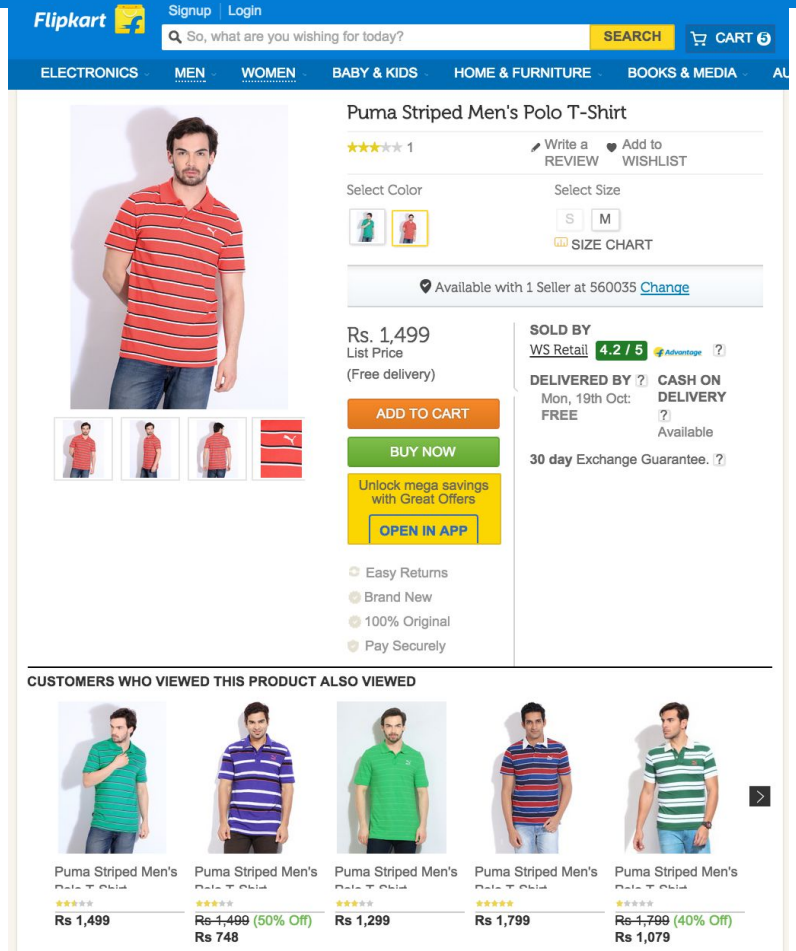
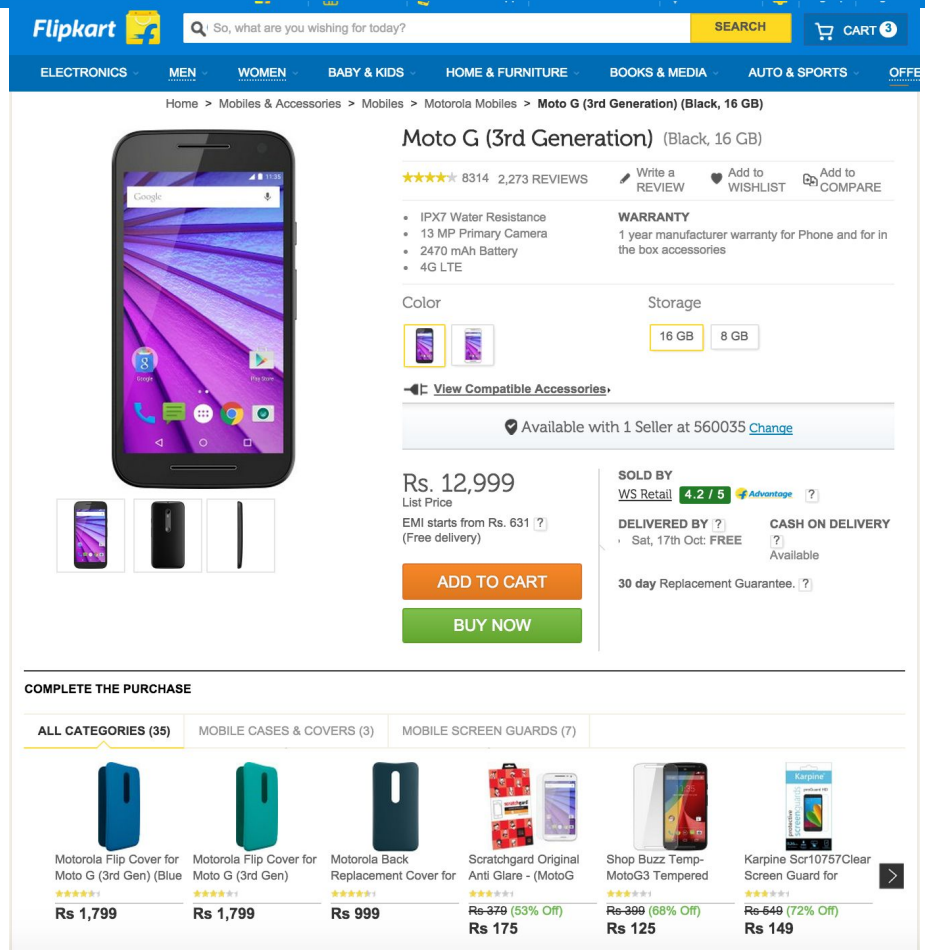
Recommendation Systems

Akash Khandelwal, Avijit Saha, Mohit Kumar, Vivek Mehta

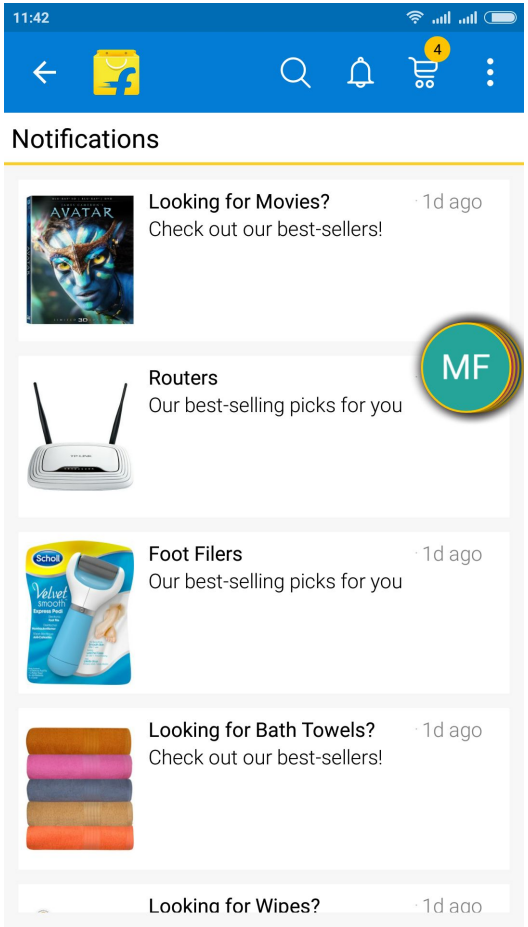
Happy Festival Season!!!

Recommendations - Overview





Recommendation Overview



- Related product recommendation (Flipkart)
- Movie recommendation (Netflix)
- Web page ranking (Google)
- Social recommendation (Facebook)
- News content recommendation (Yahoo)
- Priority inbox & spam filtering (Google)
- Online dating (OK Cupid)
- Computational Advertising (Yahoo)
- Event recommendation

Agenda

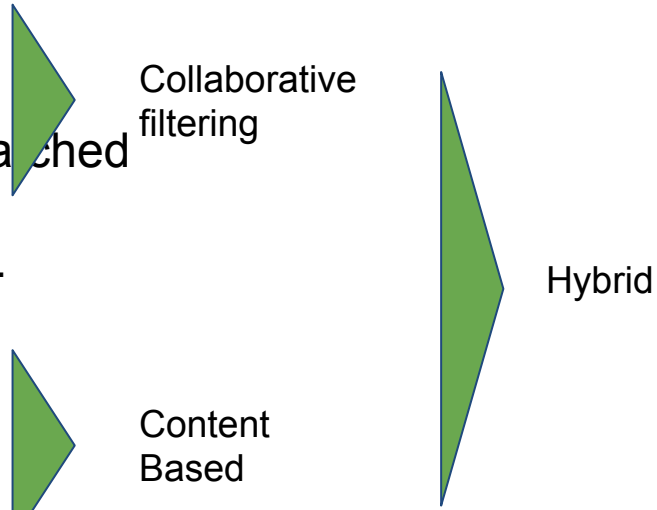
- Type of Methods
  - Collaborative filtering
  - Content based
  - Hybrid
- Collaborative Filtering
  - Neighbourhood methods
  - Matrix factorization

Reference

- Slides are prepared by leveraging internet heavily
  - http://www2.research.att.com/~volinsky/papers/ieeecomputer.pdf
  - http://www.slideshare.net/xamat
  - http://alex.smola.org/teaching/berkeley2012/slides/8_Recommender.pdf
  - http://blog.comsysto.com/2013/04/03/background-of-collaborative-filtering-with-mahout/
  - Simon funk's blog

Data To leverage

- User activity
  - Movies
    - Ratings, reviews, movies rented, watched
  - E-commerce
    - Browse, wishlist, share, ratings, etc.
- Item details
  - Movies
    - Actors, Directors, Genre, etc.
  - E-commerce
    - Product attributes, images, description



CF - Neighbourhood Methods

Neighbourhood Method

CUSTOMERS WHO VIEWED THIS PRODUCT ALSO VIEWED

GAS Printed Men's Round Neck T-Shirt

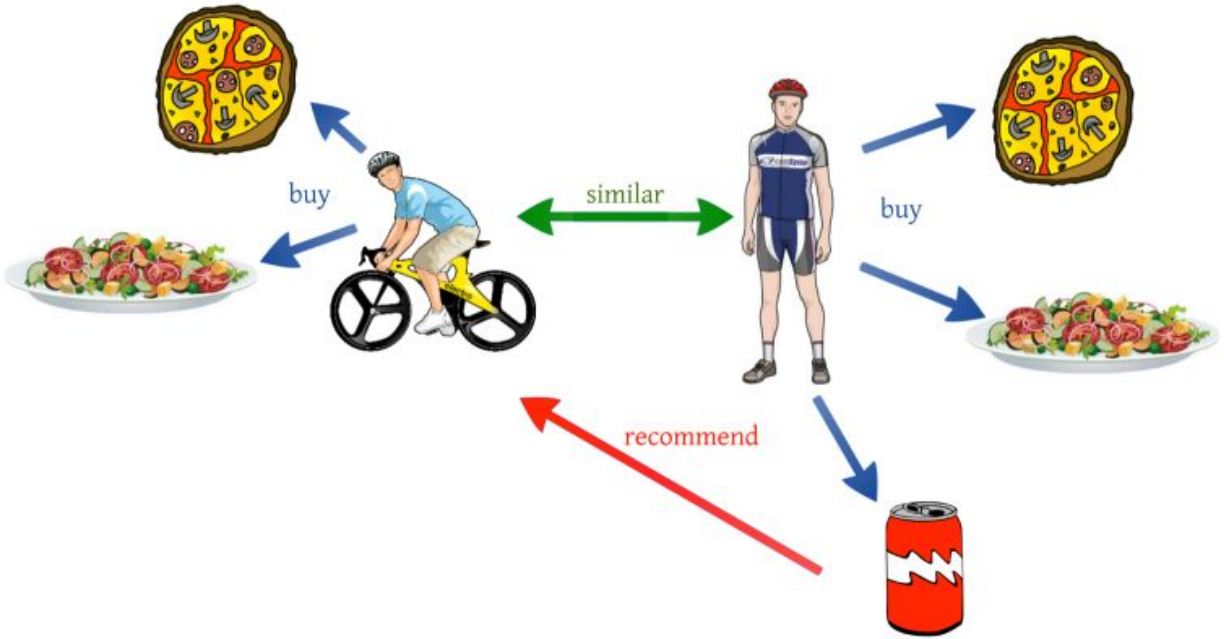
GAS Printed Men's Round Neck T-Shirt

GAS Printed Men's Round Neck T-Shirt

GAS Printed Men's Round Neck T-Shirt

Gas Printed Men's Round Neck T-Shirt

Neighbourhood Method



Neighbourhood Method - Setup

Users

Items

3		2	1		
4	4	?	5	3	
1	2	2		4	
2		2		4	

(user, user) similarity

- r_{ui} = ?
- For user *u*, find similar users (u')
- Find likeness of item *i* for similar users.

$$r_{u,i} = \frac{1}{N} \sum_{u' \in U} r_{u',i}$$

$$r_{u,i} = k \sum_{u' \in U} \text{simil}(u, u') r_{u',i} \qquad k = 1 / \sum_{u' \in U} |\text{simil}(u, u')|$$

$$r_{u,i} = \bar{r_u} + k \sum_{u' \in U} \text{simil}(u, u')(r_{u',i} - \bar{r_{u'}})$$

(user, user) similarity

- Pearson correlation

$$\text{simil}(x,y) = \frac{\sum_{i \in I_{xy}} (r_{x,i} - \bar{r_x})(r_{y,i} - \bar{r_y})}{\sqrt{\sum_{i \in I_{xy}} (r_{x,i} - \bar{r_x})^2 \sum_{i \in I_{xy}} (r_{y,i} - \bar{r_y})^2}}$$

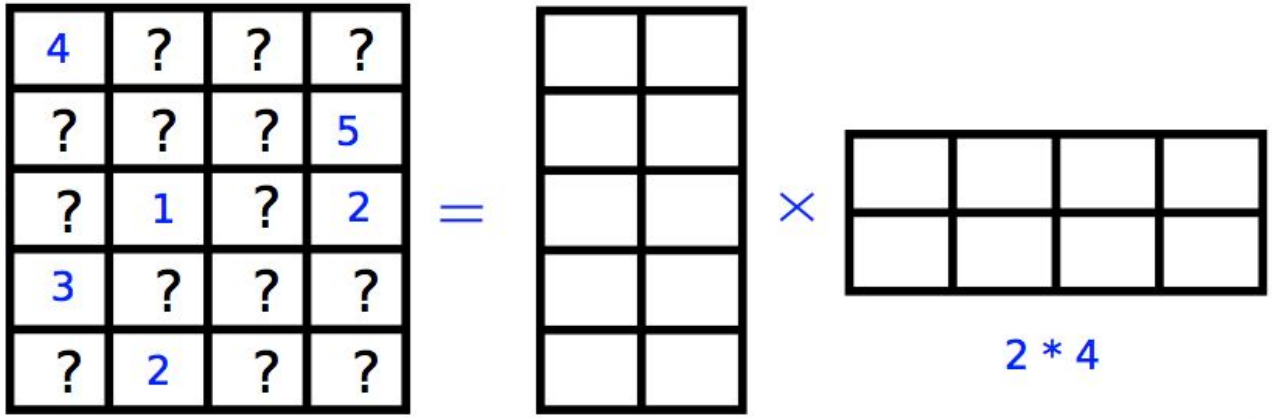
(item,item) similarity

- Similar to (user, user) similarity
- Find users who have liked a product
- Find other products liked by these users

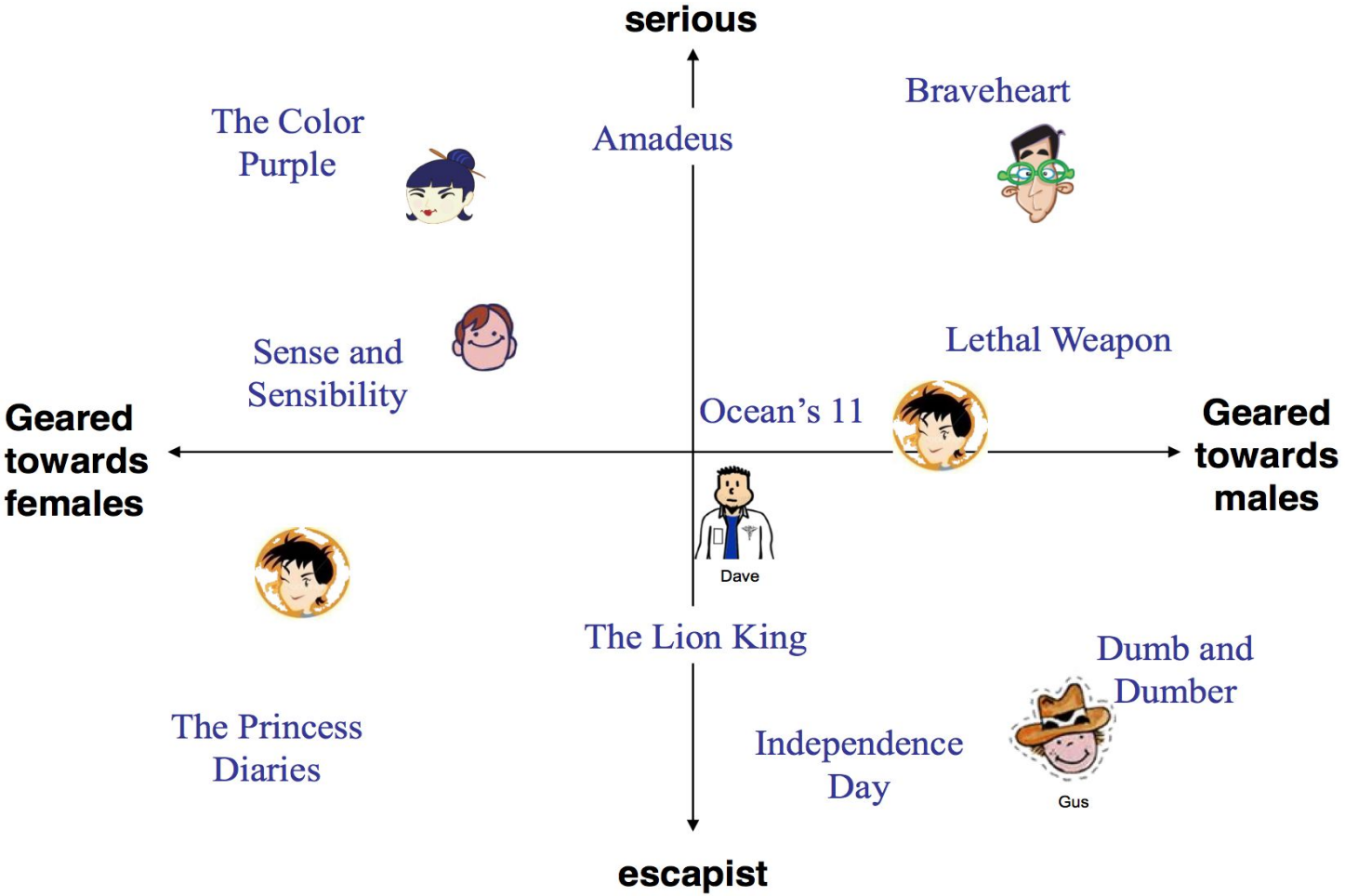
Properties

- Pros:
  - Intuitive and easy to explain to user
  - Easy to setup
  - Produces good-enough results
- Cons:
  - Scalability
  - Sparseness -> coverage -> cold start

Matrix factorization



$$\hat{r}_{ui} = q_i^T p_u.$$


serious
The Color Purple
Amadeus
Braveheart
Sense and Sensibility
Lethal Weapon
Geared towards females
Ocean's 11
Geared towards males
Dave
The Lion King
Dumb and Dumber
The Princess Diaries
Independence Day
Gus
escapist

Matrix factorization

- Similar to SVD, but missing values!!
- Overfitting is one major problem
- SGD for learning latent factors
- Objective function:

$$\min_{q^*,p^*} \sum_{(u,i)\in K} (r_{ui} - q_i^T p_u)^2 + \lambda(\| q_i \|^2 + \| p_u \|^2)$$

Regularization Constant

Stochastic gradient descent

- Compute error

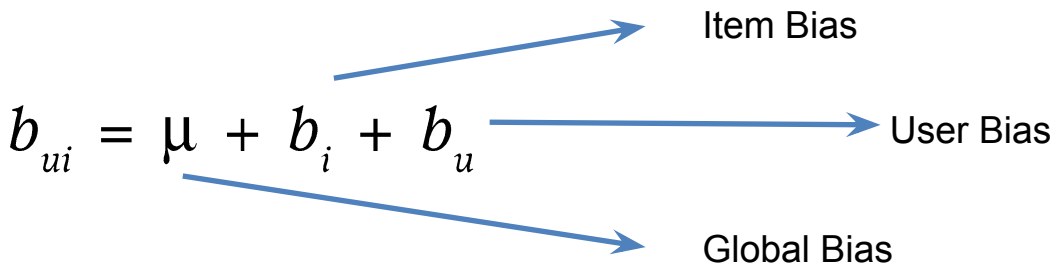
$$e_{ui} \overset{def}{=} r_{ui} - q_i^T p_u.$$

- Adjust the factors

$$q_i \leftarrow q_i + \gamma \cdot (e_{ui} \cdot p_u - \lambda \cdot q_i)$$
$$p_u \leftarrow p_u + \gamma \cdot (e_{ui} \cdot q_i - \lambda \cdot p_u)$$

Learning rate

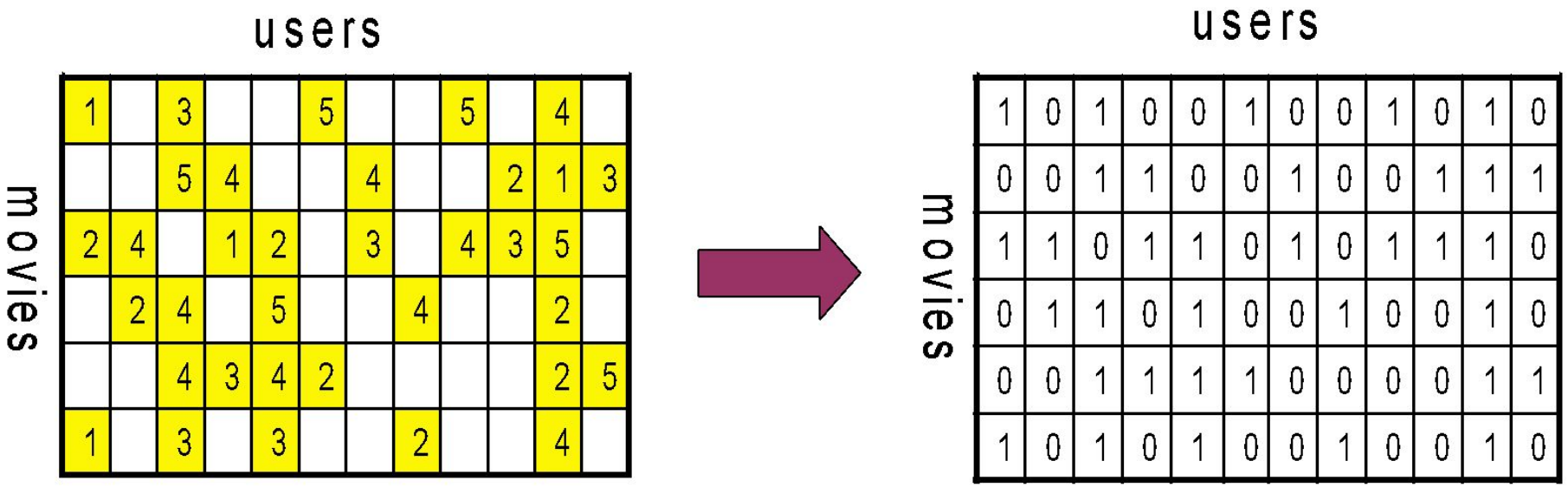
Adding Biases



$$\hat{r}_{ui} = \mu + b_i + b_u + q_i^T p_u$$

$$\min_{p^*, q^*, b^*} \sum_{(u,i) \in K} (r_{ui} - \mu - b_u - b_i - p_u^T q_i)^2 + \lambda (\| p_u \|^2 + \| q_i \|^2 + b_u^2 + b_i^2)$$

Implicit feedback



Implicit feedback

- Objective function with implicit feedback formulation

$$\min_{p^*,q^*,b^*} \sum_{(u,i)\in\kappa} c_{ui}(r_{ui} - \mu - b_u - b_i - p_u{}^T q_i)^2 + \lambda\left(\| p_u \|^2 + \| q_i \|^2 + b_u{}^2 + b_i{}^2\right)$$

Thank You