and Economics
Magyar Tudósok krt. 2.
Budapest, Hungary
pila@mit.bme.hu

Domonkos Tikk *,†
Dept. of Telecom. and Media Informatics
Budapest University of Technology and Economics
Magyar Tudósok krt. 2.
Budapest, Hungary
tikk@tmit.bme.hu

ABSTRACT

The Netflix Prize (NP) competition gave much attention to collaborative filtering (CF) approaches. Matrix factorization (MF) based CF approaches assign low dimensional feature vectors to users and items. We link CF and content-based filtering (CBF) by finding a linear transformation that transforms user or item descriptions so that they are as close as possible to the feature vectors generated by MF for CF.

We propose methods for explicit feedback that are able to handle 140 000 features when feature vectors are very sparse. With movie metadata collected for the NP movies we show that the prediction performance of the methods is comparable to that of CF, and can be used to predict user preferences on new movies.

We also investigate the value of movie metadata compared to movie ratings in regards of predictive power. We compare

1. INTRODUCTION

The goal of recommender systems is to give personalized recommendation on items to users. Typically the recommendation is based on the former and current activity of the users, and metadata about users and items, if available.

There are two basic strategies that can be applied when generating recommendations. Collaborative filtering (CF) methods are based only on the activity of users, while content-based filtering (CBF) methods use only metadata. In this paper we propose hybrid methods, which try to benefit from both information sources.

The two most important families of CF methods are matrix factorization (MF) and neighbor-based approaches. Usually, the goal of MF is to find a low dimensional representation for both users and movies, i.e. each user and movie is associated with a feature vector. Movie metadata (which

More data or better models?

Norvig: “Google does not have better Algorithms, only more Data”

The Unreasonable Effectiveness of Data

Alon Halevy, Peter Norvig, and Fernando Pereira, *Google*

Many features/ low-bias models

[Banko and Brill, 2001]

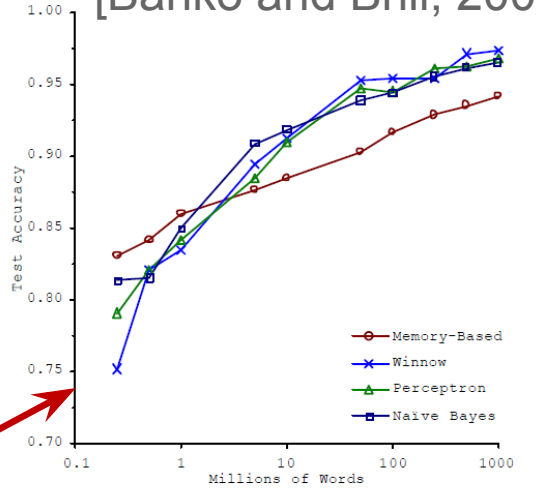
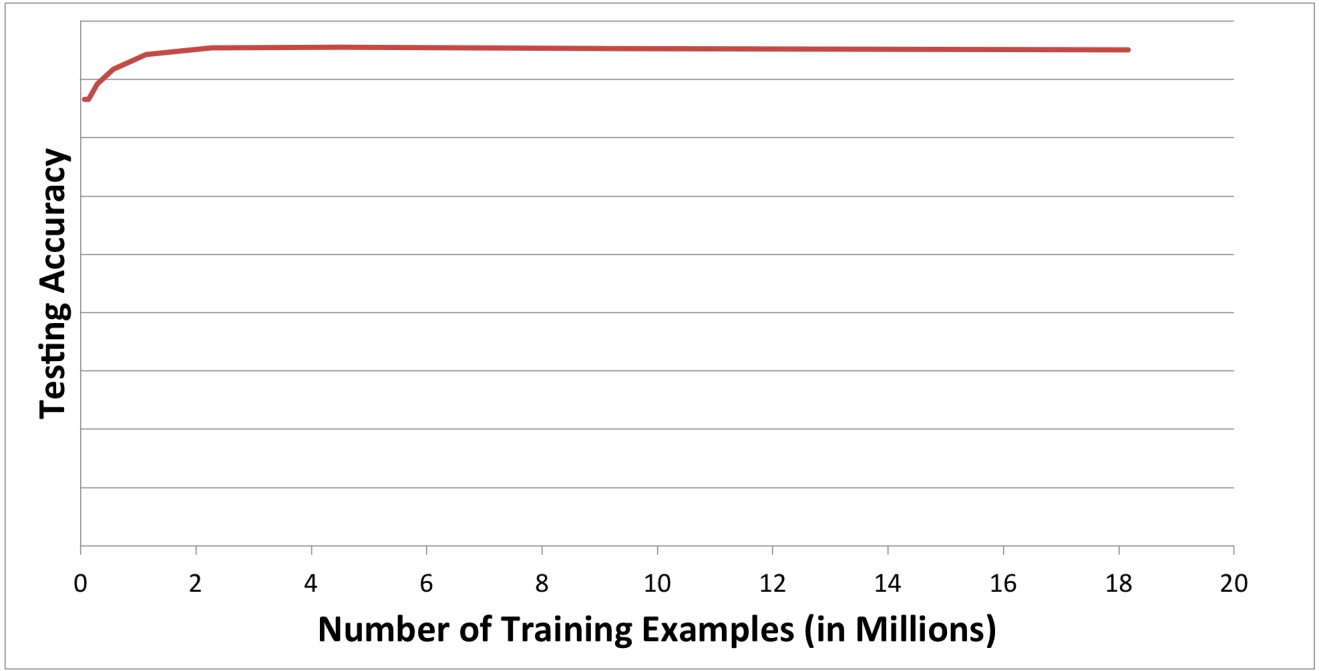


Figure 1. Learning Curves for Confusion Set Disambiguation

NETFLIX

More data or better models?



Sometimes, it's not about more data

NETFLIX

More data or better models?

WIRED MAGAZINE: 16.07

SCIENCE : DISCOVE

The End o The Data De es the Scientific M olete

By Chris Anderson 06.23.

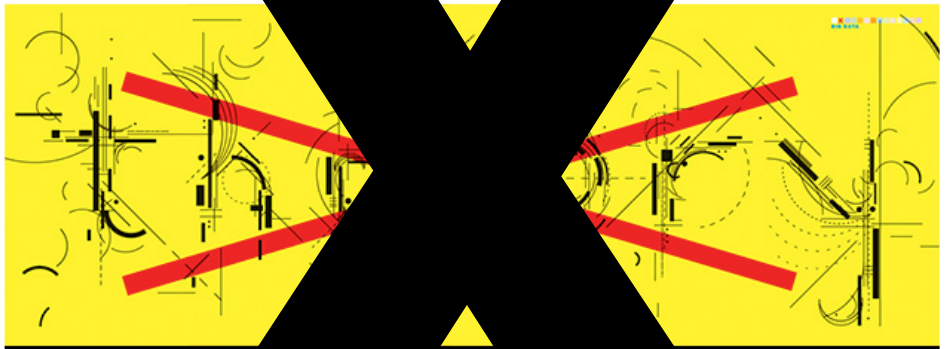


Illustration: Marian Bantjes

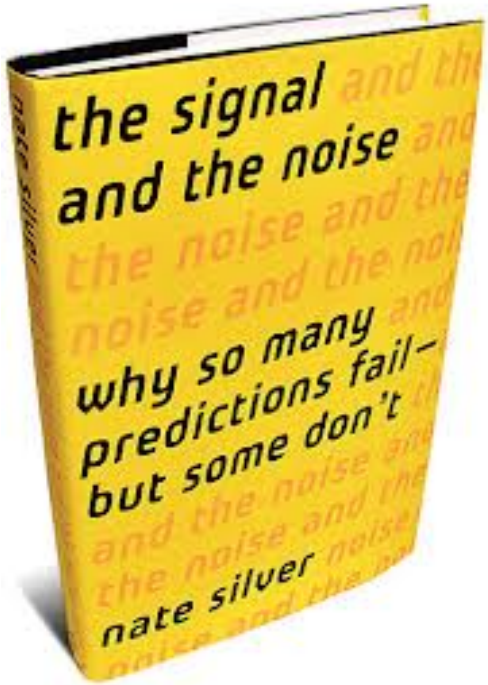
THE PETABYTE AGE:
Sensors everywhere. Inf
processors. Our abilit

"All models a e are useful."

So proclaimed stati o years ago, and

NETFLIX

“Data without a sound approach = noise”



NETFLIX

More data +

Smarter models +

More accurate metrics +

Better approaches

Lots of room for improvement!

NETFLIX

Xavier Amatriain (@xamat)
xavier@netflix.com

Thanks!

NETFLIX

We are hiring!