INTRODUCTION TO DATA SCIENCE

This lecture is
based on course by E. Fox and C. Guestrin, Univ of Washington

28/01/2020

WFAiS UJ, Informatyka Stosowana
I stopień studiów

Recommending system: films

Mchiane learning: recommending system

- Personalizacja

YouTube

100 Hours a Minute

What do I care about?

Information overload

↓

Browsing is "history"

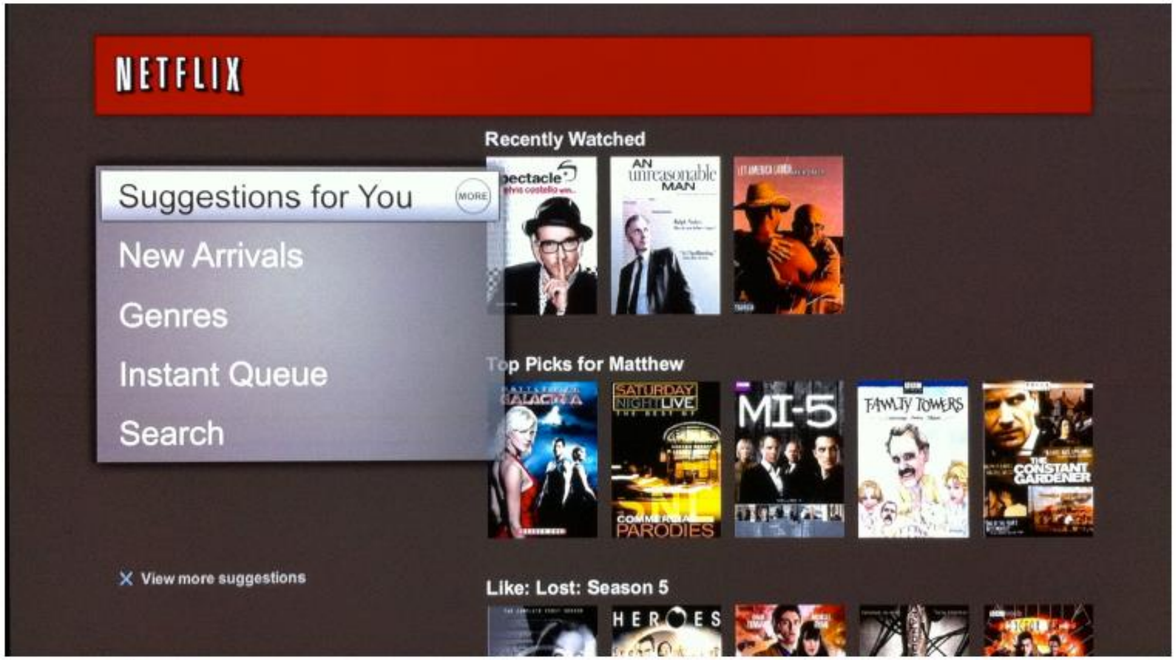
– Need new ways to discover content

Personalization: Connects *users & items*

viewers

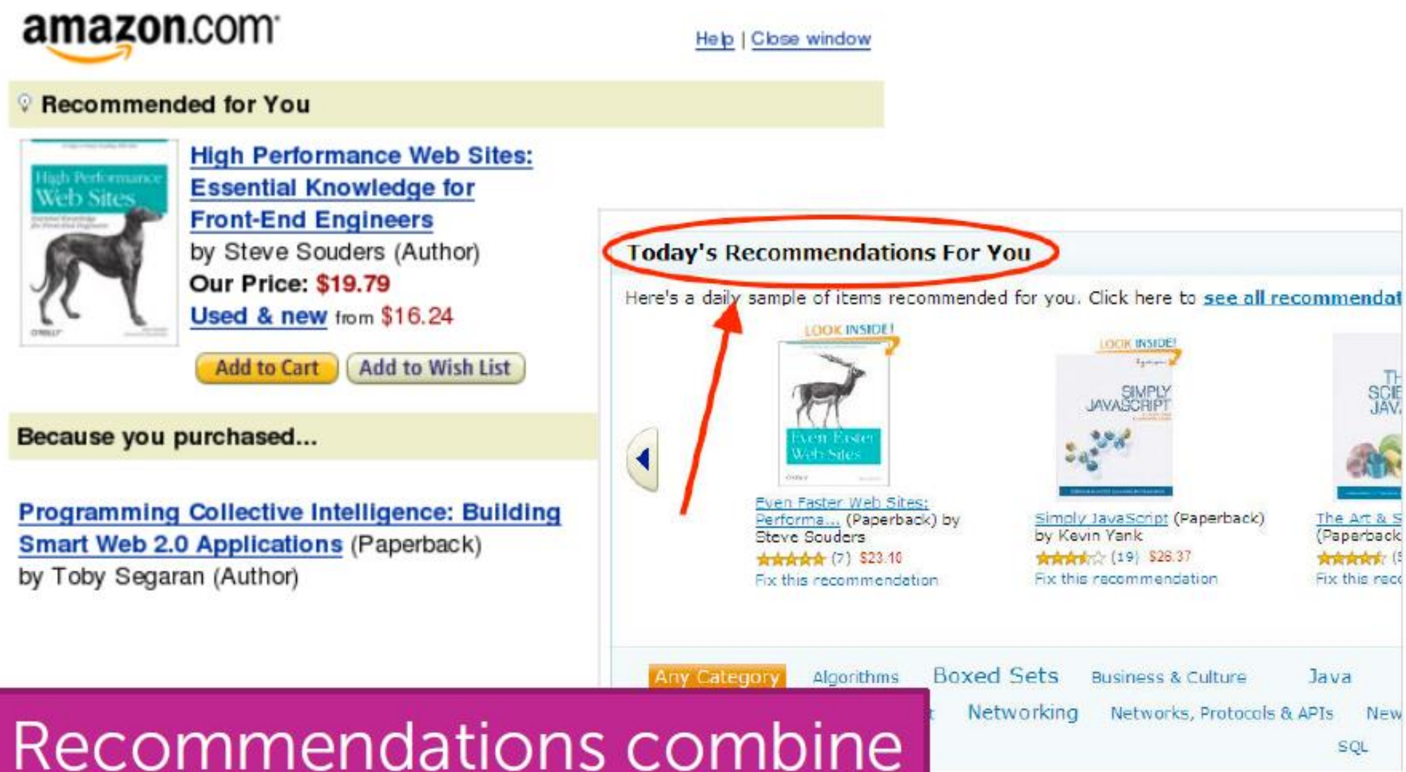
videos

Recomending system:



Connect users with movies they may want to watch

Recomending system:



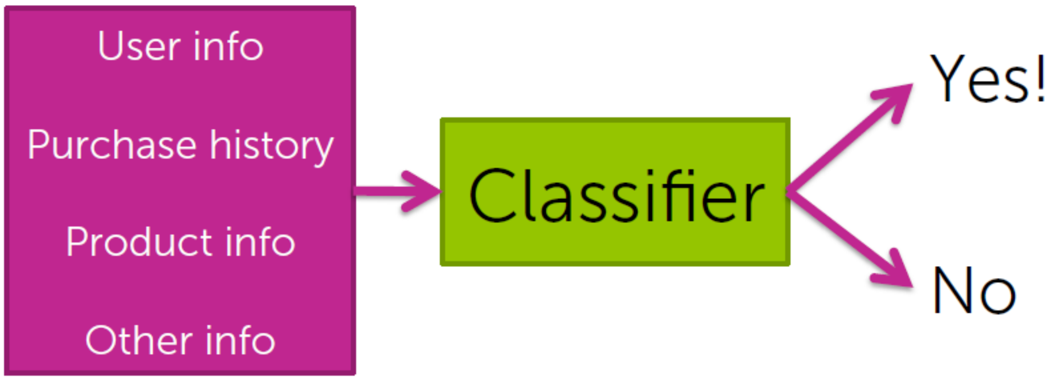
Recommendations combine global & session interests

Recommending system: popularity?

- **Popularność?**
  - **Ranking wg. liczba oglądań**
  - **Nie ma personalizacji**

Recommending system: classification

- Classification?
  - What is probability that I will buy this product?
  - Personalisation: purhase history, monthly and yearly trends, etc.



Recommending system: correlations

- **Analyse correlations. Customers who bought product A also bought product B**
  - **Correlation matrix**

User  purchased *diapers*

1. Look at *diapers* row of matrix

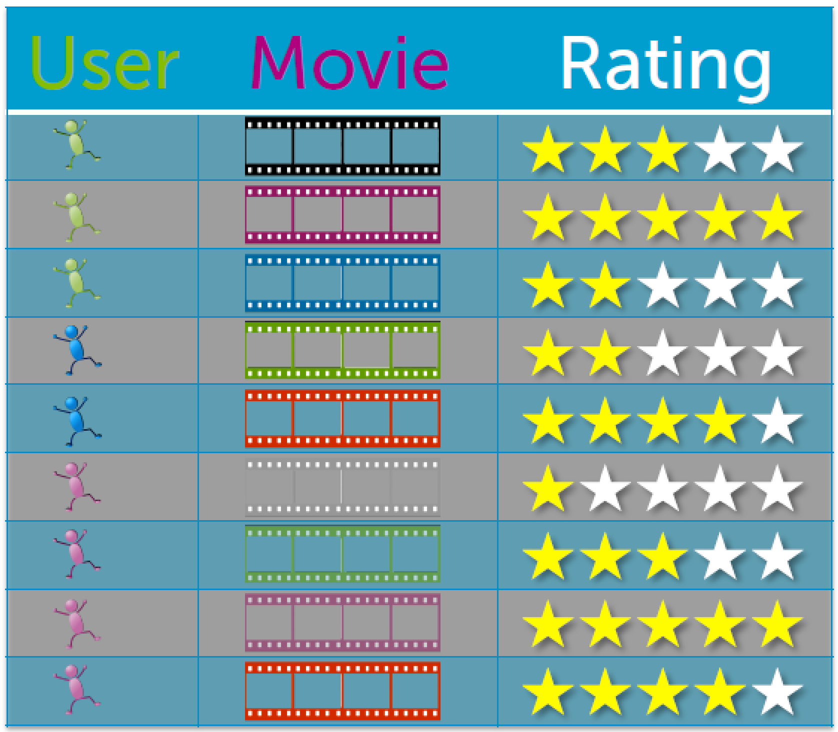
2. Recommend other items with largest counts
   - *baby wipes*, *milk*, *baby food*,...

Recommending system: correlations

- Analyse correlations. Customers who bought product A also bought product B
  - Should we normalise correlation matrix?
  - How to quantify that products are „products”?
- Limitation of correlationss:
  - It is not looking at the purhasing history (trends in time)
  - How to add a new customer (no info on correlations)?

Recommmending system: films

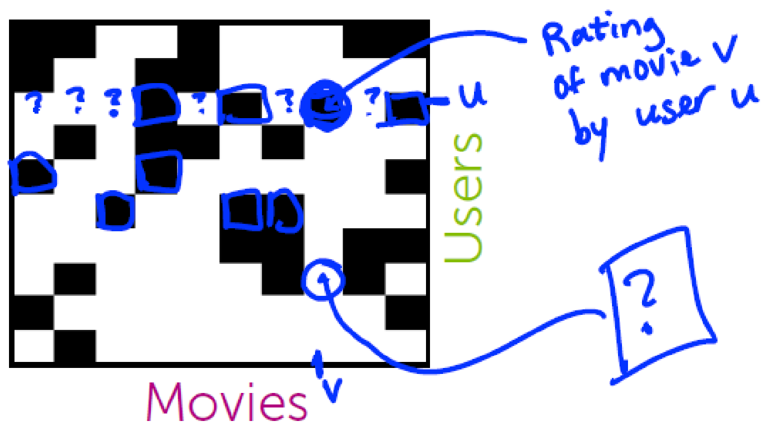
- Users watch movies and rate them



Each user only watches a few of the available movies

Recommending system: films

Rating =

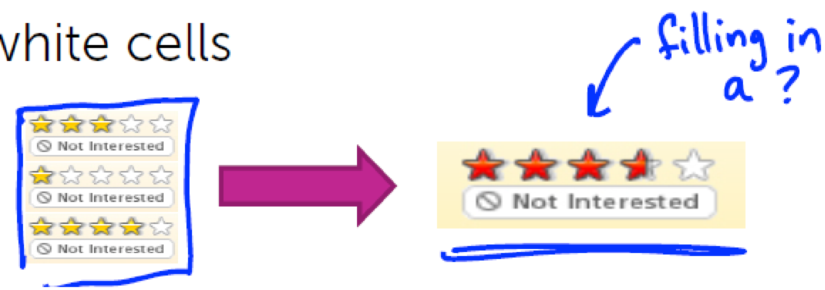


- **Data:** Users score some movies

  Rating(u,v) known for black cells

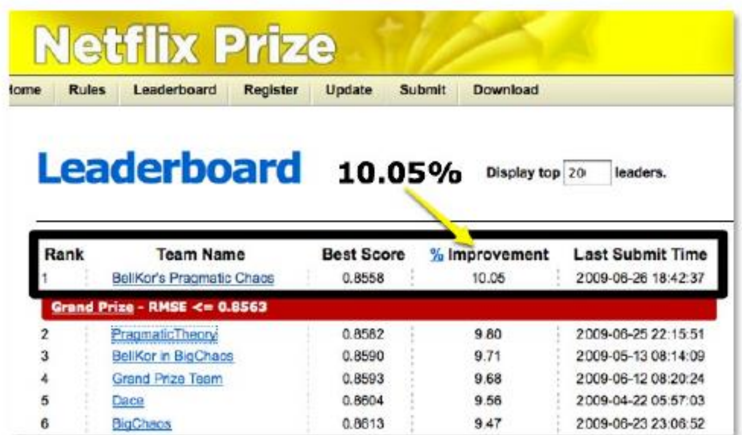
  Rating(u,v) unknown for white cells

- **Goal:** Filling missing data?



Recommending system: optimisation

- Squeezing last bit of accuracy by blending models
- Netflix Prize 2006-2009
  - 100M ratings
  - 17,770 movies
  - 480,189 users
  - Predict 3 million ratings to highest accuracy
  - Winning team blended over 100 models



Recommending system: how effective?

The world of all baby products

Recommending system: how effective?

User likes subset of items

Recommending system: how effective?

How many liked items were recommended?

Recall

$$\frac{\text{\# liked \& shown}}{\text{\# liked}}$$

$$= \frac{3}{5}$$

Recommending system: how effective?

How many recommended items were liked?

Precision

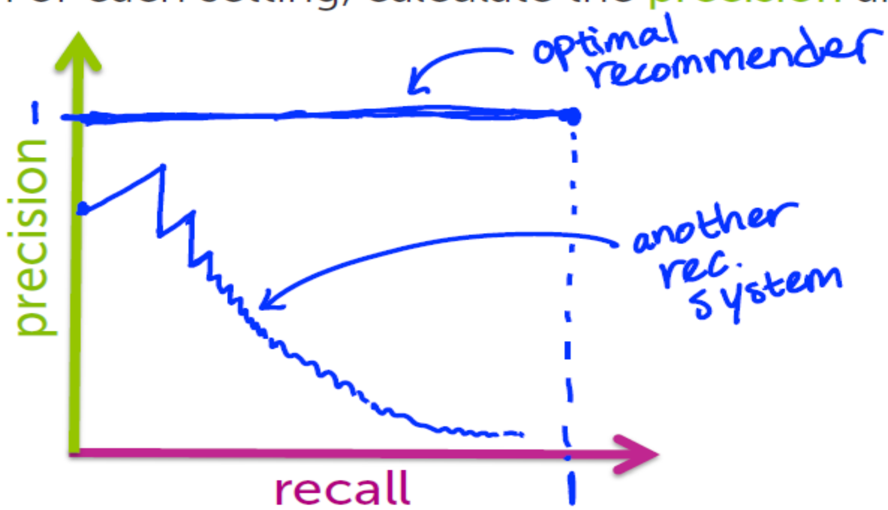
$$\frac{\#\ \text{liked \& shown}}{\#\ \text{shown}}$$

$$= \frac{3}{11}$$

Precision-recall curve

- **Input:** A specific recommender system
- **Output:** Algorithm-specific precision-recall curve

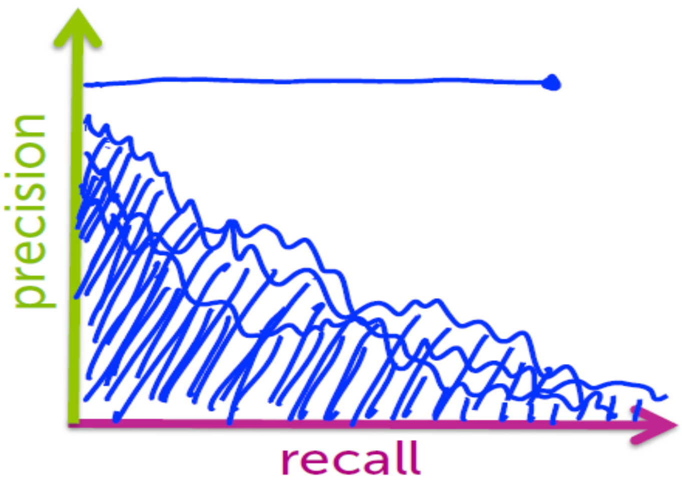
- To draw curve, vary threshold on # items recommended
  - For each setting, calculate the precision and recall



Recommending system: how effective?

Which Algorithm is Best?

- For a given precision, want recall as large as possible (or vice versa)
- One metric: largest area under the curve (AUC)
- Another: set desired recall and maximize precision (precision at k)



Recommending system

Models

- Collaborative filtering
- Matrix factorization
- PCA

Algorithms

- Coordinate descent
- Eigen decomposition
- SVD

Concepts

- Matrix completion, eigenvalues, random projections, cold-start problem, diversity, scaling up