X-Informatics Case Study: e-Commerce and Life Style Informatics: Recommender Systems I

February 4 2013

Geoffrey Fox
gcf@indiana.edu

http://www.infomall.org/X-InformaticsSpring2013/index.html

Associate Dean for Research and Graduate Studies, School of Informatics and Computing
Indiana University Bloomington

2013

Big Data Ecosystem in One Sentence

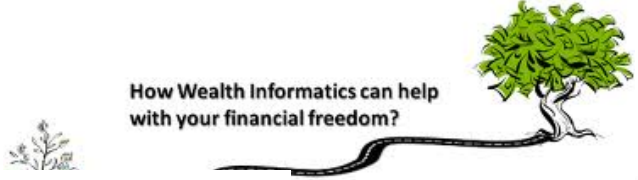
Use Clouds running Data Analytics Collaboratively processing Big Data to solve problems in X-Informatics (or e-X)

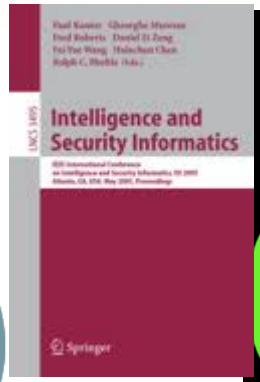
X = Astronomy, Biology, Biomedicine, Business, Chemistry, Climate, Crisis, Earth Science, Energy, Environment, Finance, Health, Intelligence, Lifestyle, Marketing, Medicine, Pathology, Policy, Radar, Security, Sensor, Social, Sustainability, Wealth and Wellness with more fields (physics) defined implicitly

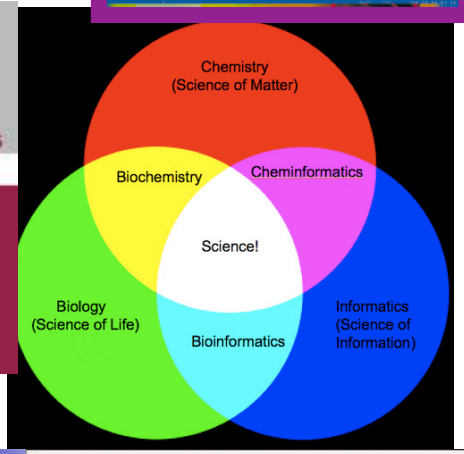
Spans Industry and Science (research)

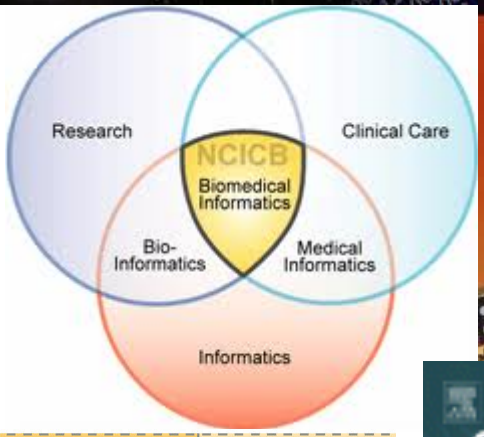
Education: Data Science see recent New York Times articles

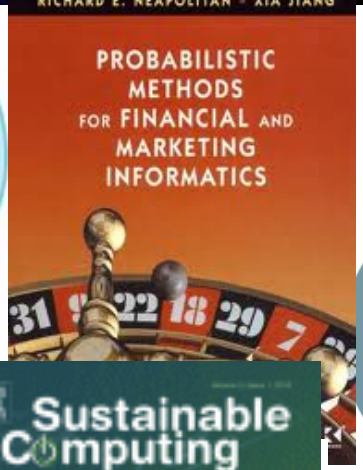
http://datascience101.wordpress.com/2013/04/13/new-york-times-data-science-articles/

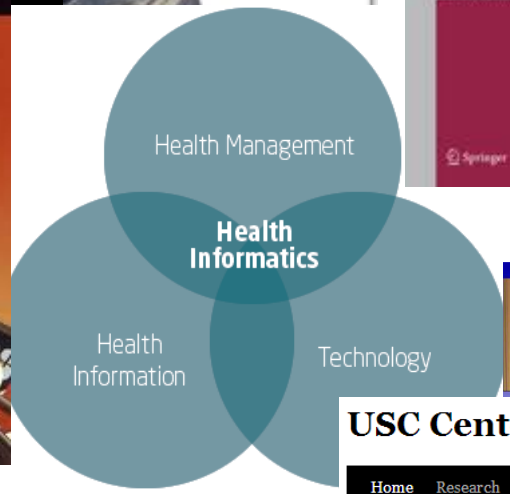


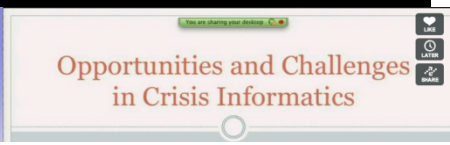


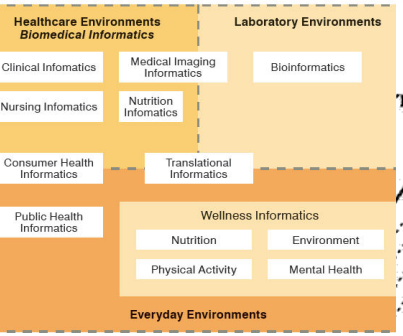


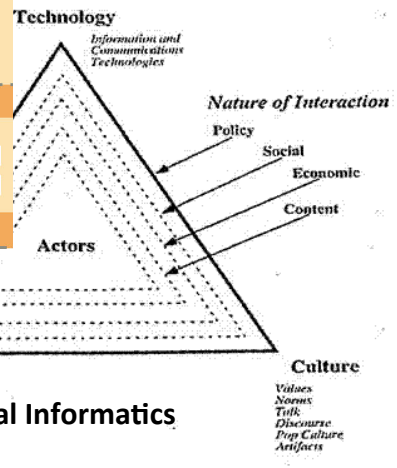








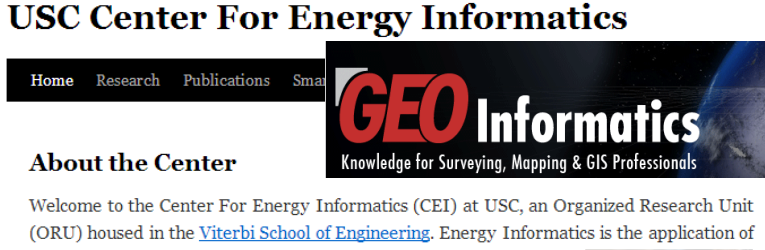


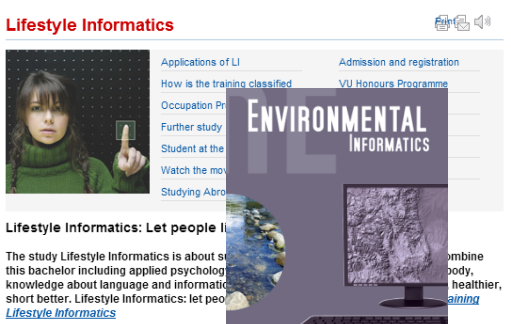


Noelia Penelope Greer (Ed.)

Business Informatics

Information technology, Management,





Recommender Systems as an Optimization Problem

Overview of Problems

- Basic problem is personalized matching of items to people or perhaps collections of items to collections of people
- **People** to **products**: Online and Offline Commerce
- **People** to **People**: Social Networking
- **People** to **Jobs** or **Employers**: Job Sites
- **People+Queries** to the **Web**: Information Retrieval (search as in Bing/Google)

Recommender Systems in more detail

- A large number of online and offline commerce activities plus basic Internet site personalization relies on "**recommender systems**"
- Given real-time action by user, immediately suggest new actions (as in Amazon **buy recommendations** on web)
- Based on past actions of users (and others) suggest **movies** to **look** at, **restaurants** to eat at, **events** to go to, books and music to **buy**
- Based on mix of explicit user choice and grouping of internet sites, present customized **Google News** page
- Given sales statistics, decide on **discounts** at "real" **supermarkets** and **placement** of related (by analysis of buying habits) products
- Identify possible colleagues at Social Networking sites like **LinkedIn**
- Identify matches between employers and employees at sites like **CareerBuilder** and **Monster**

Everything is an Optimization Problem?

- **Fit** Model to Data
  - Higgs + Background
- **Match** User to Jobs or Books or Other Users?
- **Classification** is optimizing assignment of members of an ontology (list of categories) to data
- **Keeping alive in a jungle** involves optimizing one distance from hungry lions
  - One's neural net optimizes interpretation of pixels detected by your eye to identify the lions you are trying to avoid
- **Typically minimize some function (or maximize negative of function)**
  - Interesting feature of these problems is ingenious choice of function
  - Note Physics minimizes (free) energy
  - This usually involves thinking of people and/or items as points in a space (not always a traditional vector space)

Next Part of Course

- Comments from online site http://recommenderbook.net/teaching-material/slides on goals of recommender system and overall approaches
- Introduction to Kaggle site
- Comments from Netflix http://www.slideshare.net/xamat/building-largescale-realworld-recommender-systems-recsys2012-tutorial on their approach

Recommender Systems Introduction

Problem domain

- **Recommendation systems (RS) help to match users with items**
  - Ease information overload
  - Sales assistance (guidance, advisory, persuasion,…)

RS are software agents that elicit the interests and preferences of individual consumers […] and make recommendations accordingly.

They have the potential to support and improve the quality of the decisions consumers make while searching for and selecting products online.

» (Xiao & Benbasat 2007[1])

- **Different system designs / paradigms**
  - Based on availability of exploitable data
  - Implicit and explicit user feedback See Netflix example later
  - Domain characteristics

(1) Xiao and Benbasat, *E-commerce product recommendation agents: Use, characteristics, and impact*, MIS Quarterly **31** (2007), no. 1, 137–209

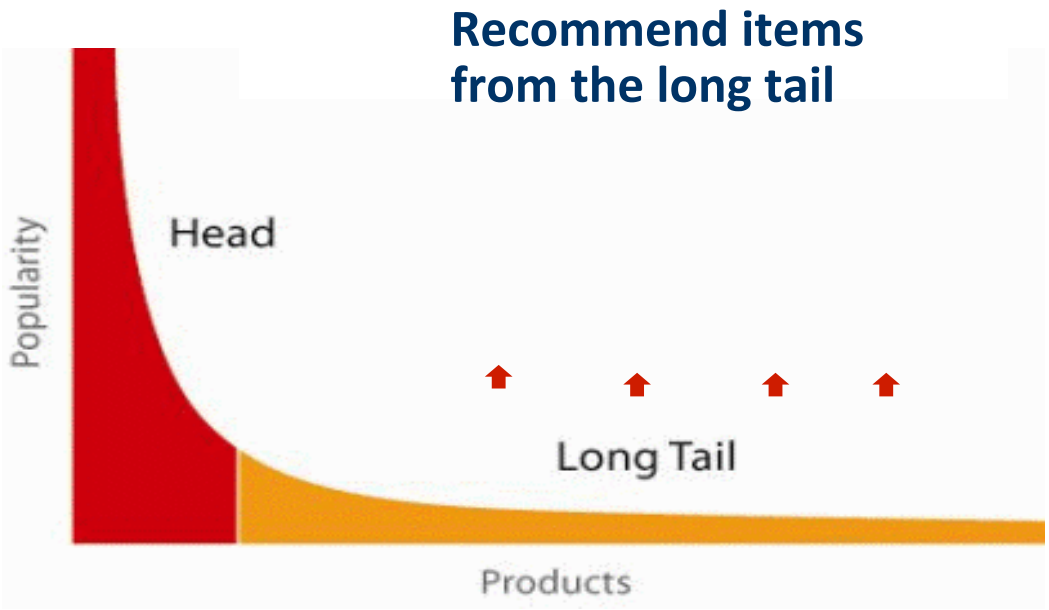
Purpose and success criteria (1)

- **Different perspectives/aspects**
  - Depends on domain and purpose
  - No holistic evaluation scenario exists
- **Retrieval perspective**
  - Reduce search costs
  - Provide "correct" proposals
  - Users know in advance what they want
- **Recommendation perspective**
  - Serendipity (unexpected success) – identify items from the Long Tail
  - Users did not know about existence

- Money is often essentially criterion
- This is goal of site owner but in long term this requires that
- User maximizes happiness or knowledge

When does a RS do its job well?

90-10 or 80-20 "rule" broadly applicable



- "Recommend widely unknown items that users might actually like!"
- 20% of items accumulate 74% of all positive ratings
- Positive means items rated > 3 in MovieLens 100K dataset (ratings 0 to 5)

Purpose and success criteria (2)

- **Prediction perspective**
  - Predict to what degree users like an item
  - Most popular evaluation scenario in research
- **Interaction perspective**
  - Give users a "good feeling"
  - Educate users about the product domain
  - Convince/persuade users - explain
- **Finally, conversion perspective**
  - Commercial situations
  - Increase "hit", "clickthrough", "lookers to bookers" rates
  - Optimize sales margins and profit

Money!

Recommender systems

- **RS seen as a function**
- **Given:**
  - User model (e.g. ratings, preferences, demographics, situational context)
  - Items (with or without description of item characteristics)
- **Find:**
  - Relevance score. Used for ranking.

- **Relation to Information Retrieval:**
  - IR is finding material [..] of an unstructured nature [..] that satisfies an information need from within large collections [..].
    - » (Manning et al. 2008[1])

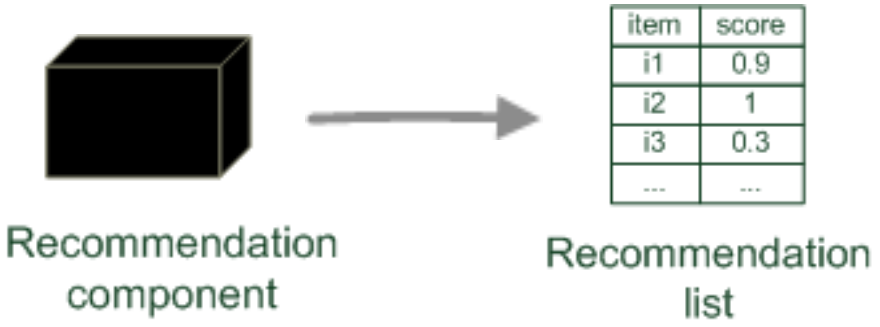
Advanced IR does use essentially recommender system to order results

Recommend links that satisfy search query

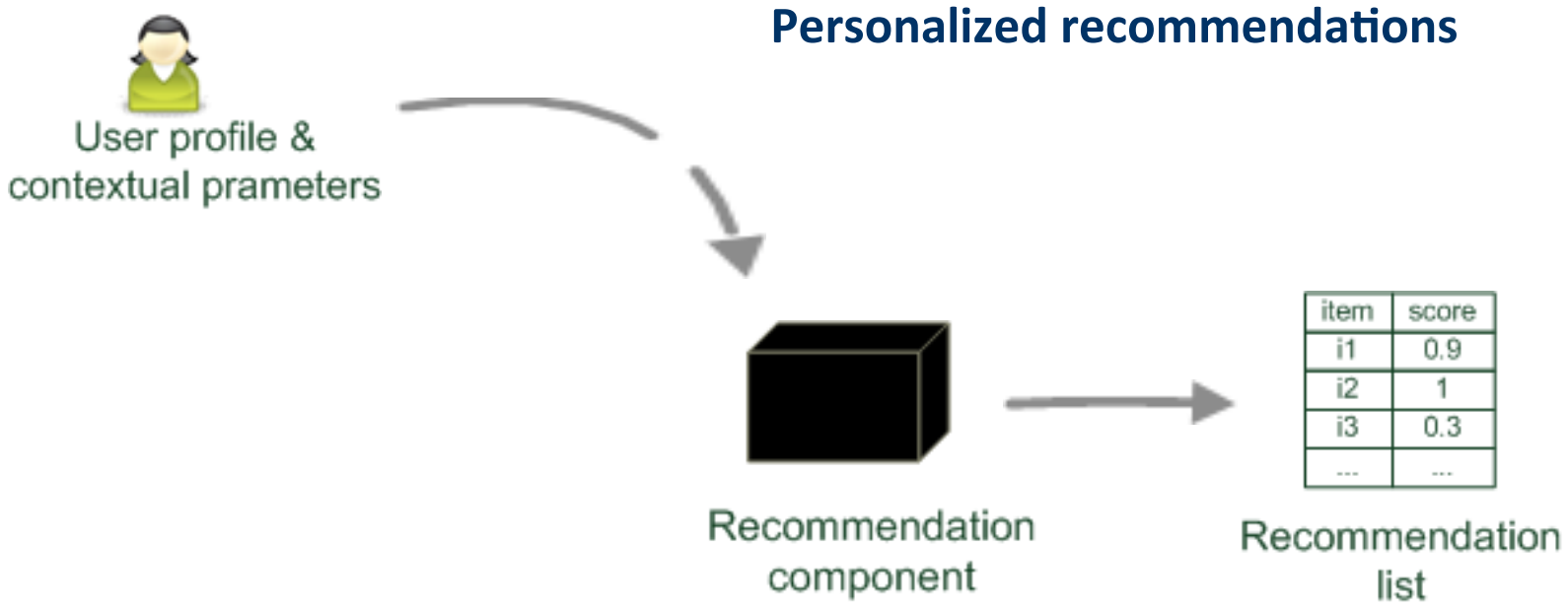
(1) Manning, Raghavan, and Schütze, *Introduction to information retrieval*, Cambridge University Press, 2008

Paradigms of recommender systems

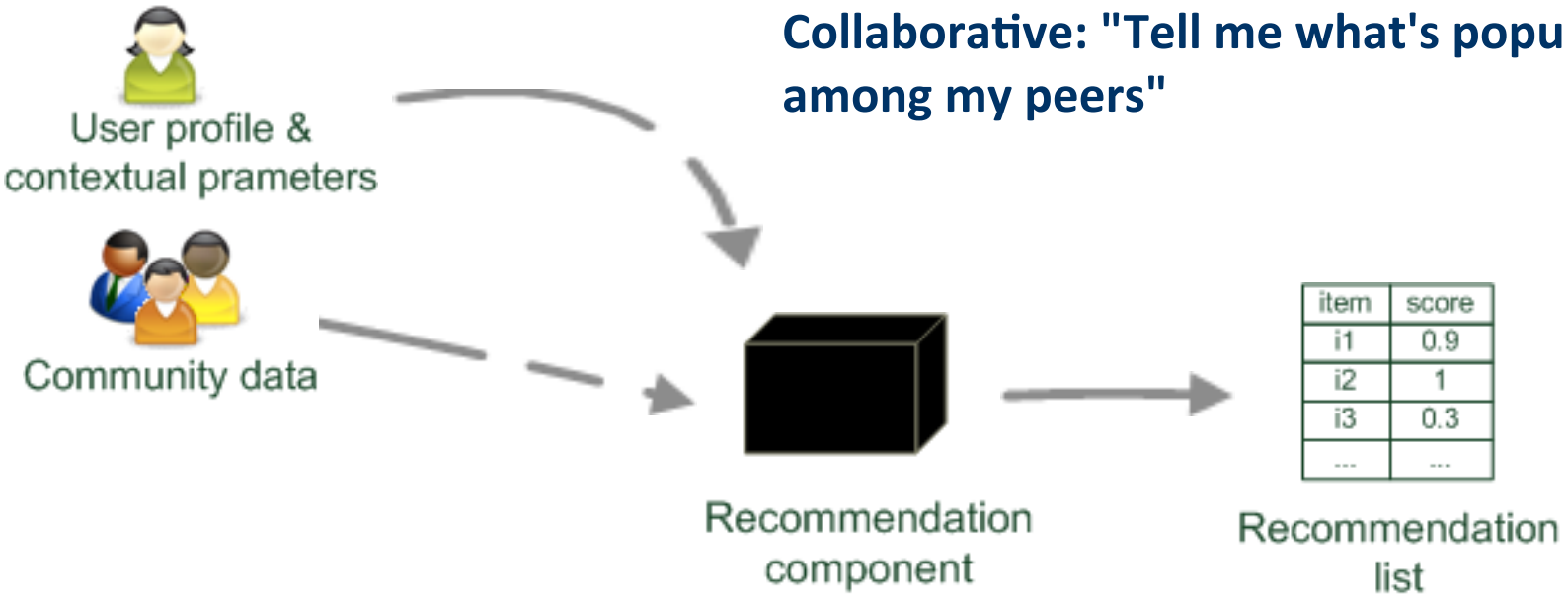
Recommender systems reduce information overload by estimating relevance



Paradigms of recommender systems



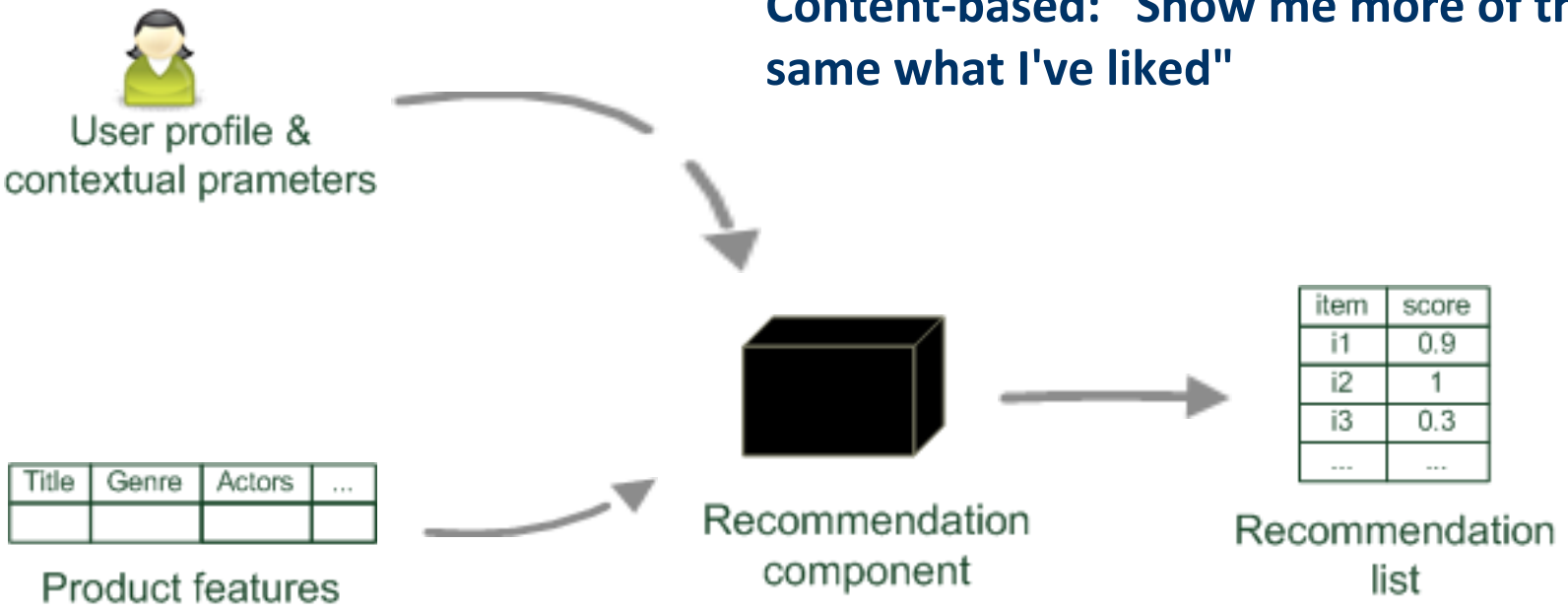
Paradigms of recommender systems



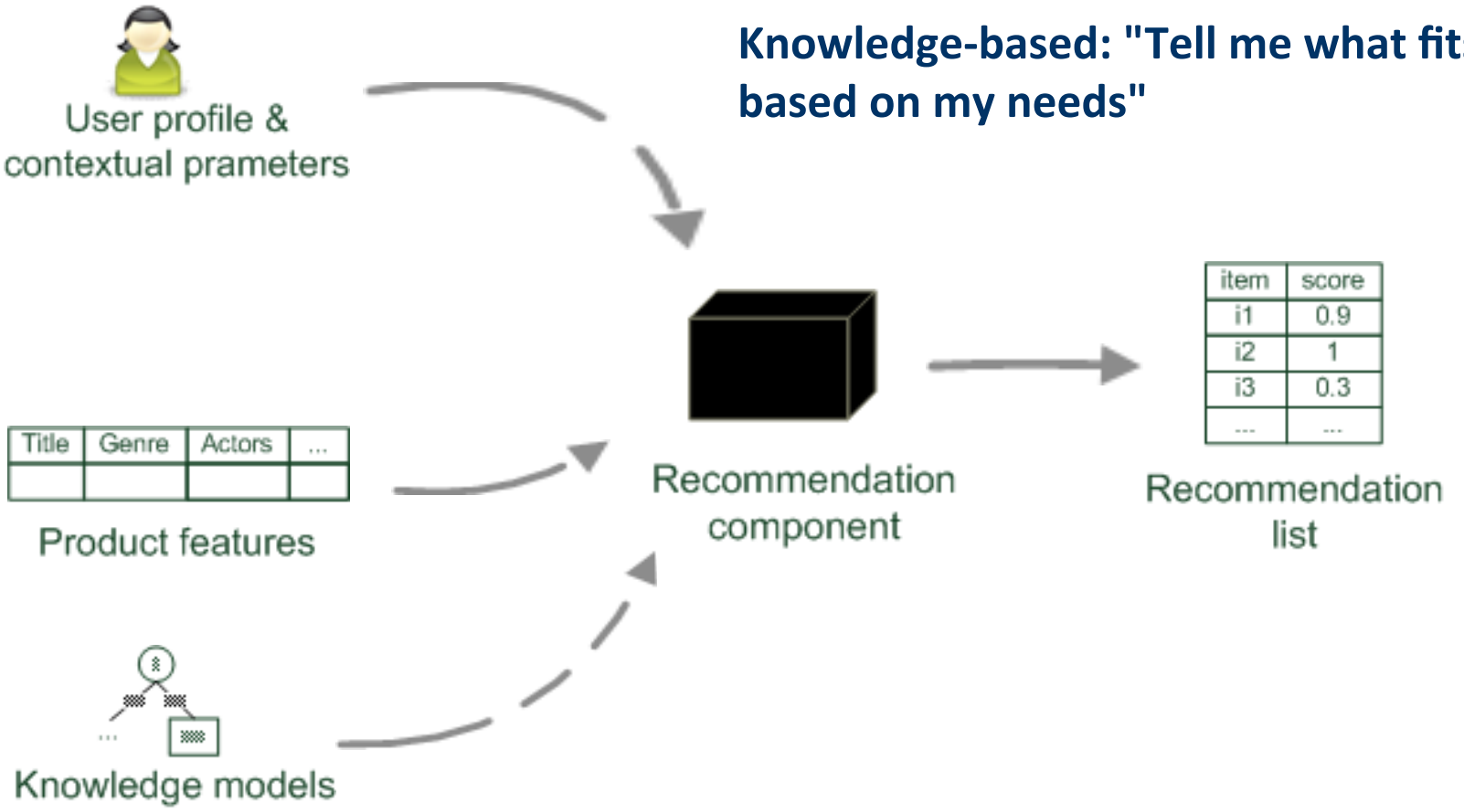
Collaborative: "Tell me what's popular among my peers"

Paradigms of recommender systems

Content-based: "Show me more of the same what I've liked"

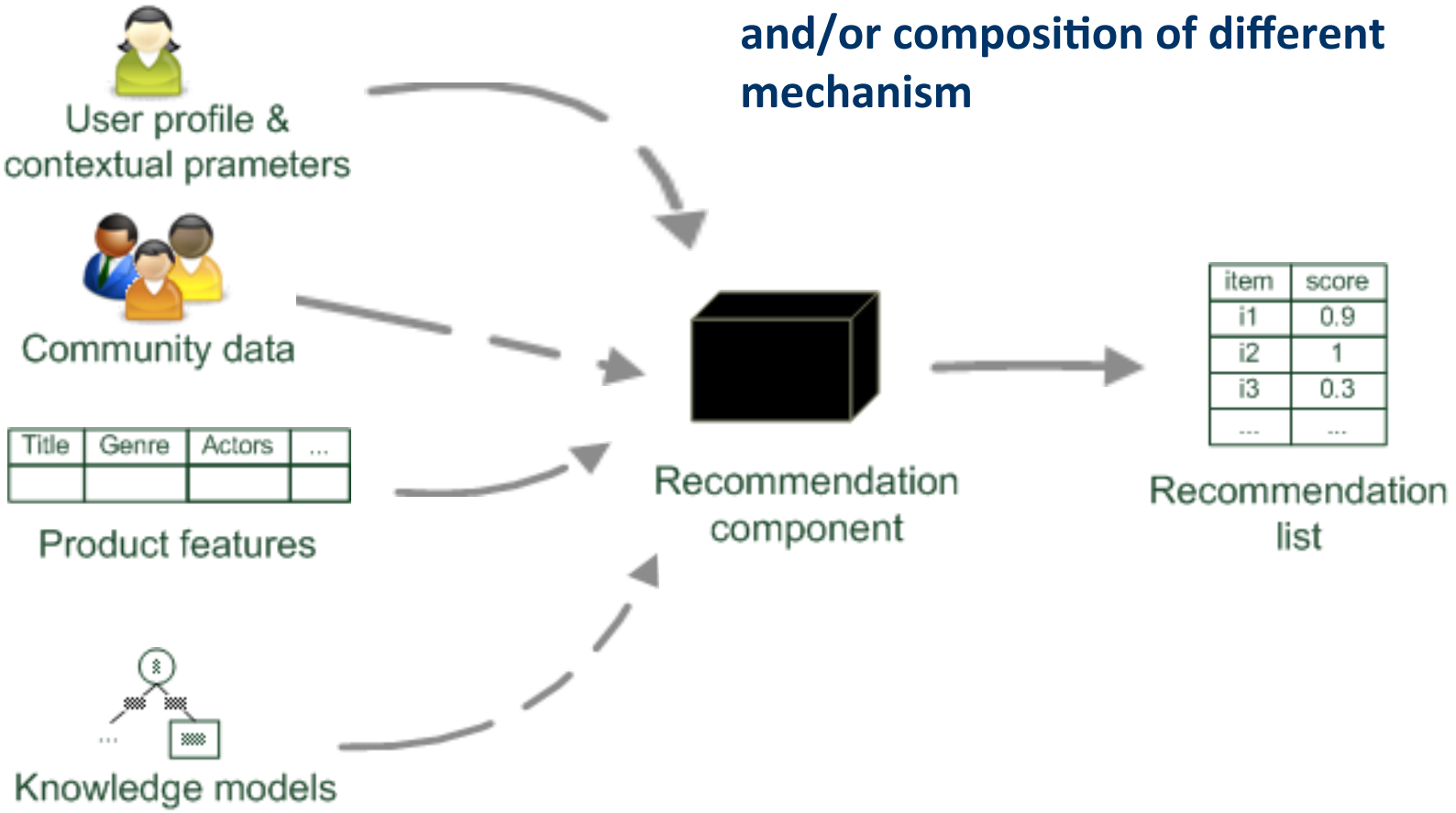


Paradigms of recommender systems



Paradigms of recommender systems

Hybrid: combinations of various inputs and/or composition of different mechanism



Kaggle Competitions

http://www.kaggle.com

e.com/about

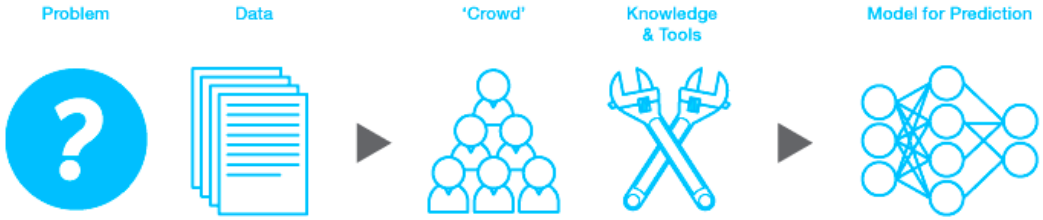
MSN Twitter Facebook CNN Wikipedia eBay NY Times ESPN Amazon Weather Channel BBC Yahoo! Finance - Po...

kaggle

Geoffrey Fox About Hosting Center All Competitions Users Forums Wiki Blog Data Science Jobs

About Us | Our Team

Kaggle is an innovative solution for statistical/analytics outsourcing. We are the leading platform for predictive modeling competitions. Companies, governments and researchers present datasets and problems - the world's best data scientists then compete to produce the best solutions. At the end of a competition, the competition host pays prize money in exchange for the intellectual property behind the winning model.



Geoffrey Fox
View / Edit Profile

http://www.kaggle.com

Host a competition for...

Analytics
Get the world's best predictive model.

Data Exploration
Find the diamonds in your data.

Recruitment
Uncover objectively brilliant candidates.

Education
Free, powerful classroom competitions.

Calling all data-driven startups.
"Let the crowd be your cofounder"
Kaggle Startup Program now open!

On the Forums

Predicting sunspot numbers
Scoring Models for Competitons
Kaggle as an academical
Medley: a new R package for
Get a "Data Creator" tag on your
Work for CIA

On the Blog

Colorado Succeeds Succeeds! ...
Explorer: Long Distance Kaggl...
Getting Started with Pandas -...

GE Quests

GE Flight Quest

Think you can change the future of flight?

11 days next deadline
148 teams
$250,000

GE Hospital Quest

Think it's possible to make hospital visits hassle-free? GE does.

13 days next deadline
$100,000

Featured Competitions

Event Recommendation Engine Challenge

Predict what events our users will be interested in based on user actions, event metadata, and demographic information.

10 days next deadline
166 teams
$5,000

Heritage Health Prize

Identify patients who will be admitted to a hospital within the next year using historical claims data. (Enter by 06:59:59 UTC Oct 4 2012)

58 days next deadline
1513 teams
$3 million

Blue Book for Bulldozers

Predict the auction sale price for a piece of heavy equipment to create a "blue book" for bulldozers.

2 months next deadline
78 teams
$10,000

On the Forums

- Predicting sunspot numbers
- Scoring Models for Competitons
- Kaggle as an academical
- Medley: a new R package for
- Get a "Data Creator" tag on your
- Work for CIA

On the Blog

- Colorado Succeeds Succeeds! ...
- Explorer: Long Distance Kaggl...
- Getting Started with Pandas -...
- Webapps for Data Scientists -...
- Newsletter: New Year, New Com...
- GE Hospital Quest: Milestone ...

76292 **participants**

205901 **entries**

Find a competition

163 competitions found, 17 active

Search competitions

- (•) All competitions
- ( ) Enterable

Status

- [x] Active
- [x] Completed

Type

- [x] Prospect
- [x] Open to all
- [x] Private
- [x] Limited

Sponsor

- [x] in-Class (student competition)

Competition Name	Reward	Teams	Deadline
in-Class **UCI Math77B: Collaborative Filtering** Did you laugh? Predict whether a joke is funny for Math 77B Collaborative Filtering at UCI!	Kudos	22	89 months
 **Titanic: Machine Learning from Disaster** Getting Started Competition, with tutorials in Excel, Python and introduction to Random Forests.	Knowledge	1669	7 months
 **Digit Recognizer** Classify handwritten digits in this "Getting Started" competition.	Knowledge	873	5 months
limited / in-Class **Predicting Who Will Get Sick** This predictive modeling competition is for students affiliated with the Rocky Mountain regional chapter of INFORMS.	Private	1	2 months
 **Blue Book for Bulldozers** Predict the auction sale price for a piece of heavy equipment to create a "blue book" for bulldozers.	$10,000	78	2 months
 **Heritage Health Prize** Identify patients who will be admitted to a hospital within the next year using historical claims data. (Enter by 06:59:59 UTC Oct 4 2012)	$3,000,000	1513	58 days
limited / in-Class **Oxford CS PoS tagging task - Hilary 2013** Danish Part of speech tagging challenge.	Private	3	50 days

UCI Math77B: Collaborative Filtering

89 months to go

Friday, January 13, 2012

Kudos • 22 teams Wednesday, June 10, 2020

Dashboard

Home/Info
Data
Make a submission
Forum
Leaderboard
My Team
My Submissions

Leaderboard

1. Ground Truth (1)
2. Igii (75)
3. Taylor (39)
4. Oliver Nowak (58)
5. jstroud (49)
6. Christine (24)
7. emilytn (4)
8. TomHall (6)

Competition Details » Get the Data » Make a submission

Did you laugh? Predict whether a joke is funny for Math 77B Collaborative Filtering at UCI!

- Jester dataset includes user ratings ranging from -10 to +10 for 100 jokes. For more detail description on the dataset, refer to here.
- There are 21,983 users for training and (non-overlapping) 3,000 users for test.
- The dataset includes 27.5% of missing values (not rated items) and these are represented as '99' in the .mat file.
- We removed three ratings per each user in testset and marked them as '55'. Thus, your goal is to predict these intentionally removed 9,000 ratings (3 items x 3000 users).
- For submittion, please refer to the 'Submission Instructions'
- RMSE will be used for evaluation.
- Once you submit a result, it will list you in the leaderboard based on the best score of your submissions.

Started: 9:22 pm, Friday 13 January 2012 UTC

Ends: 11:59 pm, Wednesday 10 June 2020 UTC (3,071 total days)

Examples of Recommender Systems

UCI Recommender Examples

1) Amazon.com (Search for your favorite book and look at what other books are recommended)
2) Bing or Google (Search for computer and view the ads that are displayed)
3) Netflix (If you have an account see how Netflix recommends new movies to you)
4) Jester (Try it out and rate some jokes)
5) MovieLens from Univ. Minesota. Find Movies you like
6) Pandora (Try adding some songs and observe what it plays for you) Watch this Time Video on the Pandora Recommender System (Then check out that Time Video webpage and observe it is recommending related videos ....)
7) Y! Music Yahoo Music Site
8) StumbleUpon (Watch this short video) Recommends web sites
9) YouTube (Observe how it recommends new video based on what you watched before)
10) and many, many more.....
11) http://www.ics.uci.edu/~welling/teaching/CS77Bwinter12/CS77B_w12.html
12) http://www.time.com/time/magazine/article/0,9171,1992403,00.html Time magazine on Recommender systems

Netflix on Recommender Systems

http://www.slideshare.net/xamat/building-largescale-realworld-recommender-systems-recsys2012-tutorial

Recsys 2012

Building Industrial-scale Real-world Recommender Systems

September 11, 2012

Xavier Amatriain
Personalization Science and Engineering - Netflix

@xamat

http://www.slideshare.net/xamat/building-largescale-realworld-recommender-systems-recsys2012-tutorial

Anatomy of Netflix Personalization

Everything is a Recommendation

http://www.slideshare.net/xamat/building-largescale-realworld-recommender-systems-recsys2012-tutorial

Everything is personalized

Note:
Recommendations are per household, not individual user

See comment at end

NETFLIX

http://www.slideshare.net/xamat/building-largescale-realworld-recommender-systems-recsys2012-tutorial

http://www.slideshare.net/xamat/building-largescale-realworld-recommender-systems-recsys2012-tutorial

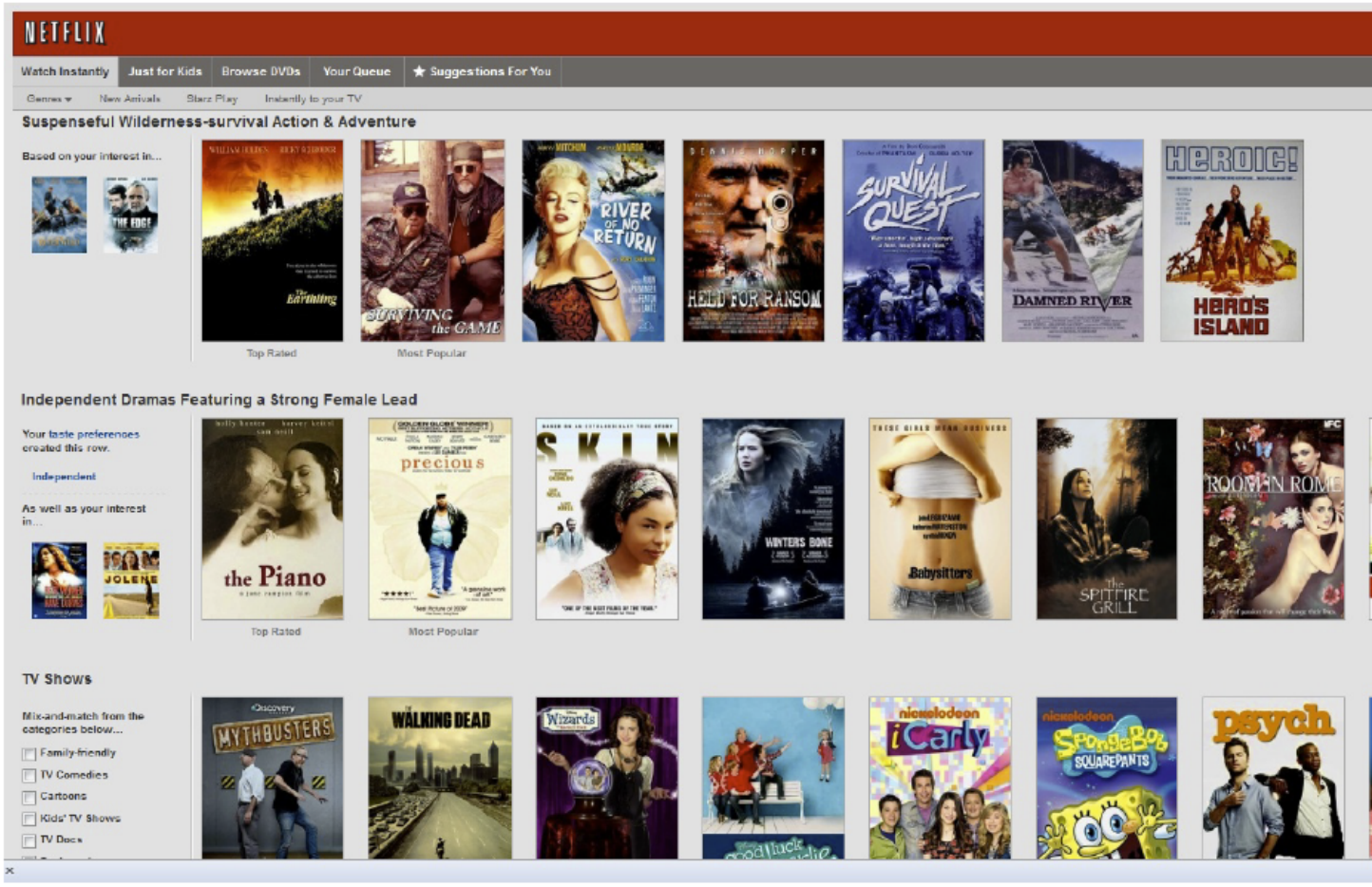
Genre rows

- Personalized genre rows focus on user interest
  - Also provide context and "evidence"
  - Important for member satisfaction – moving personalized rows to top on devices increased retention
- How are they generated?
  - **Implicit**: based on user's recent plays, ratings, & other interactions
  - **Explicit** taste preferences
  - **Hybrid**:combine the above
  - Also take into account:
  - Freshness - has this been shown before?
  - Diversity– avoid repeating tags and genres, limit number of TV genres, etc.

NETFLIX

http://www.slideshare.net/xamat/building-largescale-realworld-recommender-systems-recsys2012-tutorial

Genres - personalization



NETFLIX

http://www.slideshare.net/xamat/building-largescale-realworld-recommender-systems-recsys2012-tutorial

Anatomy of a Personalization - Recap

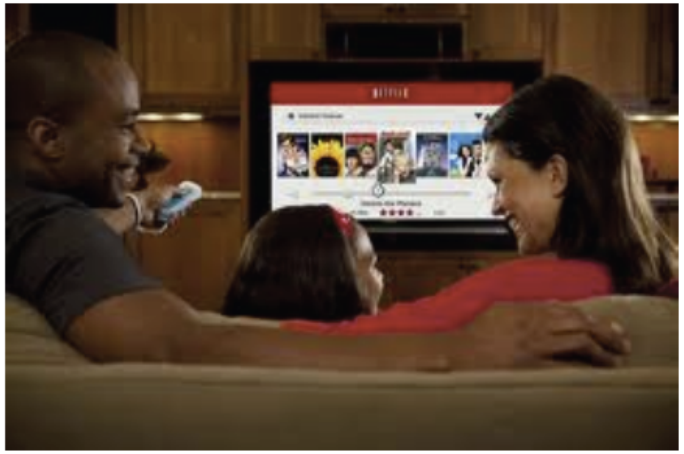
- Everything is a recommendation: not only rating prediction, but also ranking, row selection, similarity…
- We strive to make it easy for the user, but…
- We want the user to be aware and be involved in the recommendation process
- Deal with implicit/explicit and hybrid feedback
- Add support/explanations for recommendations
- Consider issues such as diversity or freshness

NETFLIX

http://www.slideshare.net/xamat/building-largescale-realworld-recommender-systems-recsys2012-tutorial

April 2013: The last two quarters have each brought more than 2 million new streaming subscriber signups. That gives Netflix a current total of nearly 29.2 million subscribers

Big Data @Netflix



- 25M+ subscribers
- Ratings: 4M/day
- Searches: 3M/day
- Plays: 30M/day
- 2B hours streamed in Q4 2011
- 1B hours in June 2012

NETFLIX

http://www.slideshare.net/xamat/building-largescale-realworld-recommender-systems-recsys2012-tutorial

Models

- Logistic/linear regression
- Elastic nets
- Matrix Factorization
- Markov Chains
- Clustering
- LDA **Latent Dirichlet Allocation**
- Association Rules
- Gradient Boosted Decision Trees
- …

NETFLIX

at/building-

der-systems-

Netflix Prize

COMPLETED

Home | Rules | Leaderboard | Update

Netflix Prize: Forum

Forum for discussion about the Netflix Prize and dataset.

Index User list Rules Search Register Login

You are not logged in.

Announcement

Congratulations to team "BellKor's Pragmatic Chaos" for being awarded the $1M Grand Prize on September 21, 2009. This Forum is now read-only.

Pages: **1**

Index » Grand Prize » **Grand Prize awarded to team BellKor's Pragmatic Chaos**

2009-09-18 09:58:04 #1

prizemaster
Administrator
From: Netflix HQ
Registered: 2006-08-29
Posts: 181
Website

It is our great honor to announce the $1M Grand Prize winner of the Netflix Prize contest as team BellKor's Pragmatic Chaos for their verified submission on July 26, 2009 at 18:18:28 UTC, achieving the winning RMSE of 0.8567 on the test subset. This represents a 10.06% improvement over Cinematch's score on the test subset at the start of the contest. We congratulate the team of Bob Bell, Martin Chabbert, Michael Jahrer, Yehuda Koren, Martin Piotte, Andreas Töscher and Chris Volinsky for their superb work advancing and integrating many significant techniques to achieve this result.

The Prize was awarded in a ceremony in New York City on September 21st, 2009. We will post a video on this forum of the presentation the team delivered about their Prize algorithm. In accord with the Rules the winning team has prepared a system description consisting of three papers, which we both make public below.

Team BellKor's Pragmatic Chaos edged out team The Ensemble with the winning submission coming just 24 minutes before the conclusion of the nearly three-year-long contest. Historically the Leaderboard has only reported team scores on the quiz subset. The Prize is awarded based on teams' test subset score. Now that the contest is closed we will be updating the Leaderboard to report team scores on both the test and quiz subsets.

To everyone who participated in the Netflix Prize: You've made this a truly remarkable contest and you've brought great innovation to the field. We applaud you for your contributions and we hope you've enjoyed the journey. The Netflix Prize contest is now closed.

We will soon be launching a new contest, Netflix Prize 2. Stay tuned for more details.

The winning team's papers submitted to the judges can be found below. These papers build on, and require familiarity with, work published in the 2008 Progress Prize.

Y. Koren, "The BellKor Solution to the Netflix Grand Prize", (2009).

A. Töscher, M. Jahrer, R. Bell, "The BigChaos Solution to the Netflix Grand Prize", (2009).

M. Piotte, M. Chabbert, "The Pragmatic Theory solution to the Netflix Grand Prize", (2009).

COMPLETED

$1M Prize

http://www.slideshare.net/xamat/building-largescale-realworld-recommender-systems-recsys2012-tutorial

2007 Progress Prize

- KorBell team (AT&T) improved by 8.43%
- Spent ~2,000 hours
- Combined 107 prediction algorithms with linear equation
- Gave us the source code

NETFLIX

Netflix competition

http://www.ifi.uzh.ch/ce/teaching/spring2012/16-Recommender-

june 26, 2009 around three years after launch of netflix prize. BellKor Pragmatic Chaos (merge of three teams) broke the 10% barrier in improving accuracy on movie prediction. NetFlix data set: 100 million + ratings, 1.4 million used as test set. 1-5 stars. RMSE on test set.

More data usually beats better algorithms

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 the teams illustrate a broader point. Team A came up with a very sophisticated algorithm using the Netflix data. Team B used a very simple algorithm, but they added in additional data beyond the Netflix set: information about movie genres from the Internet Movie Database(IMDB). Guess which team did better?

Anand Rajaraman is Senior Vice President at Walmart Global eCommerce, where he heads up the newly created @WalmartLabs,

http://anand.typepad.com/datawocky/2008/03/more-data-usual.html

20120117berkeley1.pdf Jeff Hammerbacher

http://www.slideshare.net/xamat/building-largescale-realworld-recommender-systems-recsys2012-tutorial

Ranking

- Ranking = **Scoring** + **Sorting** + **Filtering** bags of movies for presentation to a user
- Goal: Find the best possible ordering of a set of *videos* for a *user* within a specific *context* in real-time
- Objective: maximize consumption
- Aspirations: Played & "enjoyed" titles have best score
- Akin to CTR forecast for ads/search results

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

CTR = Click-through rate

NETFLIX

http://www.slideshare.net/xamat/building-largescale-realworld-recommender-systems-recsys2012-tutorial

Consumer (Data) Science

Consumer OR Life Style Informatics

NETFLIX

http://www.slideshare.net/xamat/building-largescale-realworld-recommender-systems-recsys2012-tutorial

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

Netflix foretells 'House of Cards' success with Cassandra big data engine

- House of Cards, staring Kevin Spacey, is the first major TV show to completely bypass the usual television ecosystem of networks and cable operators and premier on the streaming service Netflix.
- It may seem like Netflix took a big risk buying in unproven content rather than licensing content that was already successful. In reality, however, Netflix knew that the series would be a hit, based on data about the viewing habits of its 33 million users.
- Using the NoSQL database Apache Cassandra, Netflix was able to gather real-time data about the programmes its customers were watching, their demographics and viewing patterns, and build up an authoritative picture of the kind of content that would be well received.
- http://news.techworld.com/applications/3437514/netflix-foretells-house-of-cards-success-with-cassandra-big-data-engine/

A/B Testing from Wikipedia

- In web development and marketing, A/B testing or split testing is an experimental approach to web design (especially user experience design), which aims to identify changes to web pages that increase or maximize an outcome of interest (e.g., click-through rate for a banner advertisement).
- As the name implies, two versions (A and B) are compared, which are identical except for one variation that might impact a user's behavior. Version A might be the currently used version, while Version B is modified in some respect.
- For instance, on an e-commerce website the purchase funnel is typically a good candidate for A/B testing, as even marginal improvements in drop-off rates can represent a significant gain in sales.
- Multivariate testing or bucket testing is similar to A/B testing, but tests more than two different versions at the same time.

http://www.slideshare.net/xamat/building-largescale-realworld-recommender-systems-recsys2012-tutorial

Executing A/B tests

- Many different metrics, but ultimately trust user engagement (e.g. hours of play and customer retention)
- Think about significance and hypothesis testing
  - Our tests usually have thousands of members and 2-20 cells
- A/B Tests allow you to try radical ideas or test many approaches at the same time.
  - We typically have hundreds of customer A/B tests running
- Decisions on the product always data-driven

NETFLIX

http://www.slideshare.net/xamat/building-largescale-realworld-recommender-systems-recsys2012-tutorial

What to measure

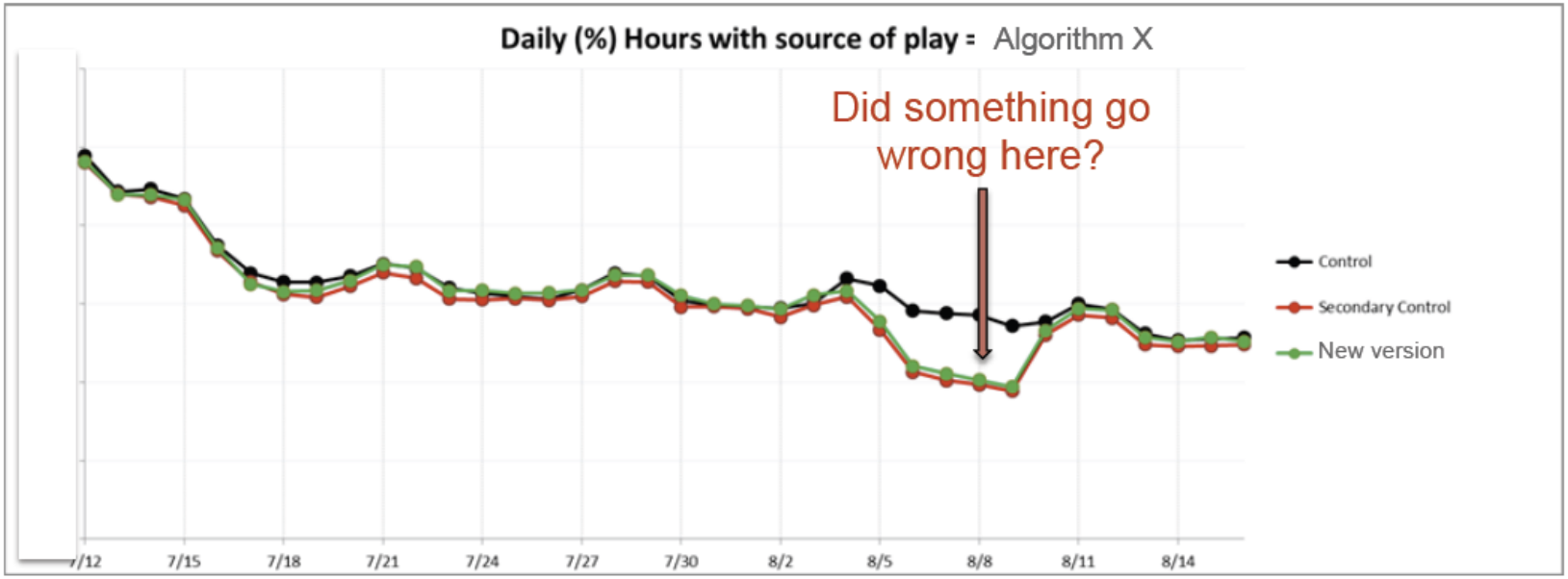
- OEC: Overall Evaluation Criteria
- In an AB test framework, the measure of success is key
- Short-term metrics do not always align with long term goals
  - E.g. CTR: generating more clicks might mean that our recommendations are actually worse
- Use long term metrics such as LTV (Life time value) whenever possible
  - In Netflix, we use member retention

NETFLIX

http://www.slideshare.net/xamat/building-largescale-realworld-recommender-systems-recsys2012-tutorial

What to monitor

- Algorithmic source for users
  - Monitor how users interact with different algorithms



NETFLIX

http://www.slideshare.net/xamat/building-largescale-realworld-recommender-systems-recsys2012-tutorial

http://techblog.netflix.com

http://www.slideshare.net/xamat/building-largescale-realworld-recommender-systems-recsys2012-tutorial

More to Recsys than Algorithms

- Not only is there more to algorithms than rating prediction
- There is more to Recsys than algorithms
  - User Interface & Feedback
  - Data
  - AB Testing
  - Systems & Architectures

NETFLIX

Interesting Unanswered Question from Student

- What is impact of users representing say a family e.g. father and daughter and so ratings are blended (if represent of interest to one family member) or inconsistent (if a given class of item sometimes rated by one and then other family member)
- Netflix recognizes issue (requires diversity in choices) but does not explain in detail