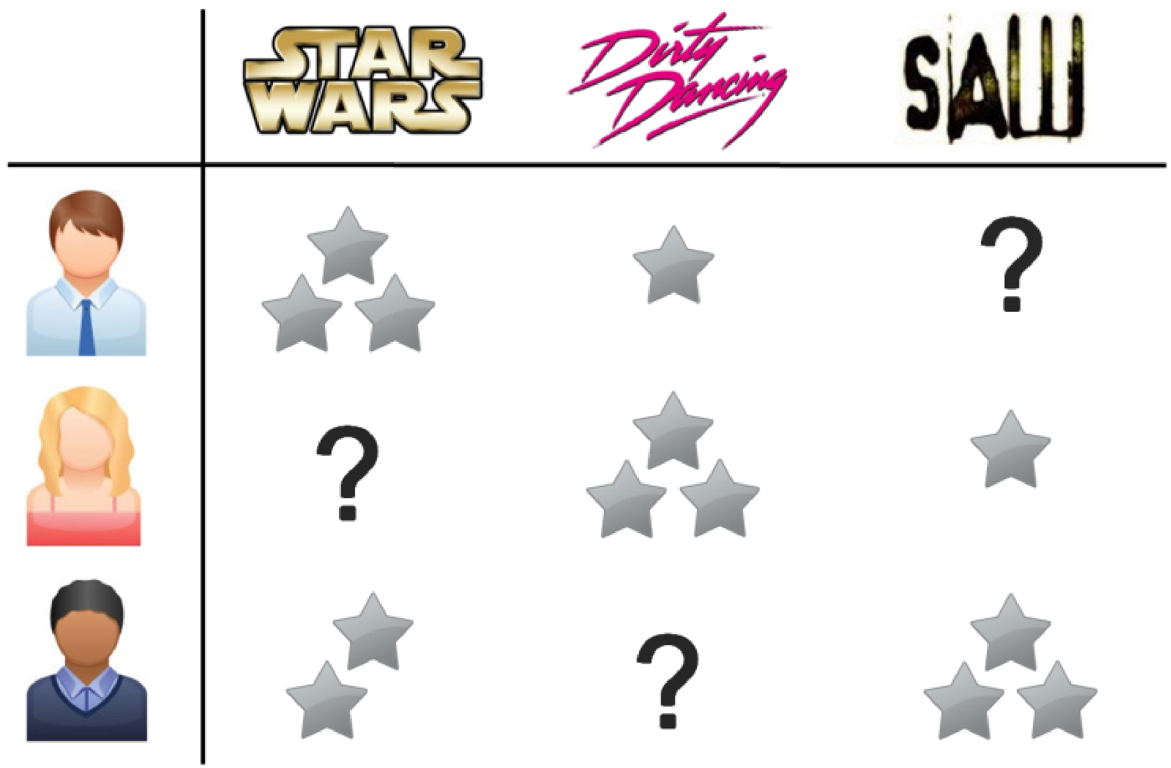
HPI
Hasso Plattner Institut
IT Systems Engineering | Universität Potsdam

Collaborative Filtering
Scalable Data Analysis Algorithms

Claudia Lehmann, Andrina Mascher



Outline

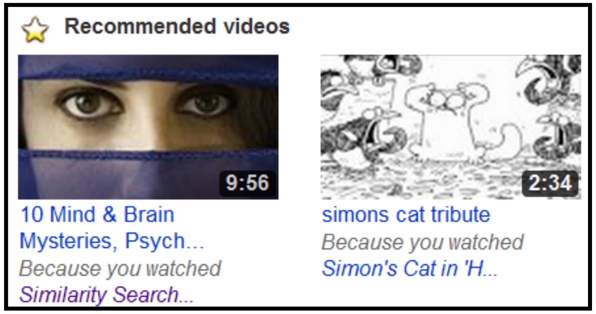
1. Recommendation Systems
   - Applications
   - Types
2. Matrix Factorization
   - Sequential SGD
   - Distributed SGD
3. Dataset

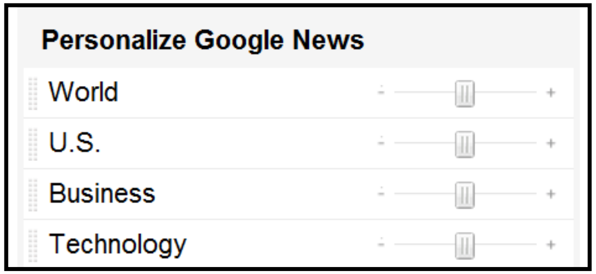
Recommendation Systems Applications

Online stores:

- Product recommendation, e.g. Amazon
- Movie recommendation, e.g. Netflix, YouTube
- News personalization, e.g. Google News

Physical stores cannot adjust store to **individual** user





Recommendation Systems
Types

Recommendation Systems

Content-based Systems

Use item and user properties

genre
actors
director
year

favorite genres
favorite actors
gender
age

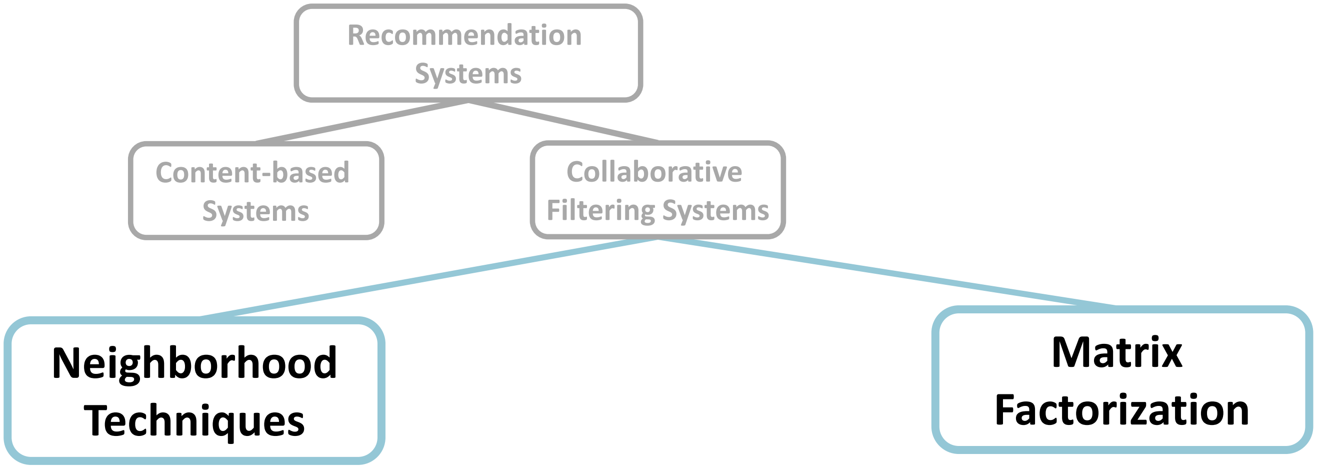
Collaborative Filtering Systems

Use user-item matrix

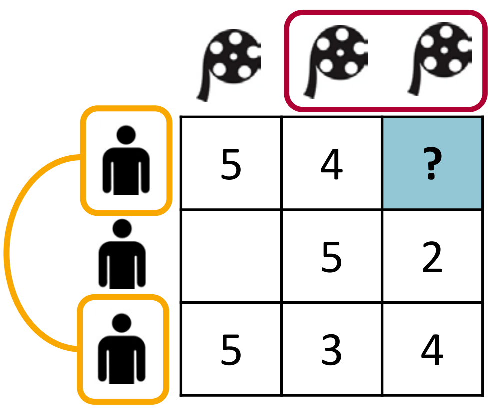
- Needs large dataset and user accounts
- Extend explicit ratings with implicit and binary data (view or buy item)

5	4	
	5	2
2		4

Recommendation Systems
Types



Use rating information from similar users and items



5	4	?
	5	2
5	3	4

Compute similarity between feature vectors of item and user

- Better recommendations
- Better scalability
- Faster
- Implicit features used

Matrix Factorization

Given a sparse matrix *V*: find factors *W* and *H*, so that *W x H* and *V* are most similar

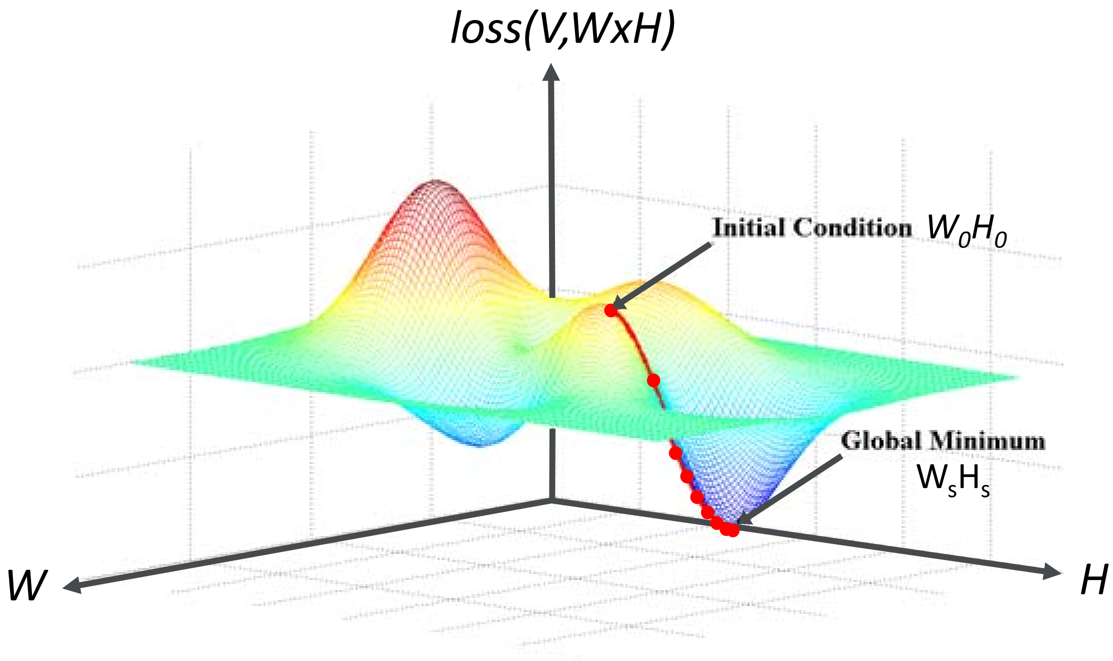
$$H = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \end{bmatrix} \qquad W = \begin{bmatrix} 1 & 2 \\ 2 & 2 \\ 1 & 1 \end{bmatrix} \qquad V = \begin{bmatrix} 5 & 4 & ? \\ ? & 5 & 2 \\ 2 & ? & 4 \end{bmatrix}$$

$$H = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \end{bmatrix} \qquad W = \begin{bmatrix} 1 & 2 \\ 2 & 2 \\ 1 & 1 \end{bmatrix} \qquad WxH = \begin{bmatrix} 4 & 5 & 3 \\ 6 & 6 & 4 \\ 3 & 3 & 2 \end{bmatrix}$$

- Machine learning with training set *V*
- Similarity of original *V* and approximated *W x H*: nonzero squared loss

$$loss(V, W \times H) = \sum_{i,j \in V} (V_{i,j} - W \times H_{i,j})^2$$

Matrix Factorization
Optimize W and H



[1]

step size
(learning rate)
decreases

$$P_{n+1} = P_n - \tfrac{1}{n} \times loss'(V, P_n)$$

partial
differentiation
results in a vector

Find minimum of loss*:*

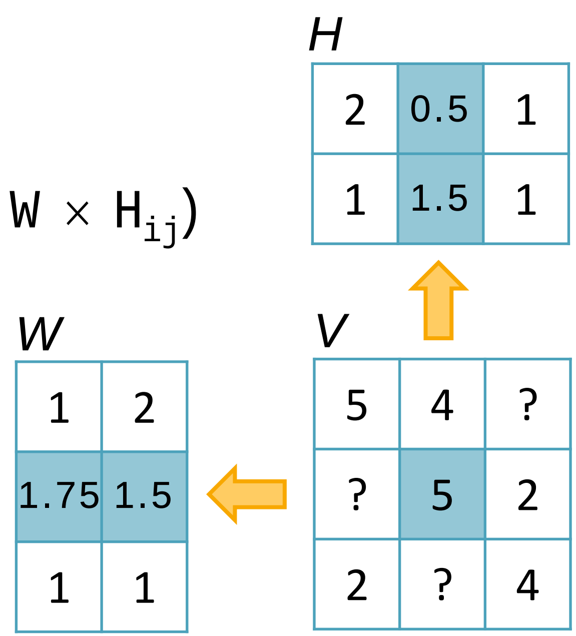
- From arbitrary start configuration *(W_0,H_0)* find path to minimum in *s* steps
  - Go into **opposite** direction of gradient
- Use different starting points to find global minimum

Matrix Factorization
Sequential SGD

Stochastic Gradient Descent (SGD) for matrix factorization:

```
while (minimum not found
       and iterations limit not met
       and loss still decreases):
   pick a random training point (i,j):
       update row in W and column in H with
```

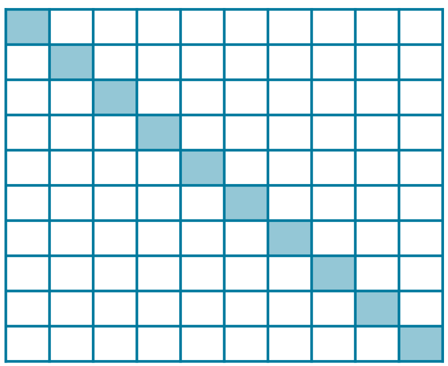
$$(W_{i*}, H_{*j})_{new} = (W_{i*}, H_{*j}) - \frac{1}{n} \times \mathrm{loss}'(V_{ij}, W \times H_{ij})$$



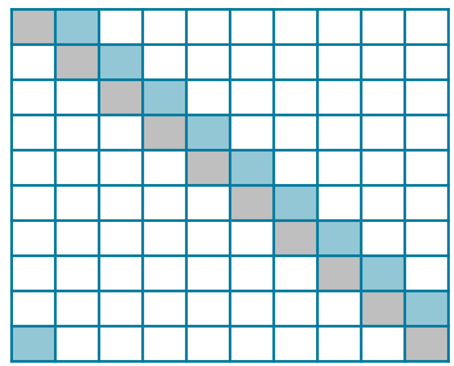
Matrix Factorization
Distributed SGD

Parallel: compute SGD on independent blocks:

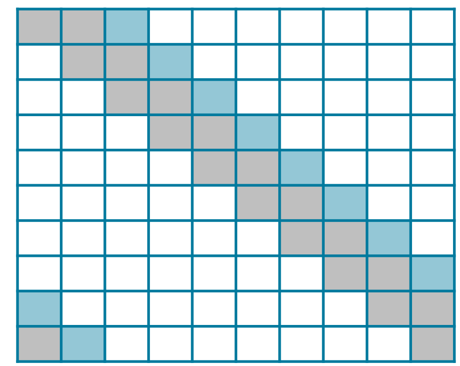
first map phase



second map phase



third map phase



while stop conditions not met:

for each diagonal shift:

MAP compute sequential SGD on blocks in diagonal

REDUCE regroup W_i and H_{j+1} with block $B_{i,j+1}$

MAP compute loss on each block

REDUCE sum up all local loss values

Dataset

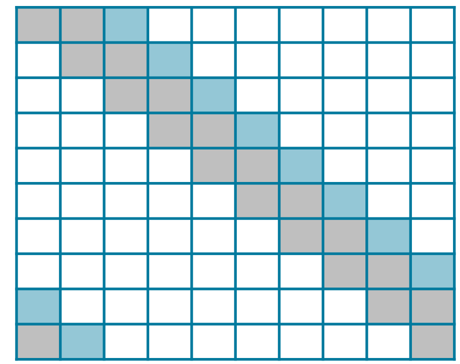
	MovieLens	Netflix
movies	10 K	17 K
users	71 K (min 20 ratings)	480 K
ratings	10 M (10 stars)	100 M (5 stars)
evaluation set	✖	✔

Netflix Spilled Your *Brokeback Mountain* Secret Lawsuit Claims

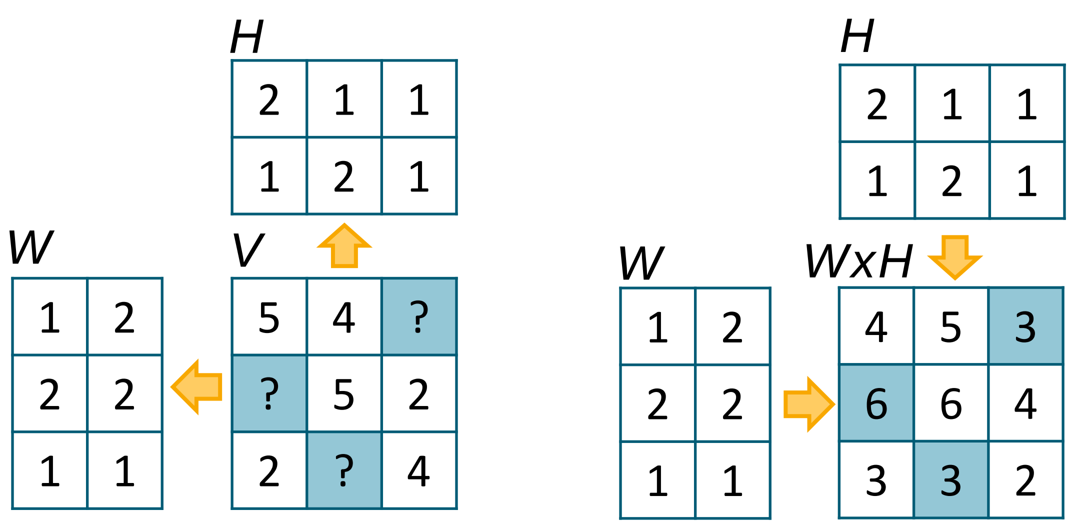
[2]

Summary

Recommendation Systems



Distributed SGD



Matrix Factorization

References

- Anand Rajaraman and Jeff Ullman. *Mining of Massive Datasets*. Cambridge University Press, 2010.
- Rainer Gemulla, Peter J. Haas, Erik Nijkamp, and Yannis Sismanis. *Large-Scale Matrix Factorization with Distributed Stochastic Gradient Descent.* IBM Research Report RJ10481, March 2011.

[1] http://www.mathworks.de/matlabcentral/fx_files/27631/1/fff.png, November 2011

[2] http://www.wired.com/threatlevel/2009/12/netflix-privacy-lawsuit/?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+wired%2Findex+%28Wired%3A+Index+3+%28Top+Stories+2%29%29&utm_content=Google+Feedfetcher, November 2011

Calculation $(W_{i*},H_{*j})_{new} = (W_{i*},H_{*j}) - \frac{1}{n} \times loss'(V_{ij}, W \times H_{ij})$

$$V = \begin{bmatrix} 5 & 4 & ? \\ ? & 5 & 2 \\ 2 & ? & 4 \end{bmatrix} \cong W_{old} = \begin{bmatrix} 1 & 2 \\ 2 & 2 \\ 1 & 1 \end{bmatrix} \times H_{old} = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \end{bmatrix} = WxH_{old} = \begin{bmatrix} 4 & 5 & 3 \\ 6 & 6 & 4 \\ 3 & 3 & 2 \end{bmatrix}$$

$loss(V,WxH_{old})$

$$\begin{array}{c|c|l} +1 & +1 & \\ & +1 & +4 \\ +1 & & +4 = \mathbf{12} \end{array}$$

$$H_{new} = \begin{bmatrix} 2 & 0.5 & 1 \\ 1 & 1.5 & 1 \end{bmatrix}$$

$$W_{new} = \begin{bmatrix} 1 & 2 \\ 1.75 & 1.5 \\ 1 & 1 \end{bmatrix} \Leftarrow V = \begin{bmatrix} 5 & 4 & ? \\ ? & 5 & 2 \\ 2 & ? & 4 \end{bmatrix}$$

$$H_{new} = \begin{bmatrix} 2 & 0.5 & 1 \\ 1 & 1.5 & 1 \end{bmatrix}$$

$$W_{new} = \begin{bmatrix} 1 & 2 \\ 1.75 & 1.5 \\ 1 & 1 \end{bmatrix} \Rightarrow WxH_{new} = \begin{bmatrix} 4 & 3.5 & 3 \\ 5 & 3.125 & 3.25 \\ 3 & 2 & 2 \end{bmatrix}$$

$loss(V,WxH_{new})$ = **11.32**

Calculation

H_{old}

2	1	1
1	2	1

W_{old}

1	2
2	2
1	1

V

5	4	?
?	5	2
2	?	4

H_{new}

2	0.5	1
1	1.5	1

W_{new}

1	2
1.75	1.5
1	1

V

5	4	?
?	5	2
2	?	4

$$(W_{i*})_{new} := W_{i*} - \frac{1}{n} \times \frac{\partial}{\partial W_{i*}} loss(V_{ij}, W \times H_{ij})$$

$$= \begin{pmatrix} 2 \\ 2 \end{pmatrix} - \frac{1}{n} \times \frac{\partial}{\partial \begin{pmatrix} W_{21} \\ W_{22} \end{pmatrix}} (5 - (\underset{W_{21}}{2} \times 1 + \underset{W_{22}}{2} \times 2))^2$$

$$= \begin{pmatrix} 2 \\ 2 \end{pmatrix} - \frac{1}{n} \times \begin{pmatrix} 2 \times (5-6) \times (-1) \\ 2 \times (5-6) \times (-2) \end{pmatrix}$$

$$= \begin{pmatrix} 2 \\ 2 \end{pmatrix} - \frac{1}{8} \times \begin{pmatrix} 2 \\ 4 \end{pmatrix}$$

$$= \begin{pmatrix} 1.75 \\ 1.5 \end{pmatrix}$$

assume step n = 8

$$(H_{*j})_{new} := H_{*j} - \frac{1}{n} \times \frac{\partial}{\partial H_{*j}} loss(V_{ij}, W \times H_{ij})$$

$$= \begin{pmatrix} 1 \\ 2 \end{pmatrix} - \frac{1}{n} \times \frac{\partial}{\partial \begin{pmatrix} H_{12} \\ H_{22} \end{pmatrix}} (5 - (2 \times \underset{H_{12}}{1} + 2 \times \underset{H_{22}}{2}))^2$$

$$= \begin{pmatrix} 1 \\ 2 \end{pmatrix} - \frac{1}{n} \times \begin{pmatrix} 2 \times (5-6) \times (-2) \\ 2 \times (5-6) \times (-2) \end{pmatrix}$$

$$= \begin{pmatrix} 1 \\ 2 \end{pmatrix} - \frac{1}{8} \times \begin{pmatrix} 4 \\ 4 \end{pmatrix}$$

$$= \begin{pmatrix} 0.5 \\ 1.5 \end{pmatrix}$$