Build a Better Netflix, Win a Million Dollars?

Lester Mackey

August 10, 2014

NETFLIX

- Rents & streams movies and TV shows
- 100,000 movie titles
- 26 million customers

Recommends “Movies You’ll ♥”

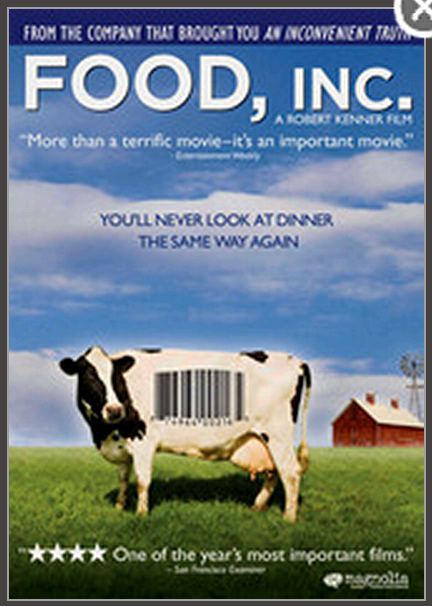
Recommending Movies You'll ♥

NETFLIX | Watch Instantly | Just for Kids | Instant Queue | Suggestions for You | DVDs | LESTER | Movies, TV

Rate what you've seen to discover suggestions for you

Haven't Seen It

Hated it! Haven't Seen It Loved it!



Haven't Seen It

Recommending Movies You'll ♥

NETFLIX

LESTER Mackey | Your Account & Help

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

Movies, TV shows, actors, directors, genres

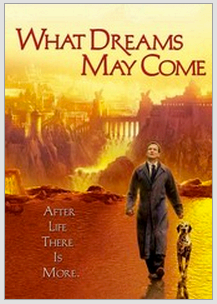
Genres | Instantly to your TV

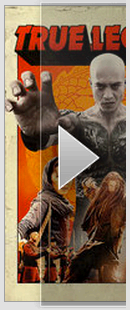
Visually-striking Sci-Fi & Fantasy

Based on your interest in…

Top Rated

Most Popular

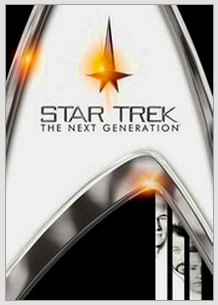




Feel-good TV Shows

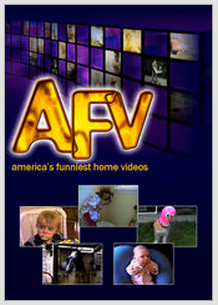
Your taste preferences created this row.

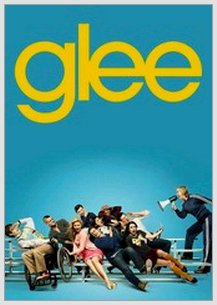
Feel-good TV Shows.



Top Rated

Most Popular

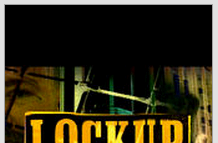


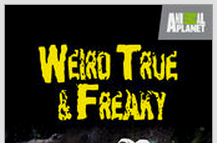


Documentaries

Mix-and-match from the categories below…

- ☐ Dark
- ☐ Science & Nature Docs





Recommending Movies You'll ♥

Sci-Fi & Fantasy

Add

Not Interested

Incepti

Becaus

Memen

Th

Batma

Inception

2010 **PG-13** **148 minutes**

Dom Cobb earns a tidy sum infiltrating the dreams c
corporate titans to steal their most closely held secre

Starring: Leonardo DiCaprio, Joseph Gordon-Levit

Director: Christopher Nolan

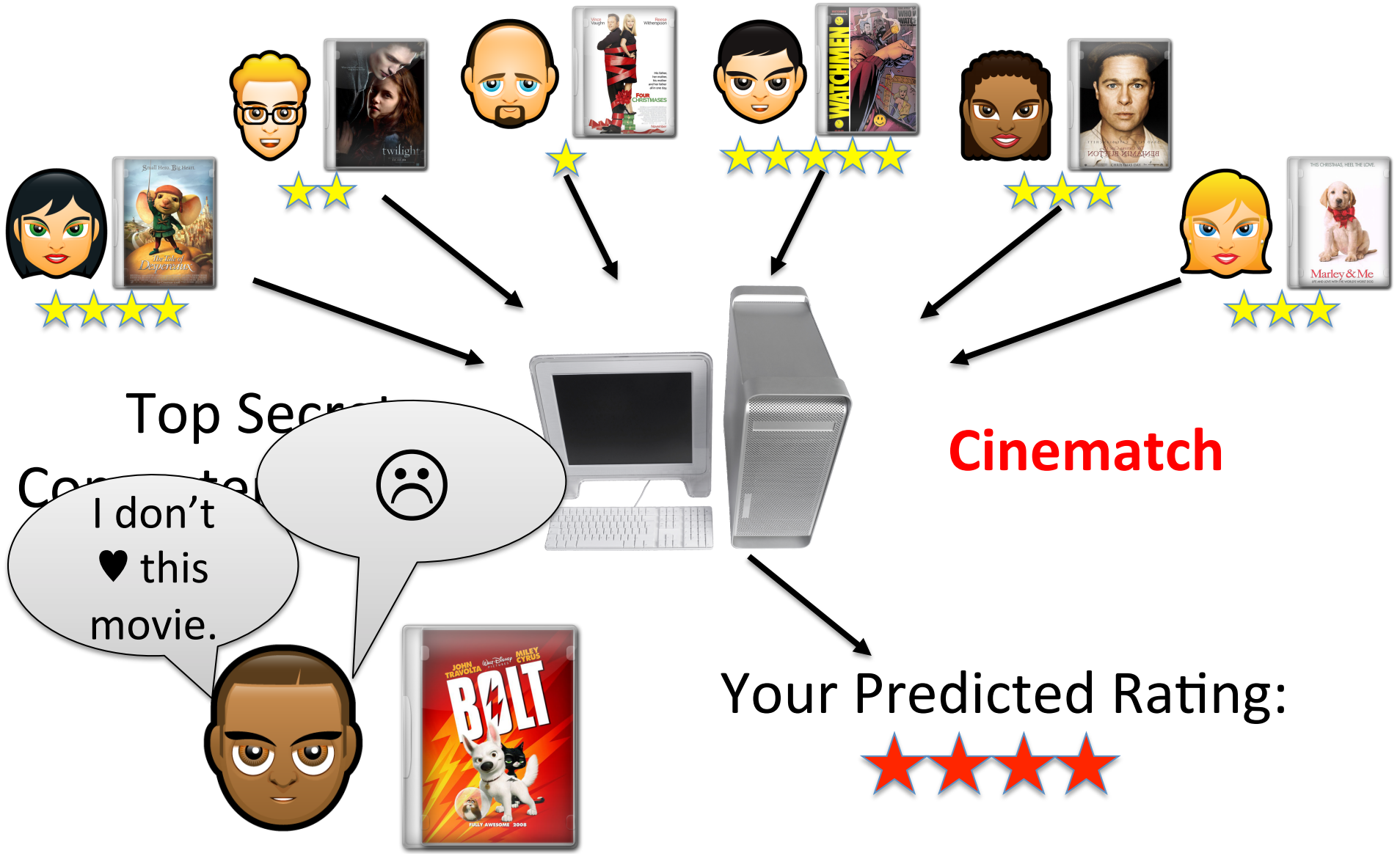
Genre: Sci-Fi & Fantasy

Availability: DVD and Blu-ray

4.4 Our best guess for LESTER

Recommended based on your interest in: *Batman Begins*, *The Matrix* and *Memento*

How This Works



Back at Netflix

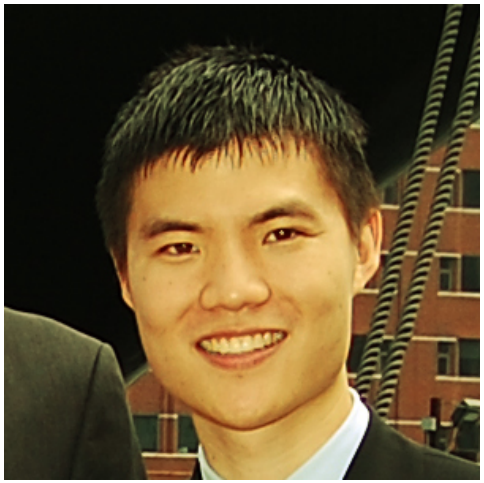
The Netflix Prize

October 2, 2006

- Contest open to the world
- 100 million movie ratings released to public
- **Goal:** Create computer program to predict ratings
- **$1 Million** Grand Prize for beating Cinematch accuracy by 10%
- **$50,000** Progress Prize for the team with the best predictions each year

5,100 teams from 186 countries entered

Dinosaur Planet



David Lin

David Weiss

Lester Mackey

Team Dinosaur Planet

The Ratings

- Training Set
  - What computer programs use to learn customer preferences
  - Each entry:    July 5, 1999
  - 100,500,000 ratings in total
  - 480,000 customers and 18,000 movies

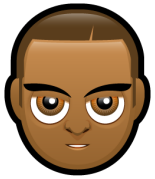
The Ratings: A Closer Look

Highest Rated Movies
The Shawshank Redemption
Lord of the Rings: The Return of the King
Raiders of the Lost Ark
Lord of the Rings: The Two Towers
Finding Nemo
The Green Mile

Most Divisive Movies
Fahrenheit 9/11
Napoleon Dynamite
Pearl Harbor
Miss Congeniality
Lost in Translation
The Royal Tenenbaums

How the Contest Worked

- Quiz Set & Test Set
  - Used to evaluate accuracy of computer programs
  - Each entry:



Rating unknown!

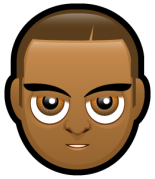
Sept. 9, 2006

- Each team predicts Quiz Set and Test Set ratings once per day
- Netflix displays Quiz score on public Leaderboard

Leaderboard (Week 2)

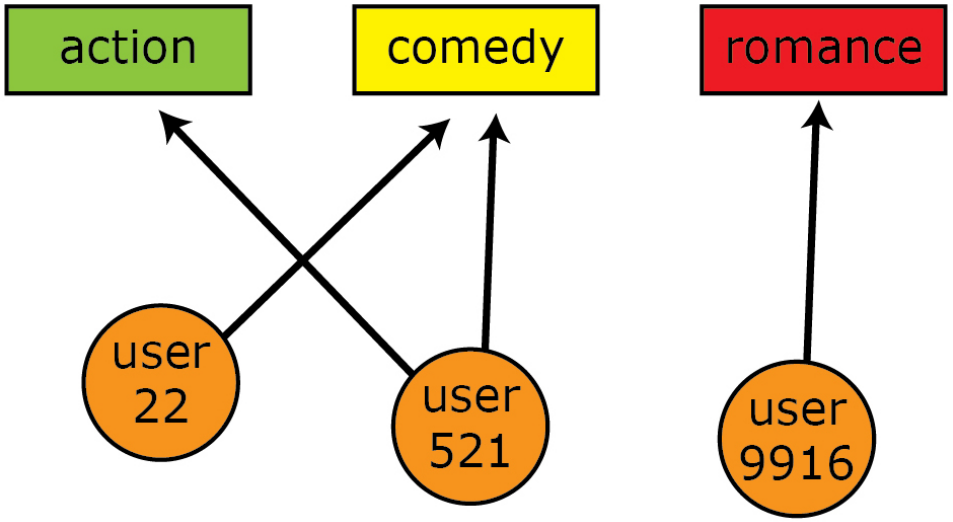
Team Name	Best Score	% Improvement
No Grand Prize candidates yet	--	--
Grand Prize - RMSE <= 0.8563		
The Thought Gang	0.9413	1.06
Progress Prize 2007 - RMSE <= 0.9419		
wxyzconsulting.com	0.9430	0.88
Sparkling_Destiny	0.9488	0.27
Cinematch score on quiz subset - RMSE = 0.9514		
Baseline0	0.9525	-0.12
CodeMonkey	0.9571	-0.60
Bjornson	0.9648	-1.41
jsnell	0.9670	-1.64

How the Contest Worked

- Quiz Set & Test Set
  - Used to evaluate accuracy of computer programs
  - Each entry:   Rating unknown! Sept. 9, 2006
- Each team predicts Quiz Set and Test Set ratings once per day
- Netflix displays Quiz score on public Leaderboard
- Test score is hidden – but best Test score wins!
  - 10% improvement ➡ **$1 Million Grand Prize**
  - Most improvement in 1 year ➡ **$50,000 Progress Prize**

A First Approach: Clustering

- Divide users (or movies) into groups based on similarities



- Use group information to predict user ratings
  - e.g. The average action-lover gives Indiana Jones a 5
- Hard clustering: each user belongs to a single cluster
- Soft clustering: each user fractionally belongs to all clusters

Clustering with Missing Data

- Centroid-based clustering
  - Represent user by incomplete ratings vector, r_u

$$r_u = (1, 5, ?, ?, 3, ?, 4)$$

  - Represent cluster by centroid vector, c_k
    - Typically, c_k is average of user vectors in cluster k
  - Minimize (estimated) distance between users and their cluster centers
- Result: -0.3% improvement over Cinematch

Matching Cinematch

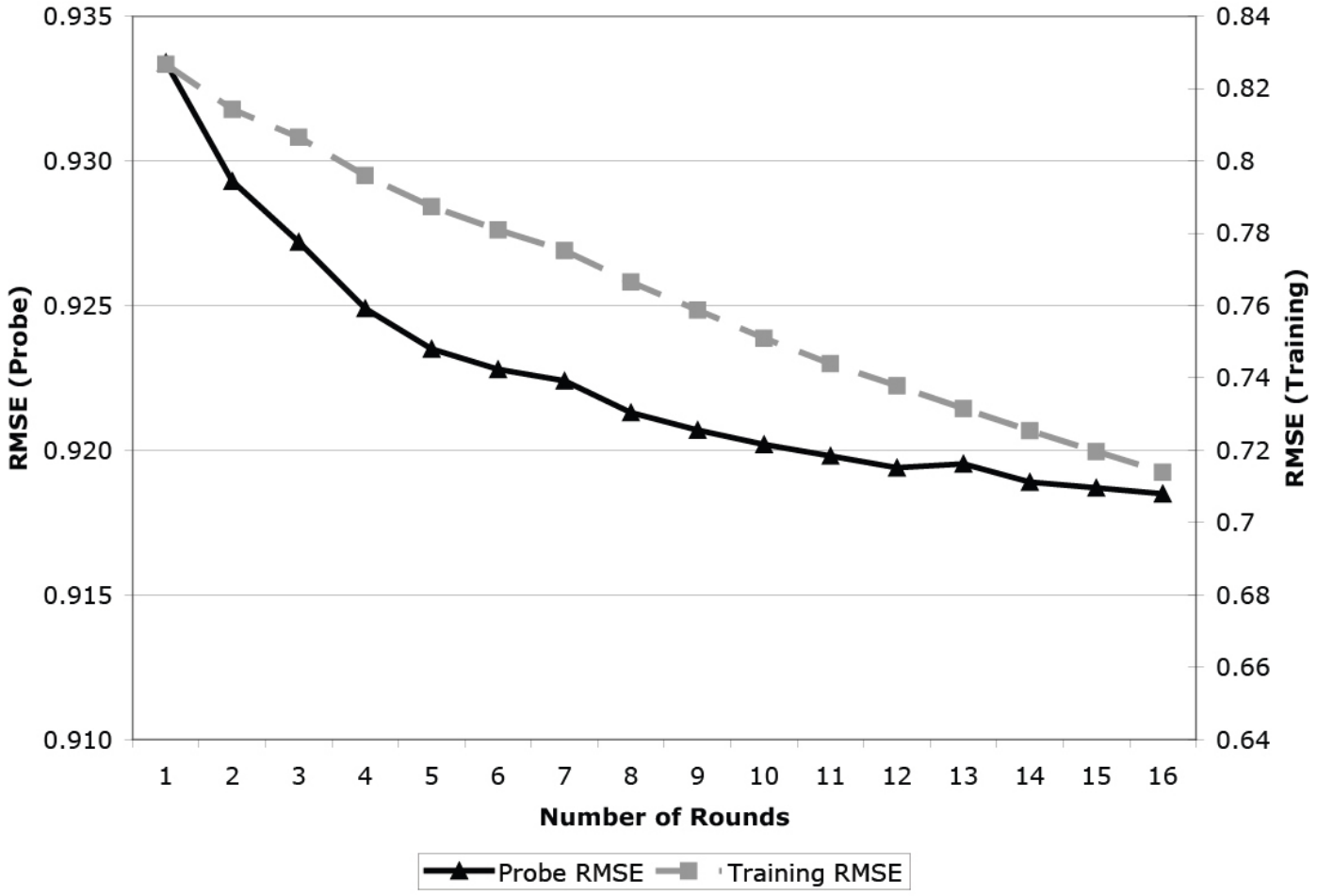
- Incorporate prior information
  - Positive ratings {3,4,5} vs. negative ratings {1,2}
    - Estimate $E[r|r \geq 3],\ E[r|r < 3],\ P(r < 3)$ and combine
  - Ordinal nature of rating data
    - Estimate $P(r < t)$ for $t \in \{2, 3, 4, 5\}$ and combine
  - Result: 0.5% improvement over Cinematch

Training on Errors

Recurring theme

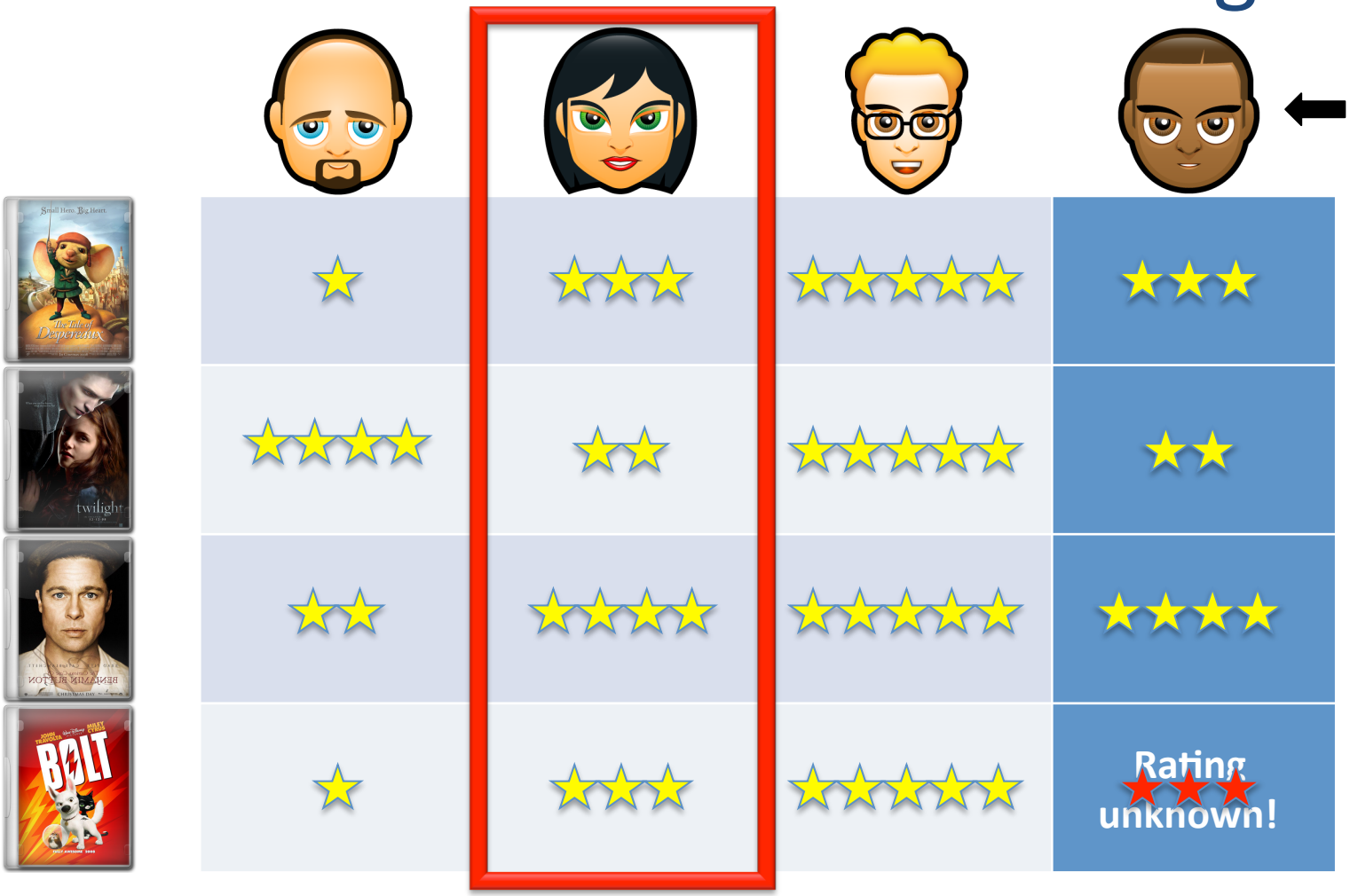
- Train one model to predict and hence correct the errors of another model
- Long history in statistics and machine learning
  - Tukey's twicing (1977)
  - Boosting (Schapire, 1990)
  - Gradient boosting (Friedman, 1999)
- e.g., Cluster on errors of clustering predictions

Clustering on Errors



Result: 3.0% improvement over Cinematch

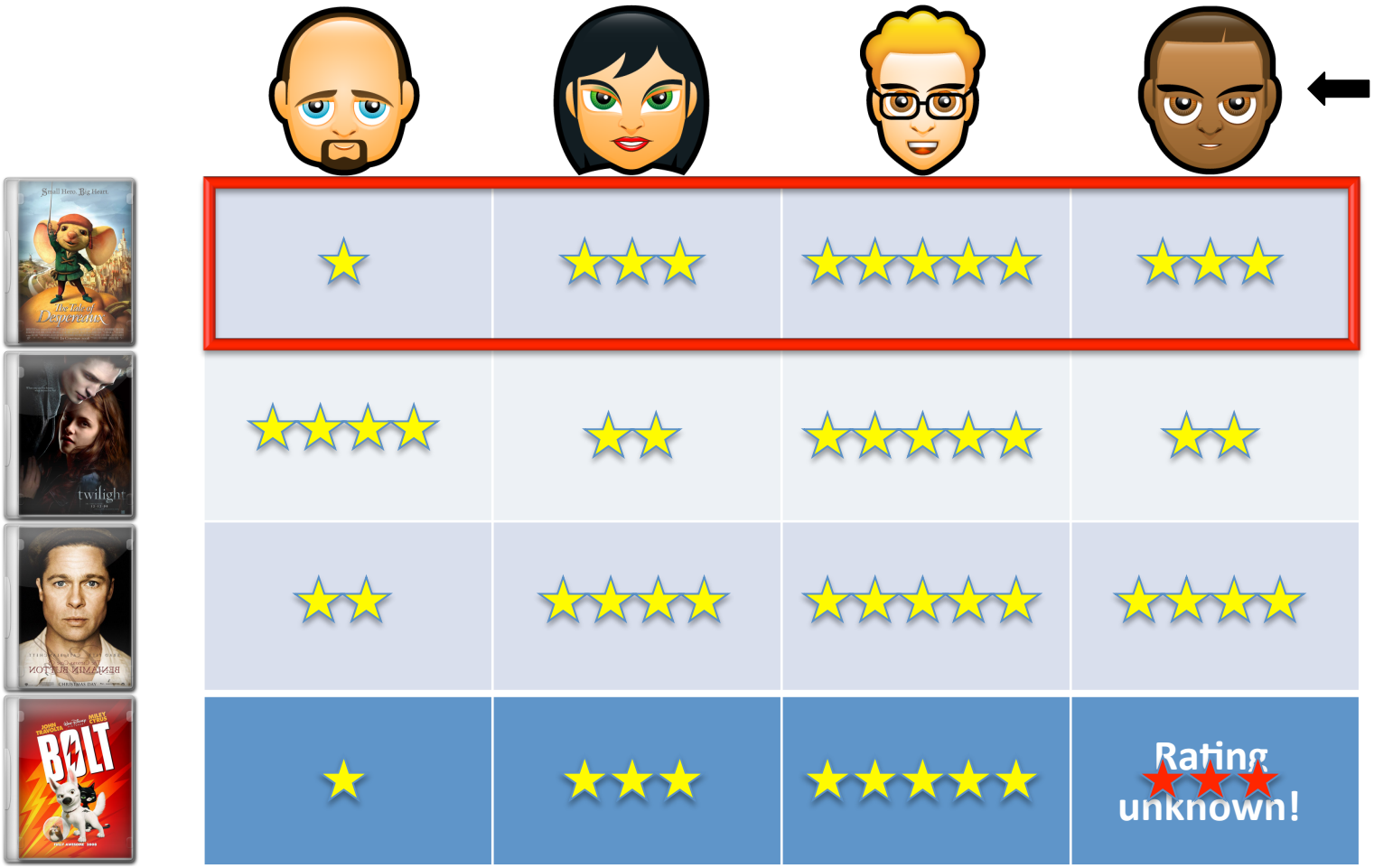
The Three Pillars: Nearest Neighbors



Nearest Neighbor Rule

- Find customer with the most similar ratings
- Use her rating as best guess for new guy's rating

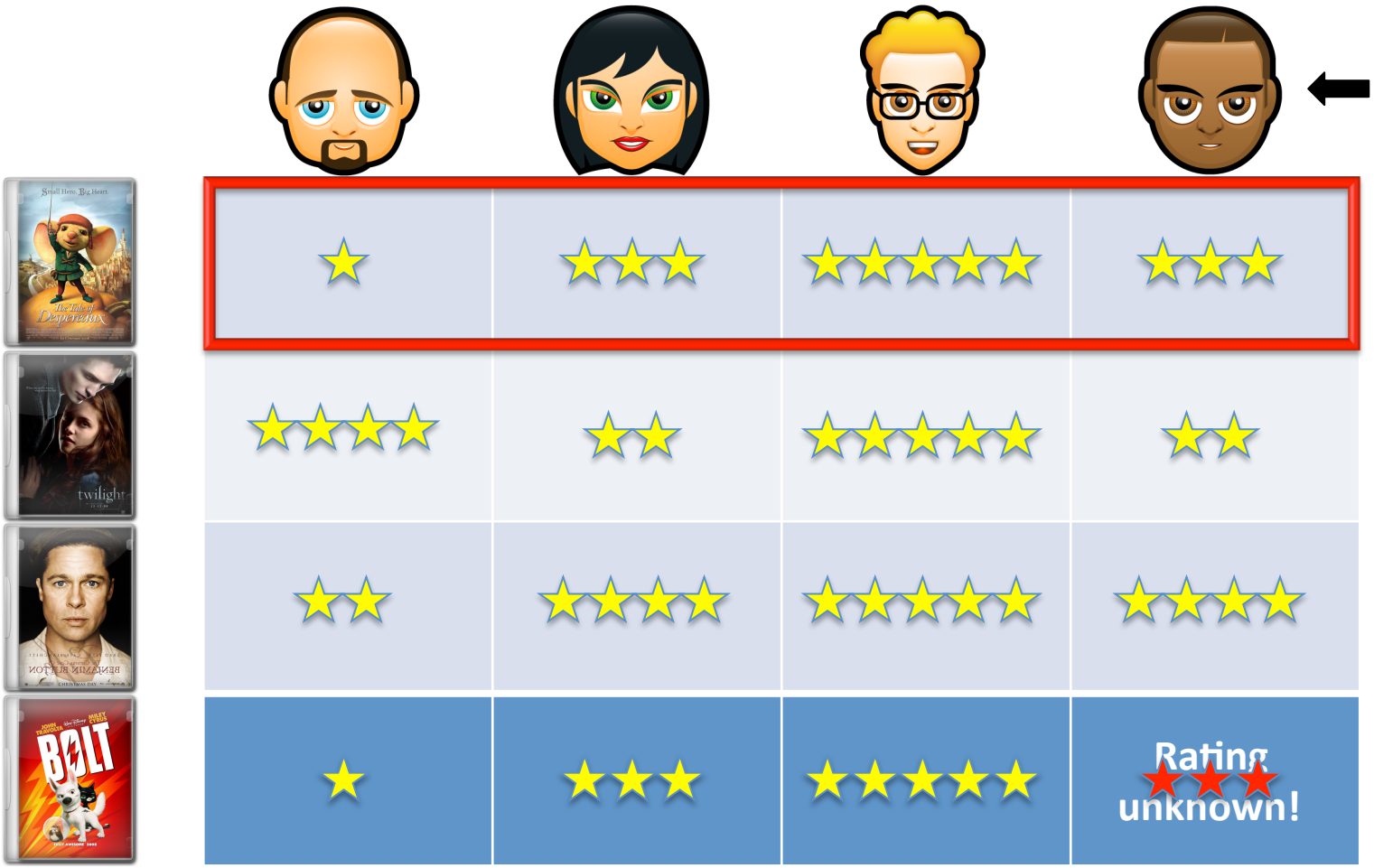
The Three Pillars: Nearest Neighbors



Nearest Neighbor Rule

- Find movie with the most similar ratings
- Use its rating as best guess for new guy's rating

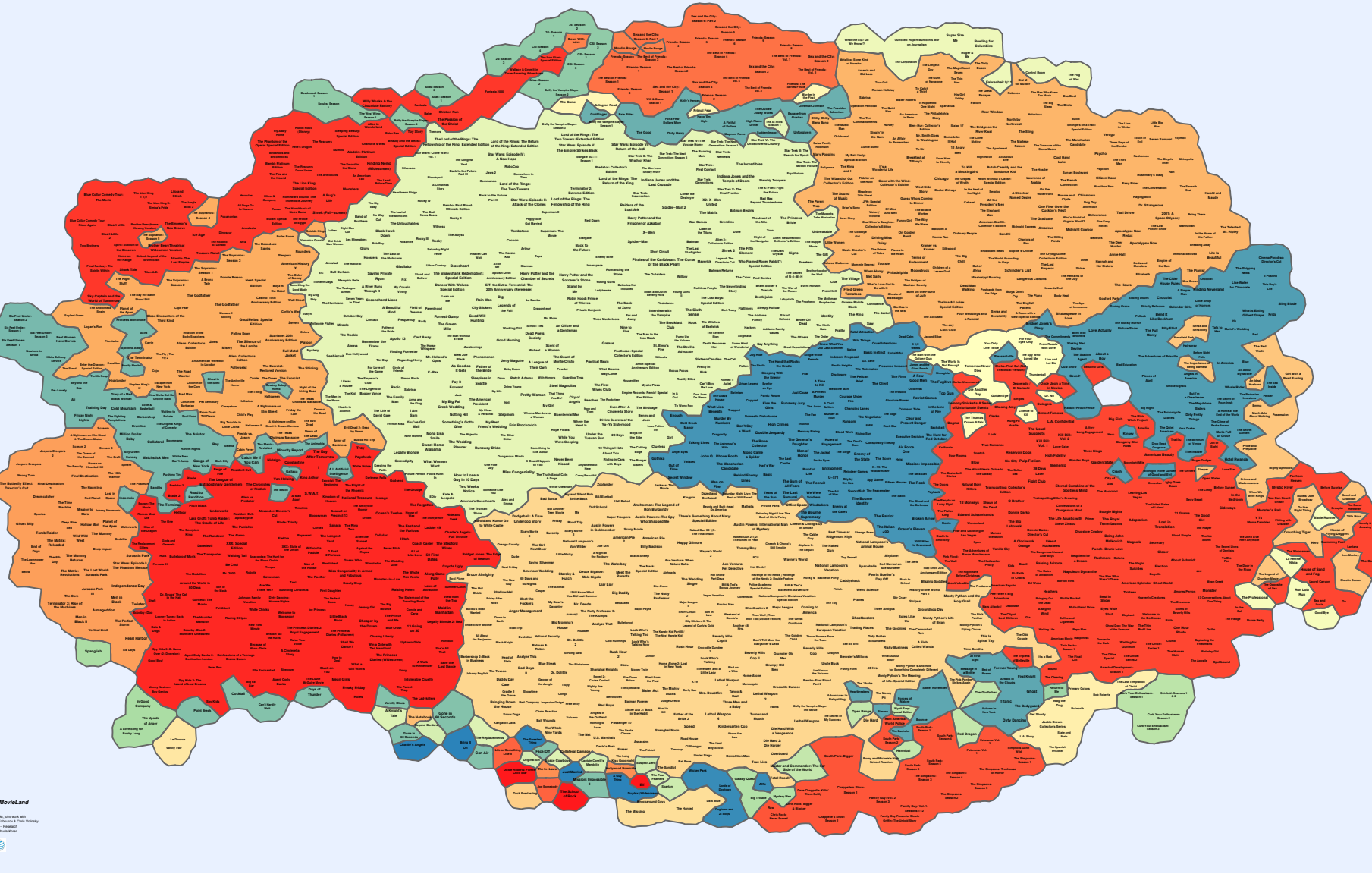
The Three Pillars: Nearest Neighbors



K-Nearest Neighbor Methods

- Classical KNN: 0.5% improvement
- KNN with learned weights: 4.6% improvement

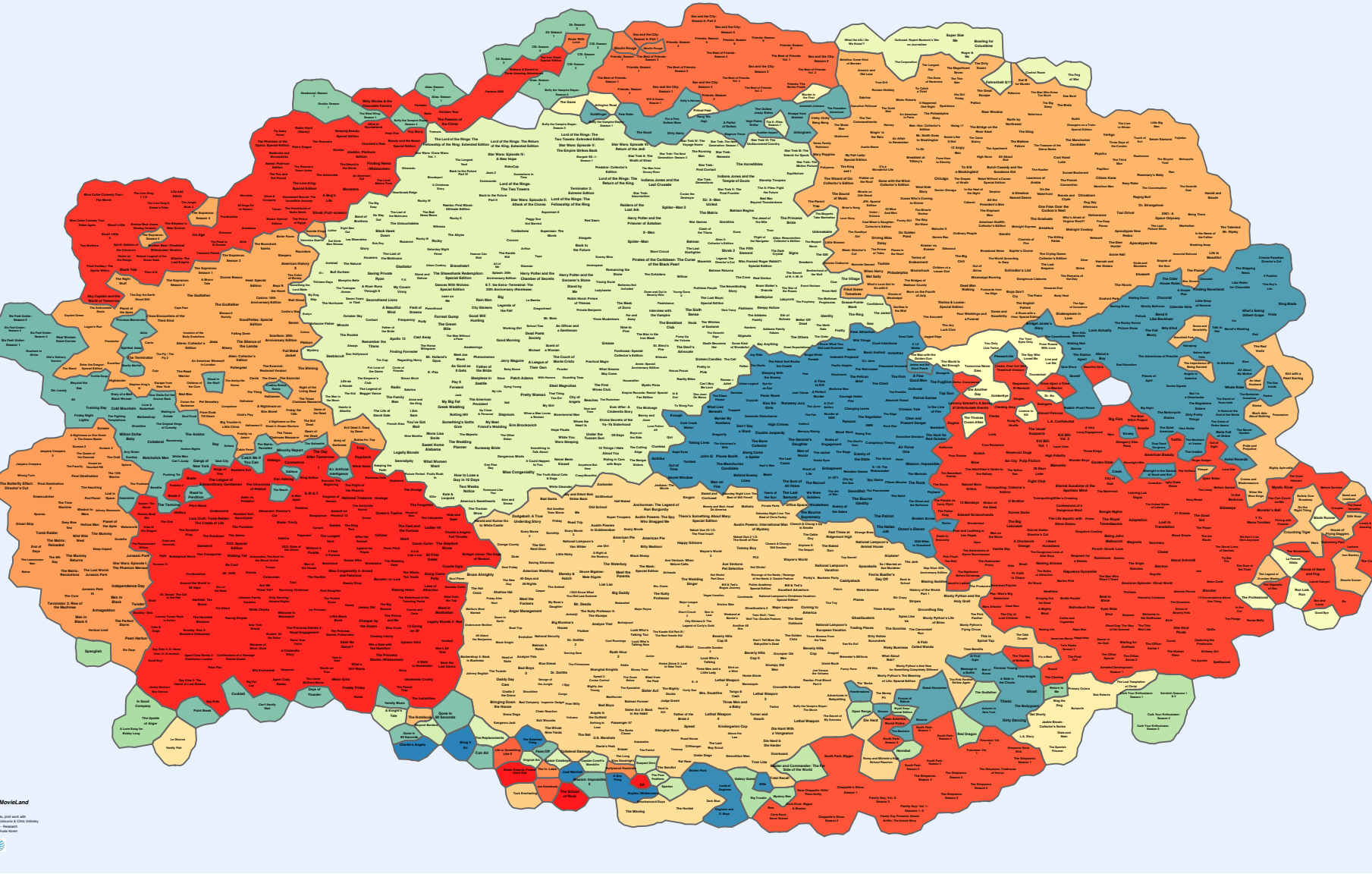
Map of Movie Neighbors



Map of Movie Neighbors

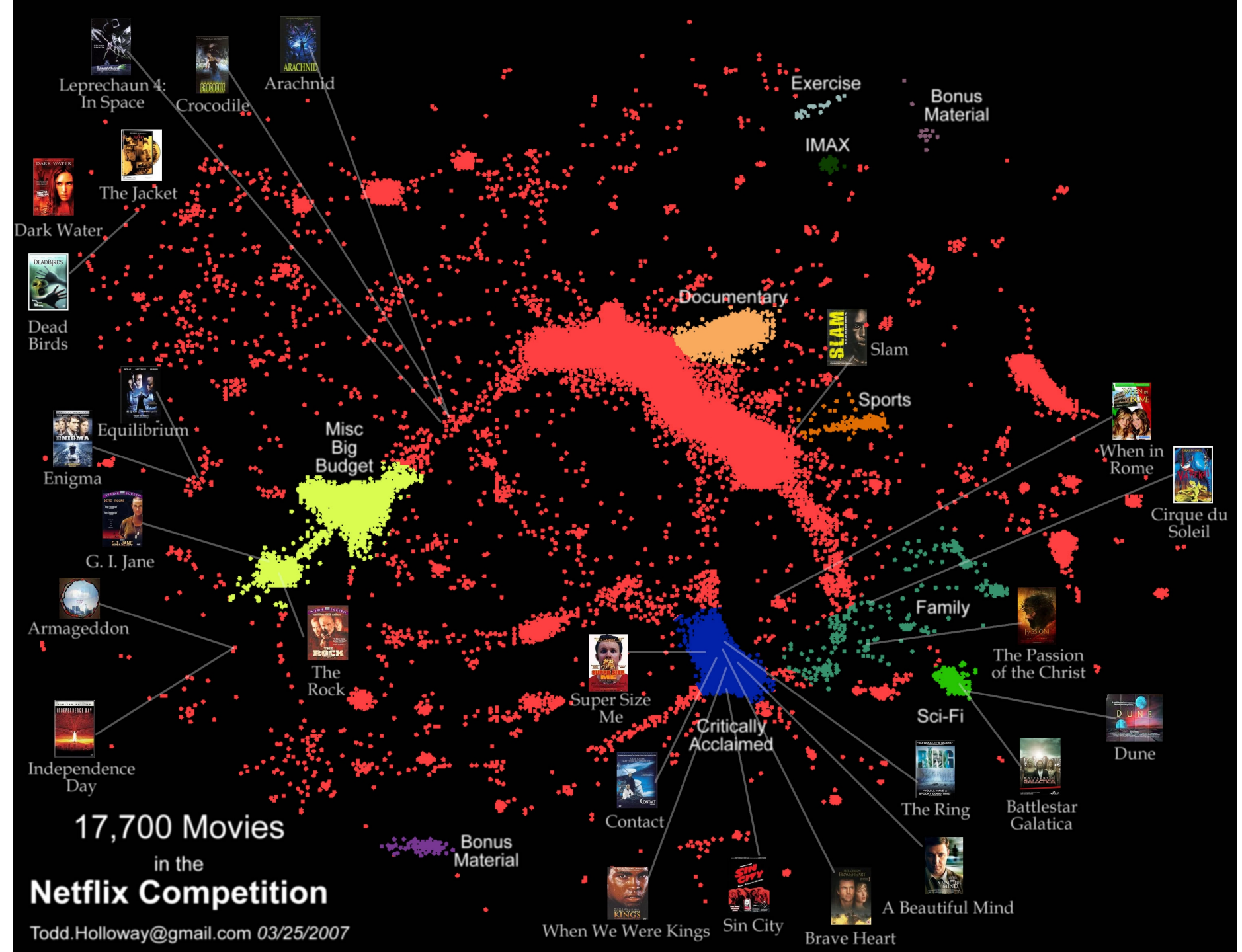
Produced by Yifan Hu, joint work with Emden Gansner, Stephen Kobourov & Chris Volinsky, Data from Yehuda Koren, Copyright AT&T

Map of Movie Neighbors



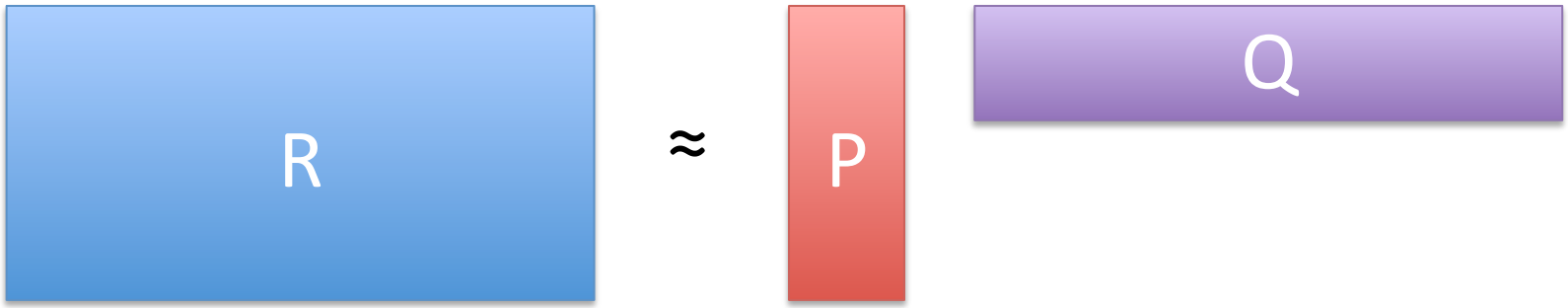
Map of Movie Neighbors

Produced by Yifan Hu, joint work with Emden Gansner, Stephen Kobourov & Chris Volinsky, Data from Yehuda Koren, Copyright AT&T


Leprechaun 4: In Space
Crocodile
Arachnid
Exercise
Bonus Material
IMAX
The Jacket
Dark Water
Dead Birds
Documentary
Slam
Sports
Equilibrium
Misc Big Budget
Enigma
When in Rome
Cirque du Soleil
G. I. Jane
Armageddon
Family
The Rock
The Passion of the Christ
Super Size Me
Critically Acclaimed
Sci-Fi
Dune
Independence Day
The Ring
Battlestar Galatica
Contact
17,700 Movies
in the
Netflix Competition
Bonus Material
When We Were Kings
Sin City
Brave Heart
A Beautiful Mind
Todd.Holloway@gmail.com 03/25/2007

The Three Pillars

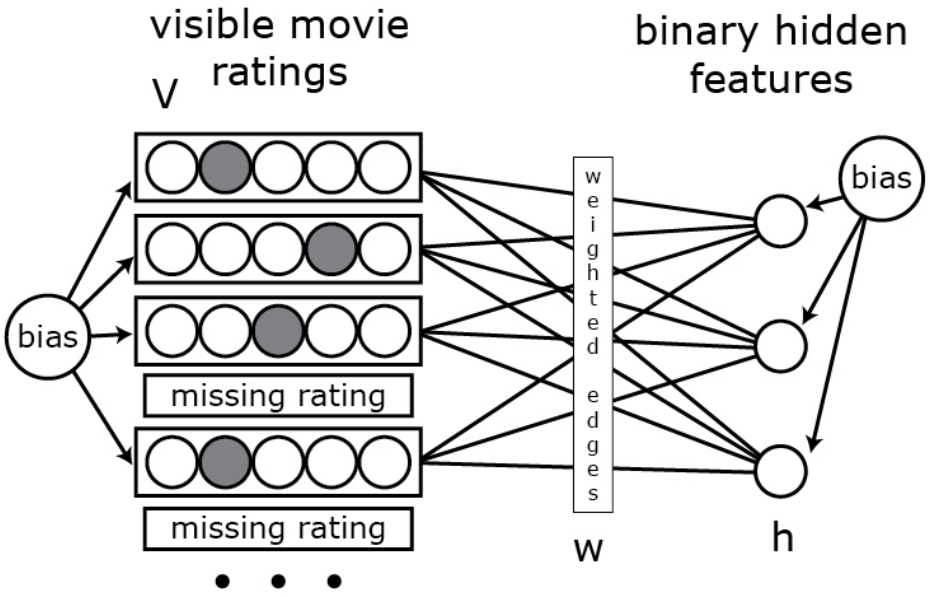
- Matrix factorization



  - Alternating least squares
  - Online/Stochastic gradient descent
    - Dec. 2, 2006, Simon Funk (Brandyn Webb)
    - Enormous impact: anyone could beat Cinematch after a few minutes of training
- Typical improvement: 4%

The Three Pillars

- Restricted Boltzmann machines
  - May 2007, Salakhutdinov and Mnih



- Typical improvement: 5%

Milestones

- Spring 2007: Dinosaur Planet enters “Top 10”
- June 2007: DP graduates from college

Model Ensembling

⬆
Recurring theme

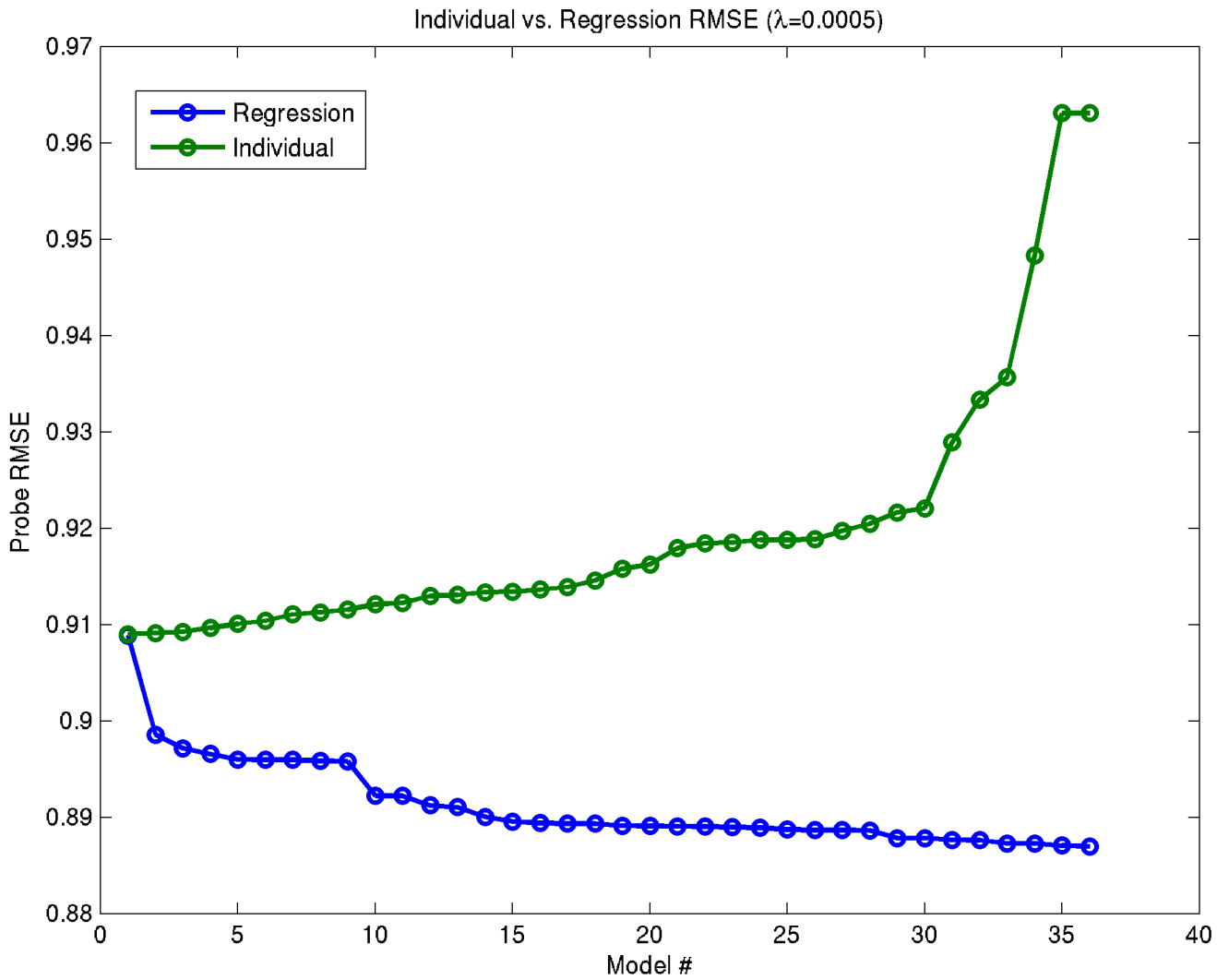
- Combining the predictions of multiple models to yield improved performance
- Motivation:
  - Diminishing returns from optimizing a single algorithm
    - Best single model improvement: 8.24% (Aron Miller)
    - The Ensemble's final improvement: 10.09%
  - Different models capture different aspects of the data
    - Global commonalities of MF vs. Local similarities of KNN
  - Variance reduction from uncorrelated inputs

Model Ensembling

- Stacked linear regression (Wolpert, Breiman)
  - Target = held-out ratings, r
  - Covariates = model predictions, P
  - Tikhonov regularization to reduce overfitting

$$\min_{\beta} \|\mathbf{r} - \mathbf{P}\beta\|^2 + \lambda\|\beta\|^2$$

Model Ensembling



Result: 2.0% improvement over best model

Model Ensembling Variations

- Add user, item, and date features as covariates
  - User rating count
  - Date of rating
  - Average inverse user rating count per movie
- Sparse regression: L1 regularizer or nonnegativity constraints
- Regress on pairwise interactions
  - Greedy selection or bagging with random subsets
- Result: 7.96% improvement over Cinematch

The First Progress Prize

- Sept. 3, 2007
  - Dinosaur Planet takes first place (from reigning champion BellKor)
- One hour later
  - BellKor takes back first place

Recurring theme

The First Progress Prize

One day before the deadline…

Rank	Team Name	Best Score	% Improvement
--	No Grand Prize candidates yet	--	--
Grand Prize - RMSE <= 0.8563			
1	BellKor	0.8728	8.26
2	Gravity	0.8750	8.03
3	Dinosaur Planet	0.8753	8.00
4	ML@UToronto A	0.8787	7.64
5	Arek Paterek	0.8789	7.62
6	basho	0.8805	7.45
7	NIPS Reject	0.8808	7.42
8	Ensemble Experts	0.8841	7.07
Progress Prize 2007 - RMSE: 0.9419			

The First Progress Prize

- Sept. 3, 2007
  - Dinosaur Planet takes first place (from reigning champion BellKor)
- One hour later
  - BellKor takes back first place

Recurring theme

- Sept. 19, 2007
  - Gravity contacts DP about potential collaboration
  - Gabor Takacs, Istvan Pilaszy, Bottyan Nemeth, Domonkos Tikk

Recurring theme

- Oct. 1, 2007
  - When Gravity and Dinosaurs Unite overtake BellKor with 8.38%
- 76 seconds later
  - BellKor ties with 8.38%
- Oct. 2, 2007
  - KorBell wins the first $50,000 progress prize with 8.43% improvement

The Power of Teamwork

- First Progress Prize
  -  joins forces with  for 8.38% improvement
  - BellKor improves to 8.43%, wins $50,000
  - 76 seconds later: BellKor ties with 8.38%
- Second Progress Prize
  - BigChaos joins BellKor for 9.44% and $50,000
- Grand Prize Team (GPT) founded by  + 
  - Anyone could join
  - The more you improve the GPT score, the bigger your share of the $1 million Grand Prize
  - Many joined and brought new techniques with them

Gaussian Missing Data Model (Roberts)

- Assume each vector of user ratings drawn from common multivariate Gaussian $\mathcal{N}(\mu, \Sigma)$
  - Incomplete vector of observed ratings drawn from marginal distribution
- Choose (μ, Σ) to maximize likelihood
  - Expectation-Maximization or gradient ascent
- Predict missing ratings as conditional expectation given observed ratings
- Result: 6.38% improvement

Feature-Weighted Linear Stacking (Sill, Takacs, Mackey, Lin)

- An adaptive approach to stacked linear regression
- Allow model ensembling weights to depend linearly on known features of the user, movie, and date
  - Did the user rate more than 3 movies on this date?
  - Log number of times the movie has been rated
  - Log number of distinct dates on which a user has rated
  - Log of average correlation between movies rated by user and movie to be predicted
- Result: 8.82% ➡ 9.46% improvement for GPT

The Last Call

Leaderboard **10.05%** Display top 20 leaders.

Rank	Team Name	Best Score	% Improvement	Last Submit Time
1	BellKor's Pragmatic Chaos	0.8558	10.05	2009-06-26 18:42:37
Grand Prize - RMSE <= 0.8563				
2	PragmaticTheory	0.8582	9.80	2009-06-25 22:15:51
3	BellKor in BigChaos	0.8590	9.71	2009-05-13 08:14:09
4	Grand Prize Team	0.8593	9.68	2009-06-12 08:20:24

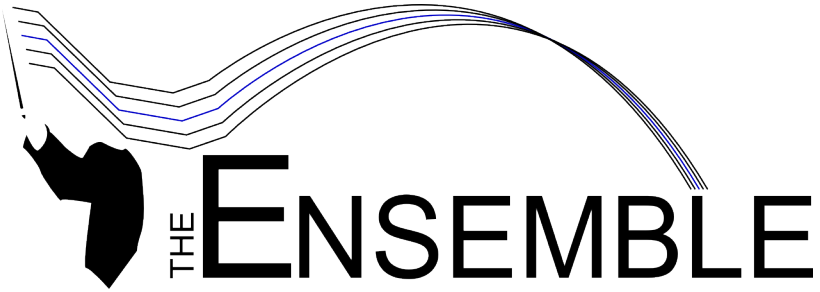
- Remaining teams had 30 days to respond
- Over the next 24 days, **30 individuals** from **11 countries** combined forces to challenge BPC

The Last Call

- June 26, 2009
  - Top 3 teams (BellKor, BigChaos, and Pragmatic Theory) combine to pass the Grand Prize threshold
  - Initiates 30 day last call period for $1 million grand prize
- June 30, 2009
  - GPT begins deeper collaboration
  - Message board to share ideas, server to share code and predictions
- July 5, 2009
  - Vandelay Industries! contacts GPT about potential collaboration
- July 7, 2009
  - Opera Solutions joins Vandelay Industries!
- July 20, 2009
  - The Ensemble is born

The Ensemble

- Grand Prize Team
  - Gravity
    - Gabor Takacs, Istvan Pilaszy, Bottyan Nemeth, Domonkos Tikk
  - Dinosaur Planet
    - David Lin, Lester Mackey, David Weiss
  - Joe Sill
  - Ces Bertino
  - Dan Nabutovsky
  - William Roberts
  - Wojtek Kulik
  - Willem Mestrom
  - David Purdy



- Vandelay Industries!
  - Greg McAlpin
  - Bill Bame
  - Bo Yang
  - Chris Hefele
  - Jeff Howbert
  - Xiang Liang
  - Larry Ya Luo
  - Aron Miller
  - Steve Pagliarulo
  - Opera Solutions
    - Bruce Deng, Peng Zhou, Priyanka Rastog, Arvind Gangadha, Jacob Spoelstra
  - Craig Carmichael
  - Mike Linacre
  - Edward de Grijs
  - Clive Gifford
  - Feeds2
    - Nicholas Ampazis, George Tsagas

Learn more at http://the-ensemble.com/

The Road to the Grand Prize

- Next to Last Day
  - The Ensemble submits
  - 10.09% improvement on Quiz Set
- Final Day, 6:18pm
  - BellKor's Pragmatic Chaos responds
  - 10.09% improvement on Quiz Set
- Final Day, 6:38pm
  - The Ensemble makes its final submission
  - 10.10% improvement on Quiz Set
- Final Day, 6:42pm: Contest closes

The Other Road to the Grand Prize

- Next to Last Day
  - The Ensemble submits
  - 10.05% improvement on Test Set
- Final Day, 6:18pm
  - BellKor's Pragmatic Chaos responds
  - 10.06% improvement on Test Set
- Final Day, 6:38pm
  - The Ensemble makes its final submission
  - 10.06% improvement on Test Set
- Tie breaker: Time of submission

The End